UNITED STATES OF AMERICA

DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

+ + +

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH

MEDICAL DEVICES ADVISORY COMMITTEE

+ + +

MICROBIOLOGY MEDICAL DEVICES PANEL

+ + +

June 29, 2011
8:00 a.m.

Holiday Inn, Ballroom 2
2 Montgomery Village Avenue
Gaithersburg, Maryland

PANEL MEMBERS:

ANGELA M. CALIENDO, M.D., Ph.D.	Voting Chair
MARGARET HAMMERSCHLAG, M.D.	Voting Member
LIZZIE HARRELL, Ph.D.	Voting Member
KELLY J. HENRICKSON, M.D.	Voting Member
KENNETH H. RAND, M.D.	Voting Member
LARRY G. REIMER, M.D.	Voting Member
KATHLEEN G. BEAVIS, M.D.	Temporary Non-Voting Member
KAREN C. CARROLL, M.D.	Temporary Non-Voting Member
PHYLLIS DELLA-LATTA, Ph.D.	Temporary Non-Voting Member
ANN GINSBERG, M.D.	Temporary Non-Voting Member
FRED GORDIN, M.D.	Temporary Non-Voting Member
CAROL HAMILTON, M.D.	Temporary Non-Voting Member
MICHAEL L. WILSON, M.D.	Temporary Non-Voting Member
JANINE JANOSKY, Ph.D.	Temporary Non-Voting Member
JoELLEN DeLUCA	Patient Representative
GARY L. DUERHRING, Ph.D.	Consumer Representative
MICHAEL L. TOWNS, M.D.	Industry Representative
SHANIKA CRAIG, M.H.A., M.B.A.	Designated Federal Officer

FDA REPRESENTATIVES:

ALBERTO GUTIERREZ, Ph.D.
Director, Office of In Vitro Diagnostic Device Evaluation and Safety, CDRH

SALLY A. HOJVAT, Ph.D.
Director, Division of Microbiology Devices, Office of In Vitro Diagnostic Device Evaluation and Safety, CDRH

MARJORIE SHULMAN
Acting Director, Premarket Notification Staff, Office of Device Evaluation, CDRH

FREDDIE M. POOLE
Team Leader, Bacteriology Division of Microbiology Devices, Office of Device Evaluation and Safety

AMANDA SENA
Press Contact

FDA PRESENTERS:

STEVEN GITTERMAN, M.D., Ph.D.
Medical Officer, DMD/OIVD/CDRH

WILLIAM BURMAN, M.D.
Interim Director, Denver Public Health

KENNETH G. CASTRO, M.D.
Director, Division of Tuberculosis Elimination, CDC

OPEN PUBLIC HEARING SPEAKERS:

MELISSA MILLER, Ph.D., D(ABMM)
FRED C. TENOVER, Ph.D., D(ABMM)
JAMES PEARSON, Dr.P.H., BCLD
KHATEREH CALLEJA, J.D.
ANTONINO CATANZARO, M.D.

INDEX

PAGE

CALL TO ORDER - Angela M. Caliendo, M.D., Ph.D. 5

INTRODUCTIONS 5

Alberto Gutierrez, Ph.D. 9

Sally A. Hojvat, Ph.D. 10

CONFLICT OF INTEREST STATEMENT - Shanika Craig, M.H.A., M.B.A. 12

GENERAL ANNOUNCEMENTS - Shanika Craig, M.H.A., M.B.A. 15

DEVICE CLASSIFICATION AND RECLASSIFICATION PROCESS - Marjorie Shulman 16

FDA PRESENTATIONS

Current FDA Regulation of *In Vitro* Diagnostics for Tuberculosis - Steven Gitterman, M.D., Ph.D. 20

Diagnostic Tests for TB: A Clinician's Perspective - William Burman, M.D. 37

The Public Health Implications of Reclassification - Kenneth G. Castro, M.D. 66

Questions from the Panel 82

OPEN PUBLIC HEARING

Melissa Miller, Ph.D., D(ABMM) 118

Fred C. Tenover, Ph.D., D(ABMM) 122

James Pearson, Dr.P.H., BCLD 125

Khatereh Calleja, J.D. 128

Antonino Catanzaro, M.D. 131

Questions from the Panel 133

INDEX (Cont.)

	PAGE
PANEL DELIBERATIONS AND FDA QUESTIONS	
Question 1	140
Question 2	199
Question 3	235
SUMMARY AND NEXT STEPS - Angela M. Caliendo, M.D. and Sally Hojvat, Ph.D.	276
ADJOURNMENT	278

M E E T I N G

(8:00 a.m.)

DR. CALIENDO: Okay, good morning, everybody. I'd like to call this meeting of the Microbiology Devices Panel to order.

I'm Dr. Angie Caliendo, and I'm the Chairperson of this Panel. And just as a way of background, I'm a Professor and Vice Chair of Pathology and Laboratory Medicine at Emory. I'm the director of the clinical labs there, with specific responsibilities as a medical director of micro and molecular diagnostics lab and also an adult infectious disease physician.

So at this meeting today, the Panel will discuss and make recommendations regarding the possible reclassification of molecular diagnostics for the rapid detection of *Mycobacterium tuberculosis* complex and the detection of genetic mutations which confer antibiotic resistance in MTB complex.

Before we begin, I'd like to ask our distinguished Panel members and FDA staff seated at the table to introduce themselves. Please state your name, your area of expertise, your position and affiliation. So, Sally, if you'll please start.

DR. HOJVAT: Yes. My name is Sally Hojvat, and I'm with the FDA. I'm the Director of the Division of Microbiology Devices in the Office of In Vitro Diagnostic Devices.

DR. HAMILTON: Carol Dukes Hamilton, FHI, in North Carolina

and Duke University. I'm an adult infectious disease physician. I've been involved as a TB controller and medical director of the TB program in North Carolina for eight years and involved as well in TB clinical research and now do international TB public health work and research.

DR. HENRICKSON: Kelly Henrickson. I'm a pediatric infectious disease physician and a molecular virologist, and I run the Midwest Respiratory Virus Program's clinical laboratory and have been involved in development of molecular devices for 21 years.

DR. DELLA-LATTA: I'm Phyllis Della-Latta. I direct the Clinical Microbiology Service for Columbia University Medical Center, New York Presbyterian Hospital, and also Children's Hospital of New York. We handle very immunocompromised patients and have a very large pediatric division. My interests have been in many areas, but particularly in TB since the epidemic in New York City in the early 1990s.

DR. REIMER: I'm Larry Reimer. I'm an adult infectious disease specialist and clinical microbiologist at the University of Utah. I've been there for 25 years, directing the clinical microbiology lab at the Veterans Affairs Medical Center there, and affiliated with ARUP and its process in terms of clinical microbiology.

DR. HAMMERSCHLAG: I'm Margaret Hammerschlag. I'm the pediatric IV specialist --

DR. CALIENDO: You have to press the mike.

DR. HAMMERSCHLAG: Hi. Margaret Hammerschlag. I'm the Director of Pediatric Infectious Diseases at the State University of New York Downstate Medical Center in Brooklyn, and I guess we see just about everything.

DR. GORDIN: Fred Gordin. I'm an adult infectious disease physician at the VA Medical Center here in Washington and part of George Washington University. I've been active in HIV and tuberculosis mycobacterial research for 25 years.

DR. BEAVIS: I'm Kathleen Beavis. I'm a pathologist and I run the microbiology and virology laboratories at Cook County Hospital in Chicago.

DR. RAND: I'm Dr. Kenneth Rand. I'm an adult infectious disease specialist and also a clinical microbiologist. I'm at the University of Florida Shands Hospital, where I direct the clinical laboratory and then provide special direction for the clinical microbiology and virology laboratories.

MS. CRAIG: My name is Shanika Craig. I'm the DFO for this Committee.

DR. WILSON: Mike Wilson. I'm the Director of the Department of Pathology and Laboratory Services at Denver Health, which includes the laboratories for the Denver Public Health Department. I'm also a Professor of Pathology at the University of Colorado School of Medicine.

DR. JANOSKY: Janine Janosky, Vice President of the Austen BioInnovation Institute in Akron. I'm also a Professor of Statistics at the University of Akron and an Adjunct Professor in Family and Community Medicine at Northeastern Ohio Medical University.

DR. GINSBERG: Good morning. I'm Ann Ginsberg. I'm originally trained in both basic molecular biology and pathology. For the last seven years, I've been the Chief Medical Officer at the Global Alliance for TB Drug Development.

DR. HARRELL: I'm Lizzie Harrell. I'm the Associate Director of Clinical Microbiology at Duke University Medical Center, where I've been for almost 33 years. I'm also a research professor of molecular genetics microbiology and pathology, and we have been involved with mycobacteria for all these years.

DR. CARROLL: Good morning. I'm Karen Carroll. I'm Professor of Pathology and Medicine at the Johns Hopkins University School of Medicine, and I am the Director of the Division of Medical Microbiology for the Johns Hopkins Medical Institutions. I've had a longstanding interest in diagnostics for healthcare-associated infections as well as tuberculosis.

DR. DeLUCA: Good morning. I'm JoEllen DeLuca. I'm the Patient Representative. Since I learned that my grandmother died of TB in the 1930s, it sort of piqued my interest in TB as a subject, and later as a teacher, nurses would come around and give the timed test, but we were

required to do the reading of it. So all along the way, I've enjoyed learning more about TB and I worked in a TB san for a summer job.

DR. TOWNS: Good morning. I'm Michael Towns. I'm the Worldwide Vice President for BD Diagnostic Systems, and I'm the Industry Representative on this Panel, and my background is in both clinical microbiology and adult infectious diseases.

DR. DUERHRING: Good morning. Gary Duerhring. I'm the Consumer Representative on this Panel. I'm also Adjunct Faculty for the Graduate Studies and Health Administration at Central Michigan University.

DR. CALIENDO: So thanks to everybody.

If you're in the audience and you haven't already signed in, please sign one of the sheets out at the table.

At this point I'd like to give Dr. Alberto Gutierrez an opportunity to say a few words.

DR. GUTIERREZ: Good morning. My name is Alberto Gutierrez. I'm the Office Director for the Office of In Vitro Diagnostic Safety and Evaluation.

I wanted to welcome you all to this Panel. This is an important part of the FDA decision making. Our mission is both to protect and to promote public health, and to do so we need to set regulatory bars that are appropriate for the risk of the devices that we regulate. And this process of trying to decide what the regulatory part should be for TB is very important.

So your deliberations today will form the basis for us to make a decision that we consider very important.

So I want to thank you. This is not usually kind of a thankless job. You know, the government brings you to beautiful Gaithersburg and doesn't provide the best accommodations in the world, but your work is really important and it does really -- it matters and it matters to us to be able to make a decision that makes sense. So thank you very much.

Sally is going to do an introduction.

DR. HOJVAT: Okay, got that sorted out.

So I just wanted to go through the objectives of the meeting from the point of view of FDA, and these are the ones we'd like to have your advice on. Oh, it's not showing up.

Okay. So objectives of the advice that you're going to give to us today. The questions we have are, Can the current Class III devices, such as the nucleic acid test and assays for mycobacterial detection from direct specimens and/or indirect assays such as the IGRAs or interferon gamma release assays be reclassified to Class II? They're currently at Class III, and we could do that through development of special controls guidance.

And, similarly, can devices for the detection of antibacterial resistance mutations in direct specimens where *M. tuberculosis* has been detected be initially classified as Class II?

You will have noticed through your reading that these are

unclassified at the moment because we have not had any submissions in. And as we talked about yesterday, they would come in as Class III if they're unclassified, we haven't seen them before, and we could down-classify them through the de novo process to Class II, if that was your recommendation.

So I think it was emphasized that it's not a meeting where we vote on whether the FDA should reclassify these devices, but it is a meeting for FDA to obtain your expert input, and expert input from open public presentations, into whether there are special controls that can be developed to mitigate the risks from potential false negative or false positive test results for devices intended to detect TB directly from patient specimens. So those are the objectives.

And what we've set up as an agenda is the following. You will be able to find what is FDA's process for reclassifying a diagnostic device. We had an introduction this morning on that. You'll hear a presentation from FDA on how we currently regulate tuberculosis diagnostics and then a presentation from an individual who uses these tests. He'll be talking about what tests are currently used here in the U.S. to diagnose latent and active tuberculosis. And from the CDC we'll have a presentation about what the public health implications of reclassifying tuberculosis diagnostics would be. And then we're open for public comment and then your committee discussions.

So hopefully that agenda will help you have a good discussion

later on this afternoon. Thank you.

DR. CALIENDO: Thank you, Sally.

We're now going to hear from Shanika Craig.

MS. CRAIG: Good morning, everyone. I will now read the Conflict of Interest Statement.

FDA Conflict of Interest Disclosure Statement, particular matter of general applicability for the Microbiology Devices Panel of the Medical Devices Advisory Committee on June 29th, 2011.

The Food and Drug Administration is convening today's meeting of the Microbiology Devices Panel of the Medical Devices Advisory Committee under the authority of the Federal Advisory Act of 1972. With the exception of the industry representative, all members and consultants of the Panel are special Government employees or regular Federal employees from other agencies and are subject to Federal conflict of interest laws and regulations.

The following information on the status of the Panel's compliance with Federal ethics and conflict of interest laws covered by, but not limited to, those found in the U.S. Code 18-208 and 712 of the Federal Food, Drug and Cosmetic Act are being provided to participants in today's meeting and to the public.

FDA has determined that members and consultants of this Panel are in compliance with the Federal ethics and conflict of interest laws.

Under U.S. Code 18-208, Congress has authorized FDA to grant waivers to special Government employees who have financial conflicts when it is determined that the Agency's need for a particular individual's services outweighs his or her potential financial conflict of interest. Under U.S. Code 712 of the Food, Drug and Cosmetic Act, Congress has authorized FDA to grant waivers to special Government employees and regular Government employees with potential financial conflicts when necessary to afford the Committee essential expertise.

Related to today's discussion, members and consultants of the Panel who are special Government employees have been screened for potential financial conflicts of interest of their own as well as those imputed to them, including those of their spouses or minor children and, for purposes of U.S. Code 18-208, their employers. These interests may include investments; consulting; expert witness testimony; contracts/grants/CRADAs; teaching/speaking/writing; patents, royalties; and primary employment.

For today's agenda, the Panel will discuss and make recommendations regarding the possible reclassification of molecular diagnostics for rapid detection of *Mycobacterium tuberculosis* complex and the detection of genetic mutations which confer antibiotic resistance to *Mycobacterium tuberculosis* complex.

The discussion will include the appropriate information and acceptable performance characteristics that will be required to assess the

safety and effectiveness of rapid diagnostic tests for *M. tuberculosis* complex and whether these can be sufficiently specified to support the reclassification.

Based on the agenda for today's meeting and all financial interests reported by the Panel members and consultants, no conflict of interest waivers have been issued in connection with U.S. Code 18-208 and 712 of the FD&C Act. A copy of this statement will be available for review at the registration table during this meeting and will be included as a part of the official transcript.

Dr. Michael Towns is serving as the Industry Rep, acting on behalf of all related industry and is employed by Becton Dickinson Diagnostic Systems.

The Agency would like to note for the record that Dr. William Burman, who is a guest speaker with us today, has acknowledged a professional relationship with the Tuberculosis Trials Consortium.

We would like to remind members and consultants that if the discussions involve any products or firms not already on the agenda for which an FDA participant has personal or imputed financial interest, the participants need to exclude themselves from such involvement and their exclusion will be noted for the record. FDA encourages all other participants to advise the Panel of any financial relationships that they may have with any firms at issue.

Ms. JoEllen DeLuca has been appointed to serve as a Temporary Non-Voting Patient Representative for the duration of the meeting on June 29th, 2011.

For the record, Ms. DeLuca serves as a consultant to the Gastrointestinal Drugs Advisory Committee in the Center for Drug Evaluation and Research. This special Government employee has undergone the customary conflict of interest review and has reviewed the materials to be considered at this meeting.

This appointment was authorized by Dr. Jill Hartzler Warner, Acting Associate Commissioner for Special Medical Programs, on June 28th, 2011.

Thank you. And all Panel members, please remember to turn your microphones on before you speak and turn them off when you're done. Thank you.

Now, before I turn the meeting back over to Dr. Caliendo, I'd like to make a few general announcements.

Transcripts of today's meeting will be available from Free State Court Reporting, Inc., at 1378 Cape St. Claire Road, Annapolis, Maryland, zip code 21409, telephone number (410) 974-0947. Information on purchasing videos of today's meeting can be found on the table outside of the meeting room.

I would like to remind everyone that members of the public

and the press are not permitted in the Panel area, which is beyond the speaker's podium.

The press contact for today's meeting is Amanda Sena. She's raising her hand.

I request that all reporters please wait to speak to FDA officials until after the panel meeting has concluded.

If you are presenting in the Open Public Hearing session today and have not previously provided an electronic copy of your slide presentation to the FDA, please arrange to do so with Ms. AnnMarie Williams at the registration desk.

Finally, please silence your cell phones and other electronic devices at this time. Thank you very much.

Dr. Caliendo.

DR. CALIENDO: Okay. So we'll now hear the reclassification discussion from the FDA. And I want to remind Panel members that after every presentation there will be time for questions.

So our first speaker is Marjorie Shulman, and she'll be discussing the FDA reclassification process.

MS. SHULMAN: Hi. Good morning. I'm just going to give a brief overview of the classification and reclassification procedures.

My name is Marjorie Shulman. I'm Acting Director of the Premarket Notification Staff with the FDA.

So, previously, devices were divided into two arenas, pre-amendment or post-amendment devices. All that means is when the device was introduced to market for the first time, and the date is May 28th, 1976 of the Medical Device Amendments. So previous to that, some devices were regulated as drugs.

So for a classification of a pre-amendment, they can be classified after the Food and Drug Administration has received a recommendation from the device classification panel, published the panel's recommendation for comment along with a proposed regulation classifying the device, and then published a final regulation classifying the device in the *Federal Register* and the C.F.R.

So reclassification of a pre-amendment device. FDA may reclassify a pre-amendment device in a proceeding that paralleled the initial classification, and that was done in the late '70s, early '80s, when they classified all the known devices; and it's in a proceeding that parallels the initial one, and it's based upon new information respecting a device either on FDA's own initiative or upon the petition of an interested person.

So post-amendment devices, once introduced into the market after May 28th, 1976, are automatically classified into Class III, and they remain in Class III and require a premarket approval unless and until FDA issues a substantial equivalent letter, which would classify it into either I or II, or the device is classified into I or II through the automatic Class III

Designation Program, also known as de novo.

A reclassification of a post-amendment device can be initiated either by industry or the Food and Drug Administration, and FDA, for good cause shown, can refer it to a panel for a recommendation. Then the panel would make the recommendation respecting the approval or denial of the petition.

So the goal here is a device should be placed in the lowest class whose level of control will provide reasonable assurance of safety and effectiveness. And there's three classes of devices: Class I, which are general controls; Class II, general and special controls; and Class III, premarket approval.

Class I mainly includes devices for which any combination of general controls are sufficient to provide reasonable assurance of safety and effectiveness.

So general controls include prohibition against adulterated or misbranded devices, good manufacturing practices, registration of the manufacturing facility, listing of the devices which are made there, record keeping, repair, replacement and refund, and banned devices.

Class II is for devices that cannot be classified into Class I because the general controls, by themselves, are insufficient to provide reasonable assurance of safety and effectiveness, but for which there is sufficient information to establish special controls to provide such assurance.

So special controls include, for example, performance standards, postmarket surveillance, patient registries, development and dissemination of guidelines, guidances, tracking requirements, and then recommendations and other appropriate actions.

Class III are for devices for which insufficient information exists to determine that the general controls, Class I, and the special controls, Class II, are sufficient to provide reasonable assurance of safety and effectiveness of the device and the devices are life sustaining and/or life supporting, substantial importance in preventing impairment of human health, or present a potential or unreasonable risk of illness or injury.

There's also a provision for restricted devices. So under the provisions of Section 520(e) of the Food, Drug and Cosmetic Act, the Food and Drug Administration is authorized by regulation to restrict the sale, distribution, or use of a device because of its potentiality for harmful effects or the collateral measures necessary to its use. FDA would determine there cannot otherwise be reasonable assurance of its safety and effectiveness.

So a restricted device can only be sold, distributed, or used either upon the written or oral authorization by a licensed practitioner or under such other conditions specified by the regulation.

Also, if the device is restricted for use by persons with specific training or experience in its use, or by persons for use in certain facilities, FDA must determine that such a restriction is required for the safe and effective

use of the device. For example, a device such as cardiac pacemakers and heart valves require a practitioner's authorization.

Hearing aids are restricted by a regulation which limits their sale to persons who obtain a medical evaluation of their hearing loss by a physician within six months prior to the sale of the hearing aid, and the labeling of the hearing aids must provide information on their use and maintenance.

So that is the end of that. Are there any questions?

(No response.)

MS. SHULMAN: Thank you.

DR. CALIENDO: Thank you, Marjorie.

Okay. So we're now going to move and hear a series of presentations from the FDA. We're going to hear three presentations, and then, at the end of the three presentations, we will have time for some questions.

And our first speaker will be Dr. Steve Gitterman.

DR. GITTERMAN: Got it, thank you. On trial and learning.

I'll be succinct. I have three simple goals for my presentation, and the first is very easy, which is an overview. The second is to give the Committee background of exactly what we're talking about. We're going to basically just go into some detail regarding what exactly is in a PMA submission and what exactly is in a 510(k) submission. So when we're talking

about reclassification, people could get a sense of exactly what's required in one option versus the other. And the last thing I'd like to go into, again not in great specifics regarding this particular issue, but what would be in a special controls guidance? Because as we'll see in a second, what is the focus of the meeting?

And the focus of the meeting is can Class III devices, such as NAAT assays for mycobacterial detection from direct specimens and/or indirect assays such as IGRAs assays be reclassified to Class II devices through the development of a special controls guidance? And, of course, we all have to be on the same page regarding what exactly is in a special controls guidance.

And the second piece would be similarly because they're unclassified at this point, can we develop -- can devices for the detection of antibacterial/antimycobacterial resistance mutations in direct specimens where *Mycobacterium tuberculosis* has been detected be initially classified as Class II?

And as I'm reading this it's just to belabor the obvious to make sure there's no, perhaps, misconception. By direct specimens we're talking about a specimen taken directly from the patient to the lab, not something that has been cultured or has gone through -- excuse me -- that has been identified as mycobacterial. This would be a patient specimen that is then directly analyzed.

What is the current approach? And I would say again, just to reemphasize something that has been brought up before, the intended use is what determines the classification, not the specific technology that's being used. It's not being classified because it's a different technology or it's a specific technology. It's based on the safety and effectiveness for the specific intended use.

And just reviewing again what Ms. Shulman had discussed a little bit before, we have Class I exempt devices, which include staining, which everybody knows, which is mycobacterial growth, and these are exempt; traditional culture media, L-J media, et cetera; automated systems and the associated media. Then we have Class I, which is not exempt, and this is the identification of the specific mycobacteria from cultured isolates.

Class II would be drug susceptibility testing from cultured isolates, not from direct specimens; again, the distinction made earlier.

Class III are NAAT assays from direct specimens and also indirect assays, such as IGRAs for the indirect detection, would mostly go, of course, for latent tuberculosis.

And, again, as mentioned earlier, unclassified devices, which are the detection of resistance mutations directly from specimens for which FDA has not received a submission to date, which is why it remains unclassified.

I will just say this will not be in -- well, actually this will be in

the test, but it will be multiple choice, and you'll only have to fill in a dot whether it's one, two, or three. So all of the above is wrong right off the bat.

But, again, I can't even see it, so I suspect no one else in the room can see it, which is probably good because somebody's going to raise their hand and say we've forgotten something. But this is basically what's been cleared by FDA previously. These are the Class I devices, and it's a fairly long list.

There's a tremendous amount of expertise around the table, and people are saying, I don't use that, I won't use that. It's just a list. These are the AFB stains and the devices that are cleared for identification. These are devices that are cleared for growth monitoring. Some of these are going to look very familiar to people at the table and in the audience.

And there's the Class II devices for antimycobacterial susceptibility. And, again, a number of these, I'm sure, are familiar to people.

Then there's the Class III devices, and this is getting more to the heart of the meeting. And, of course, we have the Gen-Probe device, the AMPLICOR device, the QuantiFERON and the T-SPOT. The latter two are indirect tests. And the latter two -- I'm sorry -- for the indirect detection of tuberculosis or immunologically based assays. And, of course, the first two, which are amplification-based assays.

Now, going to the second part that I wanted to raise, which is what exactly is in a 510(k) and in a PMA submission? I recognize this is a

review for a number of folks in the audience, but again, this is a baseline for everyone. And to really understand what people are going to be asked for input on, we'd hope we'd give everybody just that firm basis.

Okay, what is similar between the two applications? And this would be both in a PMA or premarket notification or a 510(k). It would be an intended use, what the test is to be used for. There's the device description, and that would be fairly detailed. What are the reagents and the test methodology used in a device? What are the controls for the device? What ancillary reagents? The test procedure; storage and shipping; the interpretation of test results and reporting; instrumentation and software. What are the performance data?

Of course, a nucleic acid assay, amplification assay, would have details regarding what the extraction procedure would be. What is the sensitivity? And again this is, of course, somewhat device or assay dependent, but what is the -- usually they'll almost certainly have the limit of detection and the reactivity, which is inclusivity. An example of that for people who know, like an influenza assay to detect all the strains of influenza. It's pretty obvious.

Analytical specificity and cross-reactivity; interference studies; the cutoff or equivocal zone of the assay has such a thing; precision within laboratory and between laboratories; the method for specimen collection, storage, shipping; carryover and cross-contamination studies; fresh versus

frozen specimen studies, which are as appropriate; and reagent/kit stability studies. And this would only be for a PMA, which actually perhaps the bullet doesn't even belong there, which we'll talk about in a minute. But this would not be required for submission in a 510(k).

Clinical performance data. And, of course, the key piece, which I believe is in one of the questions the Panel will be asked to discuss a little later, is what is the reference method? What is the standard the device was compared against for clinical truth? How do you determine somebody in fact has tuberculosis?

For mycobacterial detection, this could be, you know, of course culture with subsequent isolate isolation based on a composite of both phenotypic method, growth rate, and genotypic method.

And I put this in a question mark because of the second bullet because this may, I suspect, be a subject for Panel discussion, whether there needs to be a NAAT assay with bidirectional sequence as part of a reference standard. Again, I would be very surprised if this did not come up because there will be cases where, conceivably, a NAAT test under development is positive, but in fact the culture is negative. And, again, we don't need to go into detail with this now.

Susceptibility, in most cases, we'd expect an agar proportion method to be "the clinical truth" for an assay that would be looking at the direct detection of resistance mutations. And for indirect detection, again,

there would be comparison to confirm TB. That's the low-hanging fruit and that's pretty apparent.

But, again, I put this in question marks because this may be the subject of discussion. Should the TST be always part of "a clinical reference truth"? Or how would it be used in a new test assay that's being developed for indirect detection of tuberculosis? And, again, should there be an FDA-approved IGRA? Would a new assay, and perhaps as part of the reference method, include comparison against the T-SPOT or a QuantiFERON goal? Something to think about, which, I'm sure, people thought about before today.

In the PMA or a 510(k) application, there'd be discussion of the clinical performance data, what the protocol was, what's the inclusion/exclusion criteria, study procedures, the detailed statistical analysis plan, description of the blinding procedures, the copy of the IRB approvals, and informed consent.

Now, that would be required to be submitted as part of a PMA. Of course, you'd have to do it as a 510(k), but it would not be required for submission.

Again, in the clinical study protocol, this is, you know, what would be included, the specimen types, the clinical study sites. And this was actually a question that has come up previously that would likely be required, a minimum of three geographical sites, two of which would be expected to be

located in the United States. And the sites should be representative of where the device will ultimately be used. The device should be tested under the conditions of use.

As someone said at a previous meeting, you know, the company or the sponsor is very likely to be able to make the device work in their lab. They wouldn't be developing it otherwise. But the requirement would be for the clinical studies, that they reflect the conditions of use.

The study populations, of course, would be described, and again, it would be expected that the studies would be done in populations to be clinically important for the use of the device. Again, we just threw out some examples, but one would want to know these type of things, and I certainly think people would be -- certainly for age, HIV, disease prevalence, et cetera, would be of concern in the setting of a device for tuberculosis.

And for latent disease there would be subjects. I'm not sure what the footnote refers to. But for latent TB, subjects for major risk groups that would be indicated for screening because, of course, that would be where you'd want to be able to describe the performance in the package insert or the labeling.

Again, you know, a more statistical analysis plan would define what the target sensitivity and specificity was at the onset of the study. What were the goals of the sponsor? What was being anticipated? What were the confidence intervals? Of course, you want to know what the point estimates

were, but probably more important, as I suspect, will emerge would be the lower bound of the confidence interval. You'd want to define the sample sizes prior to starting the clinical validation studies. This is, of course, obvious to people on the Committee who have dealt with this previously.

The study results: the comparison of the device against the reference standard; what subgroup analyses; what was specified in advance; what is the sensitivity and specificity; the confidence intervals; the PPV and the NPV, as appropriate, which may not be appropriate for certain tests where there not is a good reference standard, but we don't need to talk about that right now; and again, for latent TB, the positive and negative percent agreements because there may not be a very good reference standard.

Of course, in the PMA or a 510(k) application, there'd have to be labeling. And labeling, interestingly enough, for in vitro diagnostic devices is defined by regulation, and that's essentially what is defined under 809.10(b). In the regulations we'd have the use of the name, the intended use, summary and explanation, basically a description of a lot of what I've just described that would be in the submission.

Of course, everything in the labeling has to be supported in the submission. It would not be in the labeling otherwise. And again specimen collection, many of the things I just referred to before, with bibliography and references, as appropriate.

Now, what's the difference? Everything I've mentioned before would be both in a 510(k), a premarket notification, and a PMA submission.

Manufacturing site inspection. This would occur for a PMA. It would not occur in the setting of a 510(k).

Clinical site inspections. Bioresearch Monitoring, or BIMO, would go out and do an inspection for a -- excuse me -- for a PMA, not for a 510(k), unless there was a for-cause reason to do so.

Postmarket oversight. There would not be the required submission of annual reports describing any minor changes made to labeling, manufacturing, review of published reports. That would not be required for a 510(k). That would be PMA specific.

Other issues. And I've also put an asterisk at the bottom to make a distinction that some of these are required for both a 510(k) and a PMA. But in the setting of a PMA -- I'm sorry. In the setting of a 510(k), it's not required to be included in the submission.

Same as the examples in the IRB approval. Of course, one has to get an IRB approval for clinical studies. But in the setting of a PMA, you would probably not have to submit that as part of the package.

So reagent/kit real-time stability studies, they would have to be submitted with the PMA. They would not have to be submitted, although they'd have to be done, with a 510(k); again, the IRB example which I've cited three times now; case report forms for all discontinued subjects; risk/benefit

analysis; a reduction in paperwork statement. That's a little bit ironic that you would have to put it in for the PMA but not a 510(k). But, you know, somebody had a sense of humor.

This is actually sort of a little of a digression, but there is a longer review time for the PMA versus the 510(k). There is an environmental impact assessment. That's on the test, not on tuberculosis. And a notification of transfer of ownership would have to be included in a PMA.

I'd also note for those who are really paying attention, I missed Slide 2. I was going to redo this one, but then I thought the confusion would just be too much. Everybody would be flipping their pages and say, They didn't reproduce Slide 2, Slide 4, or Slide 3.

What else is PMA specific? There has to be a specific summary of safety and effectiveness data that's different from a 510(k) summary in the PMA as opposed to the 510(k).

The conclusions drawn from the summary. These are just regulatorily defined aspects of a PMA that have to be included in the submission that "would not be identified as such in a 510(k)"; submission of manufacturing information, protocols, and controls; alternative practices and procedures; the marketing history, if it had been marketed; final printed labeling; and postmarketing. They would have to submit annual reports and device change reporting requirements.

Now, as I mentioned earlier -- and this is really much to the

heart of the meeting -- the overarching sort of aspect of the meeting is, can special controls be written that would mitigate the risks of having this as a 510(k) application rather than a PMA, or that would support and be able to ensure the safety and effectiveness of the device if, in fact, it undergoes the 510(k) process rather than the PMA process? And that, in fact, would be through the FDA drafting and publishing of a special controls guidance.

So what would be in the special controls guidance? And I know these questions have come up. There would be an introduction. I'm sure, in fact, many on the Committee are familiar with these documents. Certainly many in the audience are. An introduction. There would be a background. I'll also go into some of these in a little more detail. There would be a regulatory scope. There would be a description of what the risks to health are of the device and how the special controls guidance does mitigate those.

There would be a description of the device. There would be performance studies, what would be asked for somebody developing that type of device, and it would include both the analytical studies and, of course, the clinical studies. It would include what we would recommend in the labeling, and it would include what postmarketing measures we would recommend to ensure that the device is marketed safely and effectively.

Now, a special controls guidance may not necessarily include all of these things. It could include all of these things if you actually include a few additional sections as necessary. But this is likely what a special controls

guidance conceivably would look like in this circumstance.

What is the regulatory scope? The special controls guidance would define what the scope of the device is that is being developed. Now, of course, that's important if it's under the 510(k) process. It would be using the special controls guidance. The guidance has to identify what the universe of devices are that the guidance applies to.

And then again, Marjorie had a great prop there. She actually had the C.F.R. But if anyone has needed a sleeping aid and would like to read it, every device that has undergone the 510(k) process, with a few exceptions, will have a section specifically which will have 21 C.F.R. 866.blank, which will define the device, identify what the device is and, with additional information, what the class of the device is. And it really is amazing that -- I'm sure everybody's now looked at it -- every device we regulate is listed in some shape or form in that section of the C.F.R.

So we would define the regulatory scope, and this is just an example. I can't emphasize more, but I'm just talking about just examples there, really for illustrative purposes only, and I don't want anyone to think that we've made these decisions or that in any way, shape, or form we're trying to promote either reclassification or not reclassifying. This is really just an illustrative example and nothing more.

In the special controls guidance there would be risks to health. And, of course, what are the risks to health? A false positive result, a false

negative result, and in the specific case of tuberculosis, there would be a description of bio-risks and the risks to laboratory workers from handling of the samples. And, of course, I should mention, later on in the guidance, how these risks are mitigated.

There'd be a description of the device; again, what was submitted in the application; what we would recommend how the device be described; what the intended use of the device would be; what the reagents of test methodology would be recommended; what controls should be used for the device.

So we would be specifying, if you were to develop a diagnostic device for this use -- and of course in this case we're talking about, let's say, the first question, which would be NAAT assays -- we would identify what the controls would be that we would expect a manufacturer or a sponsor to use to identify to develop a test for this purpose that was safe and effective.

We would discuss, we probably wouldn't specify, of course, but we would discuss the need and the way a sponsor should identify ancillary reagents and the way they should be identified in the labeling. Of course, tests may differ, so we can't say to use a specific buffer or not.

We would identify how they should describe the test procedure, how they should describe specimen storing and shipping conditions, and how they would identify, perhaps, what the tests would be to identify those. And, again, we would describe how we would recommend

tests be interpreted or reported.

We'd have general study recommendations. This, again, would probably be a major -- an important section of the special controls guidance, what performance studies we would recommend; what are our general recommendations -- and this would be a section; what we would recommend regarding doing nucleic acid extraction studies -- this is in the case of NAAT assays; what we would recommend for how and what we would expect for analytical sensitivity studies; what we would recommend for analytical specificity studies; what are the obvious things we would want to look at for interference studies or cross-reactivities, cross-reactivity studies.

You know, again, we wouldn't have to be rocket scientists to see what are the things we would likely recommend in a such a guidance mod, et cetera, et cetera. But, again, we would recommend these in the special controls guidance, with the hope, if in fact that were a path that were pursued, that we could identify these with enough specificity so that, again, a device could be developed that we considered safe and effective.

We would describe how a sponsor should establish cutoff and equivocal zones, if appropriate. We would describe the nature of the precision studies, how many laboratories should be studied, samples, et cetera, in some degree for this specific use; what would have to be included and how it'd be included for specimen collection; carryover and cross-contamination studies, what we would expect for these type of devices; and

fresh versus frozen samples, if appropriate, what type of studies should be identified.

Again, I'm speaking as a generalization, but I think it's probably obvious to members of the Panel and members of the audience how these would apply to the different questions we're asking because, in essence, as people recognize, we'd likely be talking about three special control guidances in this case.

We would describe the clinical studies. A very, very important question, which would be, what would the reference methods be? That's obviously an equal question for a PMA and/or a 510(k). But, again, if we would give guidance in a special controls guidance, obviously, about what the recommended reference methods would be in the different scenarios we're talking about.

We would describe what are the contents of a protocol. Again, we would not be providing a 50-page protocol, but we would have a general guidance of what we recommend, what are the specimen types that should be included? What would the nature of the study sites be? As we mentioned earlier, clinical sites where, you know, that would be reflective of the practice of use. What the study populations would be in a case like this. And, again, this is a subject for discussion, but I'll leave it for that, then.

We would like a statistical analysis plan. That should be included when the protocol is submitted. We'd like certain ways, obviously,

that we would prefer presentation of results in a case for these type of studies.

In the special controls guidance we would have recommendations about labeling, which I put up earlier, what we think -- what recommendations we would include in sections of labeling, how we think certain things should be. Again, we're not going to be writing labels for people that couldn't do that. But there would be general guidance of sections of labeling we think would need to be emphasized in this. In a case like, you know, in a similar case, similar to some of the assays we're talking about now, labeling, of course, we would put in special controls; the caveats and perhaps the limitations of the nature of the use of these assays for tuberculosis; in cases where there's recommendation that a culture always be done because of the nature of the properties of the test. These type of recommendations would be laid out in the special controls guidance.

So, again, a sponsor developing a test for this use could come to FDA and basically come to us with an application that we could deem safe and effective, if they followed most of the recommendations.

That's really all I wanted to talk about. I'm hoping I could give a flavor of really what were the input we're seeking and basically framing the discussion.

I do have to say, a number of these people in the room, there's been an enormous number of internal discussions and a lot of very spirited

debate from a large number of people, and I'd just like to thank everybody listed on this slide for their input.

And I'll leave it to Dr. Caliendo. I assume there'll be questions later. Okay, thank you very much.

DR. CALIENDO: Thank you, Dr. Gitterman.

Okay. So our next speaker is Dr. William Burman. He's the Interim Director of Denver Public Health and he's going to be discussing current tuberculosis diagnostics, the role of diagnostic tests for latent and active tuberculosis.

DR. BURMAN: Well, thank you very much. I think I'm invited here because I'm not an expert on any of this and therefore I'll give an un-expert clinician's perspective.

Just a couple of disclosures. I worked with Tibotec a little bit on their drug TMC207, which just got a name, and both on their data safety monitoring board and as a scientific advisor. They pay my employer for my time; they don't pay me. And I've been a participant in a study of interferon gamma release assays. So I've had them done three or four times on me.

So my objectives today are really to provide the clinician's perspective. In some ways, I'm going to do a critical review of our current tests because then that says where we need new tests and how those might be used. So we'll really be looking for opportunities for better tests from what we're using right now.

And I'm going to look at two key diagnostic decision points: Does this person have tuberculosis? And if they have tuberculosis, is this problematic drug-resistant tuberculosis?

Despite what was listed, I'm not going to talk about the diagnosis of latent tuberculosis because that would require a whole other 30 or 40 minutes.

I'm going to do that through a series of cases because I'm a clinician and I think that helps illustrate the needs for new diagnostic tests. So I'll start with a few that revolve around the question of, Is this tuberculosis, and therefore, do we need to start taking actions relatively quickly related to that concern?

So this is a 33-year-old man who was admitted to our ICU with a five-day history of nonproductive cough, fevers, shortness of breath, profound fatigue. His past medical history was notable for advanced AIDS. He had recently been started on antiretroviral therapy eight weeks ago. He had a history of some things that would make you worry about tuberculosis. He had been an injection drug user and jailed twice, had periods of homelessness, and was from Iowa.

His medications were antiretroviral therapy and trimethoprim/sulfamethoxazole. He looked quite ill when he was admitted, was somewhat hypotensive and tachycardic, required large amounts of fluids to support his blood pressure initially, had temporal wasting and looked

profoundly fatigued. He had the new onset of severe pancytopenia, as you can see, and a mildly increased alkaline phosphatase. Chest X-ray showed miliary infiltrates. And about that time I started to get scared. And his sputum and BAL were AFB smear negative.

So the questions that come up in this kind of case, this is obviously something quite significant and moving very quickly. And so, what's the etiology of this severe illness? Is it tuberculosis? It is a non-tuberculous mycobacterial infection? Is it a fungal infection? Is it the unusual presentation of an AIDS-related malignancy? Is it immune reconstitution related to one of that list above?

And what should you do now? Should you start empiric TB therapy, start empiric antifungal therapy, get biopsies? My answer to all of those questions was yes.

The second case is quite different and it illustrates what comes into tuberculosis clinics. So this is a man who came into our clinic recently. He's an immigrant from Bhutan. He'd had cough for four weeks, small amounts of phlegm, no hemoptysis, some subjective fevers, no weight loss. None of his other household members were ill. His past medical history was notable for a positive tuberculin skin test, had not received treatment for that, and he was a smoker and still smoking.

His chest X-ray showed these ill-defined left upper lobe infiltrates, sputums were smear negative, the cultures are pending, his rapid

HIV test was negative, and he had no prior chest X-rays to compare to.

And so is this tuberculosis or is it atypical bacterial pneumonia, fungal pneumonia, or the presenting symptoms of lung cancer? And what should we do now, treat for TB, treat for a bacterial infection, consult a pulmonary specialist? Those are the clinical questions that come up.

And a third case, a 72-year-old woman with cough for four months, small amounts of phlegm, some hemoptysis, a weight loss that was quite significant for her. She hadn't responded to an empiric course of amoxicillin-clavulanate. Her past medical history was notable in that she did not have a known exposure to tuberculosis, although she was 72. She had a negative tuberculin skin test while a teacher, and she was currently working as a volunteer in a kindergarten, just to make the people in a TB control program sweat, and she lived in Colorado all her life and had hypertension.

This is her chest X-ray and chest CT, which shows this large, thick wall -- actually several thick walled cavities, and her smear was positive and the cultures are pending.

And therefore there's another series of questions. Of course, is this tuberculosis? Should you start TB therapy? But importantly, should you start a contact investigation? What should the message be to the school, to the parents, to the media, since this kind of case would go quickly to the media? All cases in which, I think, stress our need for good diagnostic tests for tuberculosis.

I put this slide in just to give you a sense of what TB is like in a low-burden setting like the United States. This is actually a recent study from Switzerland, but I think its results are typical for many places in the United States, including Denver, where I practice. It just gives you a sense of how does tuberculosis present in a low-incident setting?

Many patients have pulmonary tuberculosis. Of those, about half have smear positive pulmonary tuberculosis. A substantial percentage of patients have extrapulmonary tuberculosis. So just over half. And so, clearly, some have both. And the common manifestations of extrapulmonary tuberculosis are there: adenitis, pleuritis, and then a host of places because tuberculosis can show up anyplace.

Importantly, as is true in most places in the United States now, the majority of patients are foreign born and an appreciable percentage are HIV positive. So, again, a flavor of what -- of how tuberculosis presents in a low-incidence setting like the United States.

So the currently available tests for active tuberculosis are some relatively rapid tests. So I arbitrarily define relatively rapid as having the results available within a few days, so that certainly includes AFB smear, nucleic acid amplification tests, the tuberculin skin test, and IGRAs, which I'm not going to talk about further -- of course, those are not in any way definitive for the diagnosis of active tuberculosis, although they're used as adjuncts in cases in which the other tests aren't as definitive; mycobacterial

culture; and then things like histopathology. And I won't again discuss histopathology, but I'll say quite a bit about culture.

So how does smear perform? Well, these are data from a relatively recent study from the United Kingdom, and again, I think they're representative of how smear performs in the United States. And so this is a study of 282 culture positive specimens in a hospital in the United Kingdom. A substantial percentage of those did have tuberculosis, but then a fair number of people had other infections: *Mycobacterium avium* complex, *kansasii*, et cetera.

And so, overall, the sensitivity of smear for culture positive tuberculosis was about 50 percent, a fairly familiar figure, and the specificity of smear for tuberculosis was 77 percent, that is, not infrequently non-tuberculous mycobacterial infections presented with smear positivity.

There are certainly well-recognized factors that are associated with smear negative tuberculosis. Age -- children, particularly very young children, have a much higher risk of smear negative tuberculosis. Persons with HIV infection, and really probably that just is a marker for severe immunosuppression; other very immunosuppressed patients, such as transplant patients, behave very similarly in terms of clinical and laboratory presentation. Extrapulmonary tuberculosis is frequently -- in some cases it's the rule that it's smear negative. And patients with disseminated disease.

And the point about that list is that it makes the clinician

nervous because smear performs worst when you need it the most; that is, these are patients, that list is the list of patients who can progress rapidly to be hurt or even killed by tuberculosis. So like many tests, smear works worst when you need it the most.

This just shows that data for children. And, again, I'd point out, this is a study from South Africa, a center where it probably has the best TB diagnostics for children in the world, and so this is under ideal circumstances, in the best center in the world, in Cape Town. How does smear perform? Well, of cases that they thought clinically had tuberculosis, a smear was positive in only 12 percent. So really a substantial limitation of the test.

Can smears be falsely positive? Yes, the answer's clear. And this is a relatively recent and very clear demonstration of that. This was done in South Africa, and it was an evaluation of their laboratory's performance, and the way they did it was submit a 190 simulated sputum specimens into their standard laboratory workflow. And so they made these look like sputum, and they gave them false names so they looked, to the laboratory personnel, like just another submission.

Interestingly enough, three percent of those were read as positive, they were all culture negative, and something that I'll come back to later, two of those, or one percent, were reported as culture positive.

So, again, these are autoclave specimens. These could not be culture positive, in a sense, and they should not have been smear positive. So

that says something about the performance. And this is in a high-volume TB lab where they probably have a great deal of expertise in TB diagnostics, which is not the case in many laboratories in the United States.

So summarizing the sputum smear, it has poor sensitivity, and that's long been recognized. Smear negativity is common in pulmonary tuberculosis. It is the rule in many forms of extrapulmonary tuberculosis.

The specificity of smear is relatively high for mycobacterial disease, but the specificity for tuberculosis is relatively low, particularly in settings like the United States in which tuberculosis is uncommon.

And I think false positive results are not often recognized, but they are not rare, and I didn't go into other studies, which give similar numbers for false positive smears.

So what kind of clinical problems do those limitations cause? Well, the poor sensitivity can lead to delayed diagnosis. And, again, that's in patients who could have forms of tuberculosis which progress rapidly: young children, people with HIV infection, et cetera. And so that lack of sensitivity could result in clinical progression or even death, although I think most clinicians are very familiar with the poor sensitivity of sputum and don't make too many decisions based on a negative.

The delayed diagnosis of smear negative pulmonary disease in HIV negative adults is much less of a problem. I actually think, in a TB control program, the problem is that patients may be lost while waiting for their

culture results. They have forms of disease that progress very slowly, and so the likelihood of clinical progression is very low. And those patients also have a low, but not absent, risk of transmitting to others during that time period.

The suboptimal specificity for tuberculosis can be problematic in that it can lead to unnecessary treatment for TB, unnecessary respiratory isolation or hospitalization, and unnecessary contact investigations. So really pretty significant problems from the limitations of sputum smear.

I want to turn to, then, the other relatively rapid test, and that's nucleic acid amplification, and I chose this of the many studies of nucleic acid amplification available. I chose this one because of its size and because it represents the way that NAA has been used clinically. This is not a research study. So this is a review of the use of nucleic acid amplification tests in New York City over a substantial period of time.

I also chose it because of the number of patients involved. So this is more than 2,000 patients. So a relatively large study and one which illustrates the use of these tests in routine clinical practice, not in a research setting.

And so you can see that the sensitivity and specificity of these tests is high. It decreases, of course, where you need them the most, that is, patients who have smear negative disease. Interestingly enough, the specificity also drops off in patients with smear negative disease, and that can lead to some challenges.

This is a somewhat complicated slide from that same review, but let me just walk you through a couple of the highlights. This on the vertical axis is the percent of patients with a positive culture from *Mycobacterium tuberculosis*. So that's the vertical axis. Along the horizontal axis are various combinations of smear result and NAA result.

And the two I want to point you to are circled there. One is the patient who had a low-grade smear, so 1+, perhaps, and was NAA negative. A third of those patients had culture positive tuberculosis. That's a limitation.

And the other one I want to point you to is those who had high-grade sputum smear positivity, so perhaps 3+. And 10 percent of those patients had *Mycobacterium tuberculosis* isolated from culture, again, a limitation.

NAA has been used in extrapulmonary sites for the obvious reasons, that that's the particular weakness of smear. And so this is just one of many studies of that potential use. And so it compares 42 patients who were eventually judged to have TB meningitis, some of whom did not have positive cultures. You can see the yield of smear, so only 10 percent. And actually that's relatively high for meningitis.

Of the 26 patients who had nucleic acid amplification tests done, they were positive in 58 percent, and again, 74 percent of patients had a positive culture.

Of course, the problem of culture is shown below. That's the

time to positivity. So that was a median of 20 days, with a range out to 46 days. And the clinicians in the audience will appreciate that 20 days is too long to wait for a diagnosis of meningitis. Patients could be severely hurt by waiting for that result.

I want to turn to the potential use of PCR- or NAA-based tests for children. And this is a very carefully done study from Lima, Peru, published last year.

And so clinical truth, I love that phrase that Steve used. So clinical truth in this study was a scoring system, a standard scoring system that included risk for exposure, tuberculin skin test, radiographic results. So it was a composite clinical and laboratory score. And they then put persons or children, young children, into a high-risk category, a moderate-risk category, and then controls were children who had zero on this composite score.

And so on the left is the yield of culture, given those definitions, albeit not wonderful for clinical truth. So 14 percent of children in the high-risk category had culture positive tuberculosis, 1.6 percent of those at moderate risk, and none of the controls. And so shown on the right are the yield of PCR, and so it looks much higher in the two risk groups and is, at 26 and 23 percent, although bothersome is that two percent of the controls had PCR positivity. And they actually followed these controls for an additional year and none of them developed clinical tuberculosis. And so one

thinks that these really are false positives, though that's a difficult -- clinical truth in this area is really difficult to get at.

How do those results from NAA tests affect clinical care? There's limited literature on how NAA tests might affect clinical care. This is one nice study from the Baltimore TB lab, and it shows overall results which would be typical for many TB labs in urban America. So they had 638 suspects, a fair proportion of whom had tuberculosis, but then a lot of whom had other things, non-tuberculous mycobacterial infection.

The sensitivity of the NAA for culture positive tuberculosis was 92 percent. The specificity was nearly 100 percent, so it really performed quite well. And the point of the study was to compare -- had the availability of this test changed management. And so they provided the NAA result along with the initial smear. So at the same time, the clinician was given the smear result and the NAA result. And there was no apparent effect on the initial management of smear positive patients; that is, clinicians usually chose to treat those patients as tuberculosis.

The one effect was that during the NAA era, if you will, there was a decreased length of unnecessary therapy, from a median of 6 days versus 31 days in the pre-NAA era. So a moderate effect of the availability of these tests in the management of smear positive patients.

This is a small study, but the only one that I am aware of, that looked at how NAA testing might affect the management of extrapulmonary

tuberculosis. It comes from a single hospital in New York City. They had 32 patients with suspected extrapulmonary tuberculosis who underwent NAA testing. Fourteen of those were eventually diagnosed with tuberculosis, not all of whom were culture positive. The sensitivity of NAA was 86 percent. The specificity was 100 percent.

Although, when they looked back, they thought the NAA was very -- the result was infrequently used by clinicians. In one case a positive led to starting therapy and in one case a negative led to stopping therapy, and their point was that, at least as presently used, the effect of NAA on clinicians' management of suspected extrapulmonary tuberculosis was very modest.

So my view of NAA testing is, for pulmonary tuberculosis, the greater specificity for smear positive specimens is helpful and it's useful. And so Case 3, which was the older woman with terrible cavities and a smear positive specimen, I think it's useful to have NAA available because a negative is strongly suggestive, not definitive, that this is something besides tuberculosis.

The greater sensitivity for smear negative specimens, I think the clinical value of that is moderate because there is an appreciable culture positivity still in a negative test, and I think we don't understand the value of these tests in children. The major limitation of these tests is clearly their lack of the ability to assess for drug resistance, and so they always have to be used

with culture, not in place of culture.

For extrapulmonary tuberculosis, which is the greatest clinical need, that's where smear positivity becomes the exception, we don't really have good data. I think a positive result strongly suggests tuberculosis, but a negative result is insufficient as the basis for stopping therapy. I'll say, in some of those forms of tuberculosis, a negative culture is insufficient to be the basis, the sole basis for stopping therapy, and we clearly need additional data on how NAA performs in extrapulmonary tuberculosis.

Let me turn then to culture, which is always viewed as the gold standard, and I'll try to make you skeptics. To summarize a tremendous amount of literature from over 100 years, for pulmonary tuberculosis, culture has very high sensitivity. Its sensitivity appears to be unaffected in patients with HIV, whereas smear is less sensitive in HIV positive patients. I still don't think we know what to do with any diagnostic test, quite frankly, in very young children, and that includes culture.

Mycobacterial culture has high specificity. It can clearly allow you to distinguish TB from non-tuberculous mycobacterial infections. And, of course, a positive culture is the usual first step in testing for drug resistance. And so, hence, it's thought to be the gold standard.

But it has real limitations. The first of those, of course, is the time to positive results, and that can have a major -- that can cause real problems in decision making, both for the individual patient -- Do I treat? Do

I treat right now for tuberculosis? Do I isolate this person? Do I hospitalize them? Do I keep them in the hospital? And then it also has ramifications for TB control. Do I start a contact investigation? In the case of the third case, do I notify the kindergarten? Do I notify the press that there may have been a major outbreak of tuberculosis in a kindergarten, for example?

So real limitations. The limitations are important when there's rapidly progressive forms of tuberculosis, again, in patients with HIV, children, disseminated forms of tuberculosis, and then in drug-resistant tuberculosis, where even with rapid broth culture methods, the time to positivity can still be problematic in the management of the individual patient.

So this is a graph that I expect is familiar to almost everyone in this room. It's just the time to positivity in a nicely done study in Lima, Peru, a large number of patients evaluated, and it compares three different forms of culture and the time to positivity. MODS culture is not commonly used in the United States, but automatic mycobacterial culture, this was a BACTEC culture, and it shows that it's much faster than L-J culture, the classic form of solid media for a culture in tuberculosis. And the broth culture speeds up the time to growth detection, but it still is a median of almost two weeks, and that two weeks can be problematic in some situations.

I want to turn to what I think is the dirty secret of mycobacterial culture and that is it has problematic specificity -- or

sensitivity. And one of those problems is one we cause as clinicians, that is, we use drugs that are potent for tuberculosis, for infections that are not thought to be tuberculosis, and primary among those are the fluoroquinolones.

And this is a nice study done by Tim Sterling and colleagues in Tennessee, and it just shows that patients with tuberculosis now commonly get a course of fluoroquinolones mistakenly because of the thought that they have a bacterial infection. And by 2004, in this study, 40 percent of the patients diagnosed with tuberculosis in Tennessee had received a course of fluoroquinolones within 12 months, mostly very close to the diagnosis of tuberculosis.

Does that have effects? The analysis they did was to look at whether that predicted the eventual diagnosis of culture negative tuberculosis, and the answer was that it didn't, although they noted that there was a substantial delay incurred because patients get better when they're treated with a fluoroquinolone, and that led to a median delay of 80 days in the diagnosis of tuberculosis.

So I don't know if this is so much a problem of sensitivity, as a problem of the way the clinicians are working, not the laboratorians.

The major problem for tuberculosis -- for culture in the diagnosis of tuberculosis is extrapulmonary tuberculosis. And I chose this study because it's a very carefully done prospective study of tuberculosis

diagnosis in a key clinical setting, that is, very severe tuberculosis, TB pericarditis, a fatal illness if not treated quickly and correctly.

And so this is a prospective study of 233 patients in South Africa, with large pericardial effusions. They did a standard evaluation, so all patients underwent pericardiocentesis. All of them had mycobacterial smear and culture and used a contemporary broth culture system. They had other standard tests done on the fluid, including cytology. And those who weren't diagnosed within a week, using that algorithm, underwent pericardial biopsy. All of them were tested for HIV, and a third of them were positive.

So the final diagnosis, not surprisingly given that this study was done in South Africa, was that 70 percent of the patients were judged to have tuberculosis, and then the usual list of things below that. So, some with malignancy, some with connective tissue disease, a few with bacterial pericarditis, and a host of other rare causes of large pericardial effusions.

The key slide is this. So in their judgment, culture had a sensitivity of 52 percent when done on pericardiocentesis fluid. Its specificity 100 percent. But I think that's an important result. It just illustrates very important extrapulmonary forms of tuberculosis. Culture cannot be considered as definitive. A negative culture can't be considered as definitive.

And to give you an idea of who are these patients who are being diagnosed with culture negative TB pericarditis, they describe them very carefully. So some of those had positive smears, a few. Most had

positive histopathology from a biopsy. Others had positive cultures from another site. And so they were very careful about this diagnosis of culture negative TB pericarditis.

So the clinical settings in which mycobacterial culture lacks sufficient sensitivity to "rule out" tuberculosis are extrapulmonary tuberculosis, so that's particularly serositis, that is, from fluid, pericardial fluid, pleural fluid, ascitic fluid, and invasive biopsies are often thought necessary to diagnosis those forms of tuberculosis, certainly in meningitis, and then suspected pulmonary tuberculosis in young children and pretreated patients.

I want to turn then to specificity. I think we often consider culture to have 100 percent specificity. How can a culture be wrong when it's positive? The answer is it can easily be wrong. So this is a review we did some years ago of studies that carefully evaluated for cross-contamination, and we chose studies that included at least 100 patients with a positive culture for tuberculosis. We published that back in 2000, and I just recently updated that literature review to this month.

And so we found studies that addressed over 13,000 patients who had a positive culture for TB, and overall, three percent of those cultures were falsely positive as judged by DNA fingerprinting of isolates and clinical cores.

And another way of looking at that is the median percentage

per study. That was almost exactly the same, 2.9 percent. And the interquartile range of false positivity was 2 to 4 percent, with a range from 0 to 33 percent.

Importantly, clinicians acted upon those results. So the majority of patients who had a false positive culture were treated for tuberculosis, 55 percent in the reports in which that information was provided.

So to summarize the suboptimal specificity of culture, false positives have been identified in almost all studies that evaluated a reasonable number of patients. And we again defined that as greater than 100, so in 16 of the 17 studies that met those criteria. The common mechanism is cross-contamination, probably during initial batch processing, although other mechanisms can certainly occur.

The factor in the two studies that have looked for factors associated with higher rates of cross-contamination, the key factor is having specimens processed in a low-volume lab. And that's concerning in the United States because many labs in the United States would be considered low-volume labs.

And these false positives have definite clinical relevance. Most patients are treated for tuberculosis, and there are abundant reports of unnecessary hospitalizations, unnecessary procedures, unnecessary contact investigations, et cetera, all on the basis of false positive cultures.

So my summary of culture is that it's an excellent test, but it has more limitations than I think are commonly thought. The time to a positive test can be problematic. It has suboptimal sensitivity in some very important situations: extrapulmonary tuberculosis, tuberculosis in young children.

And what's less often recognized is the suboptimal specificity of culture. The frequency of false positive cultures is not widely appreciated by clinicians, despite more than a decade of studies on this topic.

And then, finally, the cost and the complexity of the training of laboratory personnel, et cetera, that are necessary to produce high-quality culture can't be overestimated.

Just to quickly give you a follow-up on the cases. Case 1, the severe illness in a man with AIDS. That turned out to be a immune reconstitution syndrome related to disseminated *Mycobacterium avium* from a liver biopsy. The second case, the cough in the refugee, that was smear negative, culture positive tuberculosis. And the third case, the elderly woman with cough, that was *M. avium* pulmonary disease.

Let me turn then to the second of those questions. Is this drug-resistant tuberculosis? And, in particular, is this problematic drug-resistant tuberculosis? And, again, I'll start with a couple of cases.

The first is a young woman who came into our TB clinic about a year and a half ago. She had had cough for quite a time. Initially she went to

an urgent care clinic. She was diagnosed with bronchitis, given an antibiotic, had no benefit. She went back, got another diagnosis of bronchitis, had another course of antibiotics, which didn't help. She began a nursing program and was, interestingly enough, TST negative, didn't have a symptom screen, and after six months of symptoms, she self-referred herself to National Jewish Hospital. She had a worrisome travel history. She had just come back from living in South Africa for six months, including living in KwaZulu-Natal.

This is her chest X-ray and it shows a large, thick walled off cavity in the right upper lobe. And she had 3+ smear positive specimens.

So I think all of us would say this almost has to be tuberculosis. So the question is, is this really drug-resistant tuberculosis? Should we start an expanded regimen for possible drug-resistant tuberculosis? And if so, how much should we expand? She comes from a part of the world where there have been terrifying outbreaks of XDR tuberculosis.

So it really is a challenging clinical question, and we would like not to use third-line TB drugs in this situation. They have substantial risks of side effects while waiting for definitive results.

And importantly, in her, she had just started a nursing program. How should we manage her close contacts? Should we treat those who are TST or IGRA positive with standard therapy for latent tuberculosis? Or should we wait until we understand the degree of resistance that this isolate might

have? And, again, importantly, what's the message for the public, what's the message for the media? To say that you have an XDR outbreak in a healthcare setting in the United States would certainly bring you attention from the media.

A second case that is quite different but it illustrates some of the clinical problems, a refugee who came in with seven weeks of diarrhea, fevers, abdominal pain, very substantial weight loss, severe fatigue. He looked very ill on presentation. Generally, although his vital signs were not particularly worrisome, he was found to have a host of laboratory abnormalities, and the most important of those was that he was HIV positive, with very advanced AIDS, a CD4 count of 6.

His initial evaluation showed a chest X-ray abnormality. A CT of the abdomen had a host of findings. The important ones were retroperitoneal adenopathy and a finding called omental caking. He had colonoscopy because of his history of diarrhea, and that showed CMV on a biopsy. His sputums were smear negative. A paracentesis fluid specimen was smear negative. He was treated with ganciclovir, got a little bit better, and then his broth AFB culture was positive at six days.

And so is this drug-resistant tuberculosis? This is a severely ill man in whom days will make the difference for whether he has clinical progression or survives. And so, if he has drug-resistant TB, what's the breadth of resistance? Should we start him on standard TB drugs? Should we

start him on expanded therapy for TB? Those are the clinical questions.

I won't go into a long and detailed evaluation of phenotypic resistance testing, and I'm not an expert in it. I'll give the clinician's perspective. Obviously the second case -- both cases really illustrate the problems of the time to having results. In both of these cases you would really like to be able to say is this drug-resistant tuberculosis very quickly, to manage the patients and their contacts appropriately.

My sense of phenotypic resistance testing, again, as a clinician who is not a laboratorian, is that it's difficult to obtain highly reproducible results for drugs other than INH and rifampin. And that leads to uncertainties about the interpretation of those tests for drugs other than INH and rifampin because the relationships between in vitro activity and clinical activity are complex, they're class specific, that is, they're very different for pyrazinamide than they are for a drug like ethambutol.

And it's clear that simple pharmacodynamic models, which are sometimes used to define cut points, don't work very well with tuberculosis. So if you look at a simple model like peak over MIC or AUC over MIC, those models perform really quite poorly for tuberculosis.

This slide just illustrates the problem of time to having results. This is a study from Tugela Ferry, South Africa, where there's been a broad community-wide outbreak of highly drug-resistant tuberculosis. And these are just the survival curves of patients with MDR and XDR-TB, showing -- I put

the line there at about one month, since that's the average time to both getting a positive culture and having drug-resistance testing available, and half of these patients have died by the time one would get conventional results from culture and susceptibility, clearly illustrating that in patients with -- most of these patients had advanced HIV disease. Tuberculosis, and particularly drug-resistant tuberculosis, can be a rapidly progressive and fatal illness.

MDR-TB is what keeps all of us in TB control awake at night. This is the WHO map showing the prevalence of primary MDR-TB, that is, patients who have not previously been treated. And so you can see, in the countries of the former Soviet Union, more than six percent of the patients who are presenting for the first time with tuberculosis have MDR-TB. And there are significant hot spots elsewhere in the world, and the developing one in South Africa and neighboring countries.

Not only that, it appears that in those settings, MDR-TB is becoming more common. So these are data from Tomsk, Russia, showing a substantial increase in the prevalence of primary MDR-TB. So in the last year addressed in this study, more than 10 percent of patients presenting, having never been treated before, had multidrug-resistant tuberculosis.

The prevalence of MDR-TB is increasing rapidly in Africa, and that's particularly worrisome because of the HIV epidemic that just fuels any TB epidemic, in particular, an epidemic of drug-resistant TB.

And so this study just shows orange is bad. Orange shows settings in which more than two percent of all patients, that includes retreatment patients, have MDR-TB. And so you can see, over a very short time period, from 2004 to the present time, how rapidly drug resistance appears to be developing in sub-Saharan Africa.

So my view is that this is the key challenge in global TB control. This global increase in highly resistant TB creates tremendous challenges. Certainly most of those challenges are in these particular settings, but they also create challenges for domestic TB control programs, as I've tried to illustrate with these cases.

And I think there's a critical need for rapid resistance testing. We need it to ensure appropriate therapy for the individual, and we need it to give appropriate TB control activities to the community: contact investigation, treatment of close contacts, decisions about hospitalization, isolation, et cetera.

We have tests coming along -- and again, I'm not an expert on these tests -- that certainly have promise. And so this is a study of one of those assays done in a laboratory in South Africa, where there's an appreciable risk of multidrug-resistant tuberculosis.

And so you can see that the sensitivity and specificity of this assay for rifampin resistance is quite high. The sensitivity for INH resistance is low. And that reflects the complexity of the genotypic complexity of INH

resistance. Rifampin resistance all resides within a single gene. INH resistance resides in at least two, and maybe more than two, genes within TB.

So the pros and cons of rapid genotypic resistance testing. The obvious pro is that you can get the results back quickly, so in less than 24 hours. The studies so far suggest high sensitivity and specificity for rifampin resistance. And I really think that genotypic testing may very well clarify resistance relationships to drugs other than INH and rifampin, although that remains to be demonstrated. I'll say I think that there is substantial promise to have that clarified.

We still need to understand more about the possibility of laboratory cross-contamination. We've talked about that for culture, and it certainly exists for nucleic acid amplification kinds of tests. And we still are uncertain about the interpretation of many forms of resistance. But I'll just add, in those same situations, we're, I think, quite uncertain about the interpretation of phenotypic drug resistance testing for drugs like ethambutol, the second-line drugs, et cetera.

And I wonder if we substantially underestimate the prevalence of resistance in certain situations. And so this is an old study from Korea that looks at the activity of a 12-month regimen for retreatment tuberculosis. So they used 12 months of rifampin and ethambutol given seven days a week for patients who had previously been treated. And this analysis shows the outcomes by the baseline ethambutol susceptibility result.

And so what they show is that it meant something. So patients who had resistance at baseline, whose isolates had resistance at baseline to INH and ethambutol, did very poorly. So 67 percent of those had treatment failure, and almost all of those had the acquisition of rifampin resistance.

But look at the results for those who had phenotypic resistance to INH alone, that is, the phenotype result said that it was ethambutol susceptible. One would expect that seven-days-a-week therapy with rifampin and ethambutol would be very successful in that clinical situation, and yet 27 percent of the patients failed therapy, again, almost all of whom developed phenotypic resistance to rifampin in the process.

So this is remarkably poor clinical performance from a regimen that "should work." And I suspect this represents undetected resistance that was engendered by prior therapy.

We have a nice, clear analogy of that for HIV therapeutics, and I'll just show it quickly. This is a comparison of two different ways of looking for resistance in HIV, one of those standard genotypic testing which detects subpopulations that are present in more than 25 percent of the circulating viral swarm versus a very, very sensitive genotypic form called ultra-deep sequencing. And it shows that for -- this slide compares the proportion of patients who had never been treated before, who had isolates with some forms of resistance.

So you can see, for any form of resistance, whether it's a non-

nucleoside drug, nucleoside drugs, protease inhibitors, the resistance was more common with this very sensitive genotypic test. And, importantly, that meant something. And so the risk of treatment failure in those who had resistance only with the ultra-deep sequencing was almost three times that of patients who had susceptible virus. So it shows that there can be forms of infectious diseases in which our current tests don't have adequate sensitivity for some forms of resistance.

So I think there are a number of reasons to suspect that resistance may be underestimated by phenotypic testing. It's already been shown that some rpoB mutations have produced low-level resistance that appears to be clinically relevant and yet are not routinely detected in broth systems.

And then I really worry about this issue of minority variants, that is, resistance mutations that have a fitness cost and therefore may not be seen in phenotypic results. I suspect those may explain the worse outcomes in patients with prior tuberculosis, even when you correct for the degree of resistance that their isolate appears to have by phenotypic testing.

So to summarize rapid resistance testing, I think they should be a very high priority for TB control. We need them both in low-burden and high-burden settings. The need is greatest for rifampin, but I think it's also present for other drugs. I really suspect, in the future, that genotypic resistance testing may be superior to phenotypic testing, particularly for

these weaker drugs, ethambutol, pyrazinamide, second-line drugs. We certainly need more data, and that's speculation on my part, but that's what I would think. Certainly genotypic testing could very well be simpler in the laboratory than current phenotypic testing.

To give a quick follow-up on the cases, the young woman with cavitary pulmonary disease, who lived in South Africa, fortunately had drug-susceptible tuberculosis. We were able to make that diagnosis within just a couple of days with a rapid resistance test. And the severely ill man with AIDS from the Sudan had multidrug-resistant tuberculosis. Again, that diagnosis was made quickly with a rapid assay. And he had an excellent response to treatment for MDR-TB and then for HIV disease.

So my closing thoughts are that, as Dr. Castro will probably say, there are favorable epidemiologic trends in the United States for tuberculosis, continued decreases in cases, but there's still a need for improved diagnostic tests.

As a clinician, my high priorities are more sensitive tests for tuberculosis in these key situations: for children, for HIV positive patients, for patients with extrapulmonary-only tuberculosis. I think rapid resistance tests are critical, and we also need simpler tests with lower risk of cross-contamination.

So I'll close there. I'm happy to take questions.

DR. CALIENDO: Thank you, Dr. Burman. Very informative.

We're going to take questions at the end of the next talk.

So we're a little bit ahead of schedule, and so what we'll do is we'll go to break now and return in 30 minutes, which will be -- I'm sorry? Ten minutes? Okay. So we're back at 10:00. So it's about 10 to 10:00. I have about 12 to 10:00 right now. So if we can return back at 10:00, and then that'll give us a little more time for questions. Thanks.

(Off the record.)

(On the record.)

DR. CALIENDO: So we're going to move on to our third speaker. And this is Dr. Kenneth Castro. He's the director of tuberculosis elimination program at the CDC, and his topic will be the public health implications of reclassification.

DR. CASTRO: Thank you very much. First, I'd like to acknowledge and extend my gratitude to our sister agency for the invitation to share our perspective from a public health role that we play.

Secondly, I should tell all of you that I've been traveling with my laptop, which can be dangerous. That means I updated my presentation, so there are about three new slides that were not there in the one that you have. I understand from Shanika it's being uploaded and you will have access to it.

Then the last thing I should say is I need to acknowledge the work by Drs. Michael Iademarco, Phil LoBue, and Bonnie Plikaytis in preparing

for this.

Before I go ahead, I will give you the three communication messages that I'd like to leave you with. Number one is that the scope I'll be focusing on is on the diagnosis of *Mycobacterium tuberculosis* complex and of drug resistance. I hope to convince you that the rapid and accurate diagnostic tools for tuberculosis are crucial for achieving the goal of elimination in the United States within the foreseeable future.

Second, you will see evidence that diagnostic delays were at the heart of the resurgent tuberculosis in the 1990s and continue to contribute to ongoing outbreaks.

And, third, I hope to show you that special controls exist to mitigate the risks associated with inaccurate test results for rapid detection of tuberculosis and drug resistance.

So having given you the messages, I'm going to torture you with about 24 slides.

The perspective is that elimination is an agreed-upon national goal, and we consider it to be important and a worthwhile social objective. There's been a minimum of 15 years accrued in experience with molecular diagnostic tests. There are incredible opportunities, from the perspective of someone who's worked in tuberculosis for a number of years, with substantial advances in technology.

We must address the perceived barriers and market

impediments for these new tools, especially in the setting of declining U.S. TB rates. This is not going to be seen as the big market for tuberculosis. Public health risk of diagnostic delays, if providers lack access to new TB tools in the United States, is something that we must consider.

And you will see some information that will show the role that CDC has played with partner agencies, as well as professional organizations, in mitigating inaccurate tests through the development of evidence-based guidelines, through the support of the national lab network, and through the provision of referral lab services.

And as I mentioned, the focus then is going to be assess potential benefits of the new tests, risks of the inaccurate diagnosis, and the risk mitigation.

This graph shows the TB case rate per 100,000 population from 1993 to the year 2008, and what you will see is that, following an annual percent decline of 7.3 percent fewer cases from '93 to the year 2000, that rate of decline has slowed. And if we are going to achieve elimination by the year 2050, we need to accelerate the rate of decline to 8.8 percent, which is only going to be achieved by implementing the use of new tools, shorter treatment regimens, and ideally a new safe and effective vaccine.

In fact, I tell people jokingly at CDC that one of my goals is to reorganize our program into the national immunization program because there's such a good vaccine.

The three priorities in tuberculosis prevention control are one focused on those with pulmonary tuberculosis, because there's a source of transmission to others, and a prompt diagnosis and effective treatment are a primary public health intervention to prevent added morbidity and transmission. Early detection of drug resistance and of HIV infection are essential, given the high cost in terms of morbidity, mortality, and transmission. And, third, the diagnosis and treatment of persons at high risk of progression from latent tuberculosis infection to disease is critical for progress towards elimination.

In some of our modeling exercises, it is quite clear that we must address people with latent tuberculosis infection, and I'll tell you a bit more about that group in our country.

The scope of the challenge, as we see it, is about 75 percent of TB cases seem to originate as a result of remote latent infection that becomes reactivated or activated. Twenty-five percent appear to be due to recent transmission. Up to about 20 percent of TB cases appear to be the result of transmission from persons with AFB sputum smear negative but culture positive tuberculosis. The estimated prevalence of latent infection, based on the tuberculin skin test survey in 1999 and 2000, was 4.2 percent. That means that anywhere from 9 to 11 million persons in our country are latently infected.

Clinical realm for diagnosis and treatment is critical for public

health. We've estimated about 200,000 TB suspects per year. Also, around every case of tuberculosis, their contacts must be investigated, and those are estimated to represent about 100,000 persons per year that undergo targeted testing and diagnostic procedures for tuberculosis.

And there are delays at every step of the way. The new tools, including rapid and accurate diagnostic tests, are critical for elimination. And I'll keep repeating that as the mantra.

The type of diagnostic delays that we see originate with patients who delay seeking medical care even when symptomatic. It sometimes has to do with access, multiple provider settings, and, you know, busy persons like us, if you feel sick, you're not prone to stay away from work to go get that diagnostic procedure done. And so those are some of the aspects related to delays from the patient side.

From the provider side, we've seen that as TB becomes less frequent in our country, there are many clinicians who don't think of tuberculosis when they see someone with respiratory symptoms and are prone to diagnosing lower respiratory tract infection.

You've heard from Dr. Burman. The propensity to use fluoroquinolones as recommended for community acquired pneumonia, only to do partial treatment of an undiagnosed tuberculosis case; the specimen, getting the specimen to the lab; the pre-analytic, analytic, and post-analytic phases of diagnosis; the provider seeing and acting on the results so that

proper treatment initiates and takes place. And there are several inter-related system delays. For example, patients may initially be seen in the private sector, then move on to the public sector, but no good lines of communication, and the delay is not linear.

This slide summarizes a report by Karen Mitruka and colleagues, published in the March issue of *Emerging Infectious Diseases*, summarizing the experience with multiple outbreaks that CDC has been invited to investigate between the years 2002 and 2008.

The most frequent contributing factor was prolonged infectiousness due to the delays in diagnosis. Patients delayed seeking medical attention for symptoms and, once they did, healthcare providers did not initially suspect tuberculosis.

We also see that, in the United States, tuberculosis preys on the poor and disadvantaged, so it's becoming individuals who had a history of substance abuse, homelessness were the ones who characterize most of these outbreaks.

The new tools that offer additional opportunities for prompt diagnosis would decrease the cumulative delays, decrease transmission and associated drug resistance.

Here's a schematic diagram from that report to show symptom onset, the multiple arrows representing the time when patients would seek medical attention for TB symptoms. In the lower side you'll see the health

provider first thinking of tuberculosis and perhaps ordering a diagnostic evaluation. If a person is suspected of having tuberculosis, we often recommend that treatment should begin. The lab results will support the diagnosis, and additional information, such as drug susceptibility testing, is what enables optimal use of the regimen for the individual patient and to achieve the curative treatment.

The importance of timely lab results is reflected here because TB treatment should ideally begin, but some providers wait for confirmatory lab findings, which adds to further delays, especially when you're dealing with a slow-growing organism and phenotypic testing. If drug resistance is not initially suspected, the initial choice of treatment regimen is likely inadequate, which prolongs infectiousness and exposure risk to others in the community. And infected contacts may also be placed on inadequate regimens if they happen to be infected with drug-resistant tuberculosis.

This is a summary of turnaround times recommended for the various lab tests available in our country. So specimen delivery turnaround time should be ideally within 24 hours of collection; report of an AFB smear result within 24 hours of specimen receipt in the lab; report of nucleic acid amplification test results within 48 hours of receipt; identification of *M. tuberculosis* within 21 days of specimen receipt; and report of first-line drug susceptibility test results within 28 days of specimen receipt.

The table below this shows the lab aggregate report, the most

recent data, which summarizes the 2009 public health labs achieving the goals within the time frame. So you can see that specimen receipt within one day, only 43 percent achieve it. But then the rest of them are pretty close to targets: smear result within one day is 89 percent; positive NAAT within 48 hours, 76 percent; ID of *M. tuberculosis* within 21 days, 72 percent; but drug susceptibility result within 28 days, only 49 percent. So you can see how we could use ways to improve and accelerate this.

Example of delay related to lab testing. Twenty-eight percent of patients with negative sputum smears and positive cultures are not started on treatment until the results are available based on the national surveillance system data. That has not yet been published.

I mentioned the slow-growing bug. So even with liquid culture, it takes weeks for the organism to grow, and 72 percent of public health labs meet benchmark of identifying it within 21 days of receipt, as I just showed you. This all underscores the need for rapid -- and by rapid, we mean within hours -- an accurate test for diagnosis, especially if the smear is negative, as very nicely underscored by Dr. Burman in his presentation.

So in vitro diagnostic classification rests on theoretical risk of an inaccurate result, and here's a CDC perspective that I'm going to be showing you in a series of tables concerning four levels. One is the individual patient level, with a potential new test used by itself; the new test in the context of other diagnostic tests and recommended diagnostic algorithm; the

epidemiological context, examining the predictive value positive and negative; and the public health context, given the strategy to eliminate tuberculosis and the situation of an individual patient associated with transmission and risk to others.

So as I mentioned, one way to mitigate the risk is by relying on the national lab network. CDC presently supports, by using about $9 million annually, this network of labs. It funds about 1500 public -- I'm sorry. It includes 1500 public and private labs that provide some level of mycobacteriology service, about 300,000 specimens in CDC-funded public health sector annually. This is likely to be millions if you included the private sector.

A local legal and regulatory framework exists, and workforce training relies heavily on groups such as ASM and APHL, the Association of Public Health Labs and the American Society for Microbiology. External quality assurance is provided by a proficiency testing program that we've had for many years, CDC's Model Performance Evaluation Program.

There is infrastructure and logistics that we also pay attention to regarding specimen transport, and even defray the cost for individuals to send specimens to some of these labs. Then evidence-based guideline is developed in concert with the Association of Public Health Labs, CLSI, CMS, FDA, NIH, and OSHA.

I wanted to touch on this because we're going to be mentioning

this in about two more slides. CDC has been offering since September of 2009 MDDR service, which stands for the molecular detection of drug resistance, and this service identifies genetic mutations in nine loci that enable us to identify resistance to seven drug classes listed on the left-hand side of the slide.

This is the experience that has been accrued until fairly recently with the first 375 samples. The mean turnaround time using this service is 2.4 days with a median of 2 days, compared with 36.8 days mean turnaround time with conventional drug susceptibility testing and a median of 33 days. And we're increasingly using this. These are lab-developed tests that are tested and validated and CLIA-approved.

So now moving to the tables that I promised I was going to share with you, starting with the individual patient level, a potential new test alone. What is the risk of a false positive result? Well, there's the risk of toxicity of using first-line drugs in someone who shouldn't have been treated for TB, and in terms of drug resistance, relying on less effective and more toxic second-line drugs, as opposed to highly effective short-course chemotherapy, which parenthetically there's nothing short about six months of therapy.

Risk mitigation that exists via ATS, CDC, and IDSA guidelines rely on minimizing the harm of adverse events, and the same thing holds true for the drug-resistance testing.

What about the false negative result at the individual level? If they have tuberculosis and we fail to diagnosis it, that leads to increased morbidity and mortality. The same thing holds true for drug resistance that's not identified.

Risk mitigation is emphasizing on clinical diagnosis. As you heard from Dr. Burman, you don't rely on a single test. You have clinical suspicion and other tests that are available, and for drug resistance, clinical monitoring of response to treatment as well as conventional drug susceptibility testing.

The benefit of the new rapid test would be prompter initiation of treatment, which decreases the morbidity and mortality, decreases the risk of unnecessary testing such as bronchoscopy, more invasive procedures and therapy. Also rapid identification of drug resistance then enables prompter customized treatment, decreases the risk of acquired drug resistance and its amplification.

I can't tell you how often you see individuals who are treating with inadequate regimens and the clinicians see the regimen's -- the patient's not responding, and they add another drug, only pronouncing drug resistance in that setting by ignorance of the drug susceptibility pattern.

In the context of other tests in recommended diagnostic algorithms, you're going to find it very similar. False positive results lead to greater accuracy than smear microscopy, less overall risk, and in terms of

drug resistance, multiple other methods available decrease the overall risk.

Risk mitigation. All tests are adjunctive. They're not used by themselves. And the MDDR service we see as a way to mitigate the potential false positive result in a fairly rapid manner without having to wait the 28 days through conventional phenotypic methods.

For false negative results, there's a lower likelihood of increased morbidity and mortality, and the same thing holds true for drug resistance.

In this case, the risk mitigation is, again, via guidelines. Need for clinical diagnosis is emphasized. All tests are adjunctive to establishing the diagnosis, and the same thing holds true for drug resistance.

The benefit of using this in a diagnostic algorithm is that you improve the pre-test probability, therefore improving the predictive value positive relative to smear and less drug toxicity. In the case of drug resistance, you improve overall accuracy of conventional DST, which, as you saw, is imperfect. And it also allows for prompt change in therapy before the patient fails treatment for multiple weeks before any action is taken.

In the epidemiologic context, examining the predictive value positive and negative, considering practice and prevalence, the false positive result would lead to -- in the setting of a low prevalence and therefore lower predictive value. And I'll show you some data for that. In terms of drug resistance, also low prevalence and therefore lower positive predictive value.

However, please be reminded that, in the United States of America, 99 percent of culture positive TB cases have drug susceptible TB. So a test with an outstanding negative predictive value really assures the clinician that this person has drug-susceptible TB; get them on the standardized regimen, get them cured. Then you identify a smaller group that will need the added tests.

Risk mitigation for the false negative results are going to be focused on clinical diagnosis; again, the fact that the tests are adjunctive to establishing the diagnosis, and for drug resistance, clinical monitoring. And if there's a risk such as foreign birth, exposure to someone with drug resistance, then you could do the MDDR service and get results within a couple of days.

The benefit of this approach is reduced cost of infection control that would not be initiated when not necessary, more accurate surveillance, and for drug resistance, prompter suspicion of multidrug-resistant TB, ruling out MDR and more accurate drug resistance surveillance, which is something we've been undertaking in the country since 1993.

The final table is in the public health context and taking into consideration transmission. So the false positive leads of MTB diagnosis is cost of lower utility contact investigation as well as, in the case of drug resistance, the cost of lower utility interventions.

The risk mitigation is going to rely on culture, national

genotyping that we are relying on; for drug resistance, conventional DST as well as the MDDR service.

False negative in this setting, we believe there's little to none for establishing the diagnosis of TB as well as of drug resistance.

And the risk mitigation strategies are listed here: to prioritize cluster, guide contact investigations, and the national tuberculosis surveillance system genotyping, which is using an aberration detection to predict cluster growth and the likelihood of growth in outbreaks. The strategies to assess risk of drug resistance in clusters can also guide contact investigations.

Again, the benefit listed here: more accurate, prompter results relative to smear microscopy; you reduce transmission by earlier intervention for both the diagnosis of TB as well as establishing the diagnosis of drug resistance.

The next table that I'm sharing with you I just lifted from having participated in the Strategic and Technical Advisory Group meeting of WHO, where they were discussing the new PCR-based tests to detect rifampin resistance, and it basically shows you the graph for positive predictive value and negative predictive value relative to the prevalence of rifampin resistance.

So if you'll please recall, I mentioned that, in the United States where there's a one-percent prevalence of rifampin resistance, you would be

dealing with a 32.4-percent positive predictive value, but a 99.9 percent negative predictive value. We would end up identifying a total of 29.3 -- I'll round it up to 30 persons with positive rifampin resistance that would need to be tested with these other methods. Out of those, nine and a half would have a true positive test and the remainder would have a true negative and half a person would have a false negative test, given a population of 1,000 persons. So this is just to put in that context.

I will touch on and summarize the whole set of multiple tables when talking about new diagnostics for latent TB in one table. So the risk of a false positive result would be, for the diagnosis of latent TB, the potential for using toxic drugs in persons who wouldn't need them, and from the public health perspective, the added cost of lower yield interventions.

Risk mitigation that we would use is to tell the world, don't test anyone unless you have a pre-test probability of a higher risk. And those are the guidelines that presently exist, issued by ATS, CDC, and IDSA, for targeted testing; and from a public health perspective, strategies to determine the algorithms to be used for targeted testing based on the epidemiology of the disease.

As far as a false negative result, there is a potential for higher risk of progression to disease, if the person were latently infected. From a public health perspective, anyone who progresses to disease then becomes a source of transmission to others.

Risk mitigation, again relying on guidelines, would try to identify those at highest risk of disease progression, such as HIV-infected persons who are diagnosed with rheumatoid arthritis and being given tumor necrosis alpha blockers and rendering them immunodeficient. We would also use strategies for targeted testing.

The benefit would be decreased morbidity and mortality for the diagnosis of latent TB, if we can intervene and have effective regimens to prevent it, which we do. And from a public health perspective, it is going to be crucial for elimination to be achieved.

I did not show you the graph, but take my word for it, if we don't deal with latent tuberculosis infection, elimination is not going to be seen in my lifetime nor that of my grandchildren.

In conclusion, from a public health perspective, in considering the theoretical risks focused on device indications, the public health case argues for a larger perspective in balancing the benefits and risks.

CDC and partners, sister agencies as well as professional organizations, have a trajectory and strong precedence for establishing clinical and public health practice through guidance that would be used for mitigating risk.

The adjunctive approach to rapid TB diagnostic tests and the TB CDC-sponsored lab networks is a safety net that exists in our country to largely mitigate the risk.

Risks for devices that diagnose tuberculosis, drug resistance, and latent tuberculosis are not high but can be further mitigated with general and special controls guidance.

There is an integral relationship between public and private investment, research and development for new diagnostic tools, and public health and regulatory frameworks.

Last, but not least, without these new tools, which capitalize on advances in technology, we place the public at increased risk and delay elimination.

I will add a personal comment. I found it ironic that in my participation at the WHO TB Strategic and Technical Advisory Group, I'm looking at tests being approved elsewhere in the world, but the United States doesn't have access to them.

These are persons who contributed to the presentation I've shared with you.

And I hope to be able to answer any questions you might have. Thank you.

DR. CALIENDO: Thank you, Dr. Castro.

So what we're going to do now is open it up to the Panel to ask questions to any of the previous three speakers. So that would be Dr. Gitterman, Dr. Burman, and Dr. Castro. So I want to remind you to please state your name before you ask a question.

DR. GORDIN: So I have a few for Steve. So getting, I guess, to the main reason why we're here trying to still look at this Class III and Class II, let me ask you what, in training, we talked about yesterday. And I'm still just trying to get it from you as an official spokesperson for this.

So if you have Class III and require a PMA and you have Class II with extra guidance that you could create, as you showed us, information that would be almost of the same caliber, what would you say are still the four or five biggest differences between what would be required, you know, in a PMA versus a Class II? If you built up the Class II guidance to have more or less similar trials, data, information, et cetera, what would be the key differences?

DR. GITTERMAN: I think that's a very good question, Dr. Gordin. I will also say, too, you know, with having Dr. Hojvat at the table and having -- there's a number of people here who perhaps may want to embellish the answer.

But my perspective is, in fact, in certain cases, and perhaps this one, you could, in essence, if I'm understanding your directly, under the aspect, so to speak, of clinical performance and analytical performance, you could, in essence, require identically these same type of studies, the same type of information you would want to be submitted.

Now, you said what were the four, let's say, biggest differences would be? You do not get BIMO biomedical research monitoring of the

clinical sites with the submission. You do not get a required manufacturing inspection site of the sponsor.

But in terms of what you do not get -- well, some of these are little more nuanced. But in terms of what you would actually require in terms of the degree of specificity of the trial, the scientific rigor you would accept the development program of the sponsor to be, they in fact could be identical.

Am I answering your question?

DR. GORDIN: Yes, a couple of specific things. So under a PMA, your slide had said a PMA requires a summary of safety and effectiveness data, and that differs from a 510. That seemed odd. I would've thought, in a 510, you would still want the same summary of safety and effectiveness. Can I assume it would be?

DR. GITTERMAN: Before I misspeak, I will allow someone else, because I misspeak. But I would perhaps look at that in just more so perhaps the way the regulations are written. There's also a statement that I think Ms. Shulman presented earlier, that there has to be a statement in a PMA, that there has to be valid scientific evidence. Well, of course a 510(k) would need valid scientific evidence, but it's not exactly stated as such. What's really repeated there, Dr. Gordin, is more rote than I think, in fact, what exists in practice.

So making that distinction that a 510(k) would not have it

versus a PMA, I'm not quite sure I would look at it that way, of course. And even if it's not an explicit section of a 510(k), of course that would be looked at, you know, in any judgment that would be made regarding that.

Sally, would you like to add anything to that?

DR. HOJVAT: I mean, we do even for the 510(k) now, opposed to summaries, on the web so people can see. It's not quite as extensive as the summary of safety and effectiveness for a PMA, but most of the information that other people and manufacturers would want to know, how something was cleared or something was approved, is all readily available.

I think it's just sort of a regulation states that it needs something called an SSED. But we, in fact, are very open now and transparent in letting people know how we -- what we needed to clear something in a 510(k). And you see that in a PMA as well.

DR. GORDIN: Can I just ask one last question, then? So the only other or one other difference -- I'm sure there's many, many -- it looked like, in a Class II, you could basically say your product was the same as an already approved product. You pointed out that we're trying to get to the clinical truth. That was an interesting phrase.

Knowing how hard that is for TB, if these are switched to Class II, could you still put in guidance such that, in so many words, we're trying to get to the clinical truth, not just show you're as good as another product? That as we just went through, all of them, including all of the old

100-year-old ones, have all kinds of problems, but we're trying to get better and better tools, obviously.

DR. GITTERMAN: I'm so glad you brought that up. The answer is yes. In fact, in the special controls guidance, one would specify what the nature of the reference method would be the clinical truth. And, in fact, it can be to an absolute standard rather than relative to another device.

Really this is, you know, a very nuanced point that we can discuss probably at length around the table. But in specific answer, yes, we could write a special controls guidance that would say this is what we would expect as the reference method. And the reference method would not be a previous device. It would be a combination thereof. And I believe there's a question that addresses that later to the Panel.

DR. HENRICKSON: Kelly Henrickson.

Along those same lines, one of the controls that allow us to feel comfortable that we don't need manufacturing inspections. So what are the regulatory controls that will be in place to ensure that manufacturers follow the kinds of reproducibility? You know, what are the quality system controls that allow us not to worry about inspecting?

DR. GITTERMAN: I kept it off on purpose. I only turn it on when I have a good answer.

(Laughter.)

DR. GITTERMAN: Or press it when I know the answer. You

asked two questions, and is there a limit on questions I could be asked, Dr. Caliendo?

(Laughter.)

DR. GITTERMAN: Because they're getting tougher.

But the first half of your question, which I can address easily, which is the special controls can outline what the analytical requirements would be. That's not the main thrust of your question. So if you were to say -- I'll have to say, these are very difficult things. Before working, I would've never realized how sophisticated some of these are.

But, you know, in terms of what the Panel will want to study for interferon interferences, you know, just drugs, you know, normal biological products, et cetera, et cetera, you could specify a great number of those into some depth within the special controls guidance.

Now, Sally actually is far more familiar with the QSR requirements and, in fact, you might want to address those.

DR. HOJVAT: Every manufacturer is supposed to follow the quality system regulation. The difference is that, for a 510(k), we're trusting the manufacturer to have all that information in place at the site. If they get inspected for some reason, it could be a postmarketing problem, they are supposed to be inspected every so often, about three to five years, everything would be -- is supposed to be in place.

So the presumption with a 510(k) is you're doing it, we're

trusting you, we don't need to send in an inspection to do that double-check which we ask for with a PMA.

Does that answer? It's sort of we're trusting the manufacturers. They know this is the regulation. They have to have all the manufacturing methodology, traceability of everything in a kit.

DR. HENRICKSON: So there's these ISO -- all these kinds of manufacturing --

DR. HOJVAT: Well, ISO is the European.

DR. HENRICKSON: Okay. But there's other --

DR. HOJVAT: The quality system regulation of FDA.

DR. HENRICKSON: -- quality system regulations. Okay.

DR. HOJVAT: And they're very similar, but there are some differences. And every manufacturer is expected to have in place all of the pieces that would -- if an inspector went in, they could find it.

DR. HENRICKSON: Thank you.

DR. DELLA-LATTA: Phyllis Della-Latta from Columbia University Medical Center in New York City.

My question is involving extrapulmonary TB specimens, specimens, extrapulmonary problems in extrapulmonary TB, cases like gastric isolates as that type of specimen for pediatrics.

Would expanding the category of the specimen beyond pulmonary to extrapulmonary, would that involve an additional submission as

-- let's say, for example, to be more specific, there's a test out there that we all know, that I don't even if we're allowed to mention names of different products or whatever, but it's out for two decades and is approved for pulmonary specimens only.

For that to be expanded to extrapulmonary specimens, could -- does that involve an additional PMA or would that involve -- because that's what they needed initially, or would that involve maybe a 510(k) because it's been used successfully around the world for over 20 years?

DR. GITTERMAN: That's really a very, very good question, and some people on the side were smiling because there's some discussion about the depths which we wanted to go into that at this meeting. I'll also mention, too, after I mangle the answer, Ms. Poole, who's sitting there, who really actually is an expert in this, might want to correct me.

But the first point is the intended use for the device, regardless of whether it's a 510(k) or a PMA, will specify how it's used, the nature of the specimens to be used. Now, it's probably inconceivable that there will be a device only for the intended use for extrapulmonary TB. Certainly somebody would come in with that intended use.

So let's take the simple case. Somebody would come in with sputum or -- I don't want to make a lot of nuances, but specimen is part of the intended use, and whether you have BAL -- you know, if you wanted intended use for, let's say, the presumptive as an aid in the diagnosis of

pulmonary tuberculosis, with sputum, if you wanted to include BAL. In fact, because that is a different specimen type, that would have to be studied and that would part of the intended use.

If someone -- and again, I think it's perhaps -- the discussion now. But if at the time of initial submission, people had sufficient data to support extrapulmonary uses -- and just across the table we all recognize that these are all very different, the sensitivities, specificities, positive, negative predictive values, the nature of the actual matrix, which is studied in joint fluid, is completely different, whether it's joint fluid, peritoneal fluid, pleural fluid, CSF. You know, Dr. Burman showed a slide of CSF, which I have to say was a little optimistic for a lot of the studies that I've seen.

So the point is, if they had information regarding that with the initial submission, it would come in in the initial submission. If it's a PMA, if they subsequently generated data and wished to expand the intended use or revise the labeling on that basis, it would come in as a supplement. Once it's a PMA, no other additional intended uses for that devices -- for that device -- excuse me -- would then come in as a 510(k). It would be a PMA at that point.

Freddie, you have a much better handle on this than I do. And Freddie, if I could ask a question, which I also think you asked. If something came in as a 510(k) and then let's say -- and I'm not even suggesting this. You know, we're completely agnostic on this issue. But let's say there was a

consensus for reclassification for pulmonary tuberculosis and ultimately a 510(k) did come in, and a supplement which then expanded the intended use to other matrices such as pleural fluid or joint fluid, would that then come in as a 510(k) or would that be a PMA? Please.

MS. POOLE: Okay, Freddie Poole.

If there's a change in the indications for use and you have an existing PMA, then you would submit a PMA supplement. There are two types of PMA supplements you could submit. But as long as there was a PMA, you can't submit a 510(k). If you did submit the 510(k), we would contact you and say this is the same intended use as one that was cleared under -- approved under a PMA, so you would then have to submit a PMA supplement.

DR. DELLA-LATTA: This is a little tongue in cheek.

MS. POOLE: Yes.

DR. DELLA-LATTA: But if the intended use is to diagnose tuberculosis, a nice, broad intended use --

MS. POOLE: Um-hum.

DR. DELLA-LATTA: -- then that would go for pulmonary as well as extrapulmonary tuberculosis?

MS. POOLE: Yes, but when you did your clinical trials, you only did it using pulmonary TB --

DR. DELLA-LATTA: True.

MS. POOLE: -- as meant.

DR. DELLA-LATTA: True.

MS. POOLE: So unless you had a sufficient number, because sometimes in our labeling you would see under limitation that there were not sufficient specimens available from other type populations, so we would put it as a limitation, and they would say the performance in these type populations have not been fully studied.

DR. DELLA-LATTA: So let me just -- be ready for this. If you were to reclassify --

MS. POOLE: Um-hum.

DR. DELLA-LATTA: -- a PMA assay --

MS. POOLE: Yes.

DR. DELLA-LATTA: -- to a 510(k) first, let's say, for pulmonary --

MS. POOLE: Um-hum.

DR. DELLA-LATTA: -- and then you were to add extrapulmonary, you're saying you have to go PMA?

MS. POOLE: No, no. Then that would come as a --

DR. DELLA-LATTA: Then that would be a 510(k)?

MS. POOLE: A 510(k), yes.

DR. DELLA-LATTA: So that's an interesting possibility. Okay, thank you.

MS. POOLE: Okay.

DR. HAMMERSCHLAG: Margaret Hammerschlag, Downstate Medical Center, Brooklyn.

I just think a little clarification, though, about the use of gastric aspirates in children. It's basically used as a surrogate for pulmonary, and so the diagnosis of pulmonary tuberculosis, because young children do not expectorate for the most part. We don't get sputum from them.

So the idea is to get three in a row of these morning gastric aspirates, which represents sputum, et cetera, that's been swallowed during the night. And, of course, you're dealing with a specimen that has fun stuff in it, you know, gastric acid, et cetera, plus also commensal mycobacteria so that you AFB, so that smear is useless. And this is a major problem, particularly for the children.

I mean, Dr. Burman, the issue about kids is that this -- you know, the specimens are just different. And so really the data are extremely limited about the use of these assays under these circumstances. Yet this is a population that, because of the fact they're young children, they're more likely to disseminate and have more serious problems.

DR. GITTERMAN: Well, I can answer the question, actually. The first thing I have to just say with a smile on my face is, using the surrogate in an FDA meeting, you know, immediately has about 18 statisticians leaping on you.

(Laughter.)

DR. GITTERMAN: Luckily we've restrained the ones in the audience.

But actually it's a very good question because I would perhaps, you know, quibble with your choice of words. I wouldn't say it's a surrogate. I'd say it's a different way of diagnosing it. It's a different specimen.

But, you know, first sort of the mom and apple pie. Actually FDA is very committed, and it's an increasing recognition among all aspects of FDA about the importance of diagnostics in children. You know, for drugs, obviously there's -- you know, it's very well.

But to address your point, number one is we would address this, hopefully, in special controls, which, one, it's a different matrix. We would want to see, analytically, the performance of the test in the setting of gastric acid so we could -- you know, again, it's really good that you brought this up. One is we could specify in guidance that a sponsor test the performance, and as Ms. Poole had said, we could then list that analytical performance in that.

If a sponsor -- and they'd be strongly encouraged -- did do studies of gastric aspirate -- and, of course, no one would require studies of the same size or breadth that they would for pulmonary tuberculosis in adults, but one would describe what the performance was in that setting. And in the setting, one would hope, if this was spelled out, that in fact this would be an encouragement to do so. It's a challenge because -- I'll go back

to Dr. Della-Latta's question.

You know, what one would require and what the standard should be and what's sort of acceptable, these are not trivial issues. But I think that would be addressed very well. Now, again, there may be devices, of course, which are completely inappropriate for use in children and, you know, we would require something just for that purpose.

So does that answer your question?

DR. HAMMERSCHLAG: Somewhat. I mean, I'm not aware of any sponsor that's actually specifically looked at that specimen type. There have been studies that have been done that have attempted to use these assays sort of on their own, but it's been an investigator from various clinical sites that have tended to look at this, as well as to look at, let's say, spinal fluid because that's obviously something else we deal with maybe more often in children.

DR. GITTERMAN: Yeah.

DR. HAMMERSCHLAG: But it would need, I think, to be looked at more specifically because there's always the issue of, you know, having an adequate number of patients and the problem that we -- otherwise we have nothing. You can't do a smear on these specimens and it becomes a real problem because little kids just -- well, you know, they just don't look -- they're not large, they're not small adults.

DR. GITTERMAN: I could not agree with you more. I will say

just in passing, having seen -- some of the speakers of the Open Public Hearing, you may be able to address that question directly. Perhaps, if there's a sponsor that's presenting, what their approach would that be. So you may be able to hear from the horse's mouth. But the point is very well taken. Thank you.

DR. HAMILTON: Yes. Just to push this a little bit more --

DR. CALIENDO: Your name.

DR. HAMILTON: Oh, I'm sorry. Carol Hamilton, FHI and Duke University.

So to push this kind of in the other direction, again, for the sake of understanding the differences between classifications. So, for example, I notice that currently under Class I would be both traditional culture media and automated systems for culture, and that's in the exempt category. And, in fact, even identifying from cultured isolates, that's also in Class I. I'm just wondering, if we were sitting here today and the MGIT was being, you know, here, we were evaluating this as the first ever automated system, would that inherently be looked at in a Class I way or would be talking about Class III and Class II there? Because it's performance characteristics, as we've heard. There are certainly false positives there. There are risks of contamination, all the things we've talked about.

DR. GITTERMAN: It's sort of interesting. I've never seen mikes that turn off after a timer. That's a very good idea. Actually, I didn't realize

technology had done that.

I could partially answer your question. The answer in this isn't direct, is that if a device is unclassified, which exactly is the situation the Committee is being asked to discuss, the Panel's going to discuss today, which is the setting of specimens for direct EST, the automatic classification process makes everything Class III. And then, at the time a submission is made, FDA could decide whether to go through what's called the de novo pathway and make it Class II, or it remains as a PMA. That's for devices that unclassified.

Regarding your other point, it's not fair of me to comment on decisions that were made previously, and, you know, perhaps somebody else might want to, but points are well taken whether in fact there's a lot of considerations, both regulatory, clinical, and scientific that result in the classification.

The overall sort of rubric, which Ms. Shulman said earlier, is that devices are classified in the class which safety and effectiveness can be demonstrated -- and I'm going to use this expression, perhaps a little incorrectly -- but would be the least burdensome, so to speak. I'm just not in a position to say whether the Class I for the MGIT, in fact, would be correct or not.

But anything new, by default, that is not exempt by regulation would in fact be automatic Class III by default, unless at the time it came in we determined that it could be classified through the de novo pathway,

which is what you really are talking about for, you know, direct detection of resistant mutations associated with resistance or, in fact, the reclassification process which we're talking about for NAAT assays.

If I haven't addressed it, please, if anyone else from FDA wants to clarify it.

DR. CALIENDO: Freddie, do you want to say something?

MS. POOLE: Freddie Poole, FDA.

The ones in Class I, as Marjorie explained, with the advent of the Medical Device Act in 1976, we had convened an advisory panel, and between 1976 and 1980, that's when they classified all of the devices. And at that time TB was practically eradicated in the U.S., and so they were considered a non-significant risk. And the way the method that was used for describing the tests and all the general controls in place would be sufficient.

However, as time changed and the status of TB in the U.S. increased, then we asked for the same almost type of information that we would ask today in the Class I devices. Like when the MGIT came and the AccuProbe and all of those other tests, we started to raise our requirements for what was expected.

And then, in 1994, that was when the first NAAT test came for direct specimen testing, and we considered it and we determined that that presented a new risk of safety and effectiveness, and that risk was what would that new type of device, the information that that type of device

would provide. And there wasn't a lot of information or experience with NAAT tests at the time, and that's why we classified that and the interferon gamma release assays, those were both classified as Class III devices.

Does that answer your question?

DR. HAMILTON: Both of those answer the question in terms of the letter of the law. But in terms of kind of the spirit of what we're trying to understand, which is the benefit/risk, so to speak, of new tests and how they compare --

MS. POOLE: Yes.

DR. HAMILTON: -- to, you know, automated liquid culture systems, which themselves have a lot of complexity and require good manufacturing practices and all of these other things.

MS. POOLE: Um-hum.

DR. HAMILTON: So we're talking about these things that have this level of complexity are critically important for our diagnosis.

MS. POOLE: Um-hum.

DR. HAMILTON: And they're in this very -- you know, this Class I. And yet now we're talking about other tests that may not be any more complex or any -- you know, have maybe even better test characteristics, but we're now thinking Class III versus II.

MS. POOLE: Um-hum.

DR. HAMILTON: So I don't know that there's an answer to that.

It's maybe more of a comment for a discussion.

MS. POOLE: In fact, there were a lot of devices that were then classified as Class I. We treat them practically like a Class III, but like Sally can tell you more, but like the flu, influenza was also a Class I. But recently, with all the different variations of influenza, they've been increased, and Sally could explain how influenza, they've moved that up also, so that we could ask for more information for the influenza devices.

DR. REIMER: Larry Reimer, University of Utah.

So another kind of umbrella set of questions would be that we've talked about gold standards and clinical truth as a couple of entities here, and I wonder how the FDA defines those and where we are with new devices and what they are supposed to meet relative to gold standard versus clinical truth.

DR. GITTERMAN: Well, obviously it's going to be specific to the device. Some areas are more gray than others. In fact, I believe it's addressed in one of the questions. I could give you, you know, an off-the-cuff answer right now if we were writing -- you know, if somebody approached us around the table and said you wanted to do a test for, you know, a study of tuberculosis, we certainly would give you recommendations.

But probably more important is the process, and the process would be, if it's not straightforward, we would, no doubt, solicit the opinions of experts through a number of mechanisms, including panels, to do the best

we can.

But I don't think you -- you don't want me to answer the question, do you? You know, that could certainly give you my opinion, but free advice is worth what you pay for it. But I think, around the table, we all have a general sense.

But tuberculosis is not -- and I thought Dr. Burman did a superb job with that. You know, it's not as absolutely clearcut. I mean, walking down street, you know, I would suspect there's a fair number of clinicians who would say culture is 100 percent sensitive and 100 percent specific, and it's far from the case.

And I will ask -- you know, I will embellish it, though, and there's clearly the case we certainly see it with partially treated patients. That's the wrong expression, but patients who are early in the treatment or a NAAT assay will be positive and a culture will be negative. And as we all know, just, you know, as you've heard a lot of, you know, some of the steps in culturing, you know, the decontamination steps, they have profound effects on the sensitivity of cultures.

I mean, there's experts around the table, Dr. Wilson and others -- well, probably everyone around the table can talk about what a balance that is. You know, it's not quite an art, but it's a very sophisticated point to make sure you don't lose the yield of your mycobacterial cultures through the decontamination step.

Just a question. Is there like a line drawn that no one on this side of the table asks questions?

DR. CALIENDO: No, I'm actually going to break that line.

DR. GITTERMAN: Okay.

(Laughter.)

DR. CALIENDO: This is Angie Caliendo, and I have a question for Dr. Castro that's a completely different focus. I would like to ask you about the negative predictive value of the resistance testing.

So when I think of resistance testing that's genotypic, I always -- I kind of think that one of the issues is, is that you may miss -- you might miss mutations that you would detect phenotypically if you don't have a comprehensive understanding of the genetics. And so I was interested to hear you say that you can rely on a negative test to rule out multidrug-resistant TB.

DR. CASTRO: Thank you. That is limited to rifampin resistance because we're fortunate that mutations in the rpoB region account for 98-plus percent of rifampin resistance, unlike isoniazid, where you would need to look for mutations in INHA, the CAT G region, et cetera. So, you know, depending on the local epidemiology.

But the graph I showed really shows the negative predictive value for detecting rifampin resistance, which we see as an excellent surrogate marker of MDR because rifampin monoresistance is very

uncommon --

DR. CALIENDO: Um-hum.

DR. CASTRO: -- in our country and, for that matter, in any other place. I would certainly still want to see results for isoniazid and whether INH or CAT G mutations are present before putting my name on the MDR.

DR. CALIENDO: So is there data out there for the two mutations for INH? Do we know that, if you look in both areas and that has that kind of negative predictive value?

DR. CASTRO: I'm going to look to the back of the room to either Bev or Jamie Posey from the lab. In the experience of all of these samples that you've run, what percent of INH resistance are you picking up with MDDR, looking at the available loci? Don't rush it. Huh? About 90. Thanks.

DR. CALIENDO: Thank you.

DR. TOWNS: Michael Towns, Industry Rep. I have a question for Dr. Burman.

Thank you for your presentation, I really enjoyed it, but I was fascinated by what you presented on the false positive cultures that had occurred in approximately three percent of a large series in the numerous studies that you had. And what I was fascinated by is you said that more than half of these patients with false positive cultures were treated for TB.

So it would lead me to ask a question: How did they present

clinically with regard to did they have a moderate to high clinical suspicion of active pulmonary TB, and what were their radiographic presentations that the clinician would be putting the clinical cluster radiographic and the culture together to decide that they were going to treat these patients that had false positive cultures?

DR. BURMAN: I can't comment in a definitive way across all of these studies because obviously I don't have that information. I can comment from our own studies in which in our lab we found -- or in our city, because we looked across the entire city over about 10 years, and I can comment on the experience of those 10 or 15 false positives.

And generally we don't get cultures in someone in whom there is no suspicion of tuberculosis. Obviously you're doing it because of clinical and radiographic abnormalities. And so these patients did have abnormalities present, and some of them were extrapulmonary abnormalities that led to cultures being obtained.

Having said that, I must say that once you know about this phenomenon, often they're not difficult to pick up because it's two weeks later, they're now fine from treatment for whatever, and this comes back and if you know about the phenomenon of cross-contamination of false positive cultures, you can pick them out and say, oh, that really doesn't make sense.

Now, with all of the information we have, including some follow-up, because there's time between the time the specimen obtained and

it becomes positive such that I think they aren't so difficult to detect when you're armed with the knowledge that they occur, when you have the impression that mycobacterial culture is 100-percent specific, you say, my goodness, we can't take the risk. This person had a chest X-ray infiltrate, they had something that led to a culture being obtained, therefore we must treat it. I think that's what leads to the treatment and other ramifications. There have been large contact investigations, et cetera, all on the basis of false positive cultures.

DR. BEAVIS: Kathleen Beavis. I have a question, Dr. Gitterman, I think, unfortunately for you again, and I was hoping I could get some clarification of the word effectiveness. That word's been used a couple of times, especially prominently in our training this morning. One of the goals is to put devices in the lowest class whose level of control will provide reasonable assurance of safety and effectiveness.

And I was wondering, how is the word defined? What are some of the implications of the word effectiveness?

And I'm thinking it in the context of the different types of tests which we're being asked to consider today. If we want to talk about TB cultures, you know, as we've heard, they're not perfect, they're not 100 percent one way or the other.

But I would say that the process if fairly well understood and if somebody said to me, hey, I really think my patient has a false positive

culture, I would know where to go. I'd look in the lab for contamination. I'd look at the specimen container. Was it mislabeled up on the floor? I feel as if I understand the process, as well as with anything post-analytic that could also contribute to a false positive.

I don't have that sense for some of the tests for the TB infection. I wouldn't know how to think -- how do you want to talk about the word effective in terms of the IGRAs?

DR. GITTERMAN: I had an answer. It was an easy question before the IGRAs.

(Laughter.)

DR. BEAVIS: Well, it hasn't been addressed --

DR. GITTERMAN: Right.

DR. BEAVIS: -- this morning and --

DR. GITTERMAN: Well, let me ask. You're making an association, which is not coincidental. First, do you want me to answer the question for the non-IGRAs or is it a question only for the IGRAs?

DR. BEAVIS: I feel as if I understand how to look at whether a test for culture, traditional culture, is effective or not. I feel as if I have a gold standard against which, you know, things can be measured. And I understand where, you know, current practice and maybe the gold standard, you know, either overlap or where the intersection is not defined. I feel as if I can talk about effectiveness with those tests. I need some help in thinking about

effectiveness with IGRAs when I don't have it for that. And I'm sorry I'm being more incoherent than usual, but --

DR. GITTERMAN: No, you're being very straightforward. It's a very difficult question, and if you don't mind, it actually is a question directly -- you know, that actually is one of the three questions. It's the third, and it's a very challenging question. If you don't mind, I'll defer an answer if you want me to come back after the discussion starts. But that's a far more difficult question.

What we can talk about fairly, when the Committee gets to that later, is what the precedence then is, how studies are done with IGRAs. And you're really getting -- honestly, it's a multifaceted question. What would be the clinical truth? You know, that's your first question because, when you talk about effectiveness, it has to be against a reference standard. That by itself is a very long, fairly sophisticated question.

So if it's okay, I could certainly give a very, you know, cogent answer for the non-IGRAs, but the IGRAs, that's why they're separate questions. And, again, this would be a good opportunity to emphasize that we're really talking about three separate issues here. We're talking about, you know, NAAT tests for the detection of tuberculosis directly from clinical specimens, for the detection of resistance mutations in the same scenario, and IGRAs, as three separate questions. So if it's okay, I'll defer. But the best answer I could give you is precedent and -- I'm sorry?

DR. BURMAN: I'll comment.

DR. GITTERMAN: Oh.

DR. BURMAN: I feel guilty because I think I was supposed to talk about this and I didn't. So because of my guilt, I'll comment.

(Laughter.)

DR. BURMAN: So this is Bill Burman.

I think it is a great question because we don't have a clear gold standard for the diagnosis of latent tuberculosis. It's impossible to have an infallible gold standard in that situation because, by definition, these persons cannot have culture positivity from some site or other clinical manifestations of active tuberculosis. So that does make it indeed challenging.

Therefore, how can you try to -- to what should you compare and how? Well, there's several different strategies. One, the commonly used one for sensitivity, is to look at sensitivity among persons with active tuberculosis. Of course, that's a somewhat different question, but it gives you some confidence that immune reactivity is present, as measured by a test, is or is not present in active tuberculosis. Obviously that raises other questions. The test we've used for 100 years has a sensitivity in persons who are HIV negative with culture positive pulmonary tuberculosis of only 75 percent.

And so if we're comparing IGRAs to TST for their sensitivity in active disease, the test to which we're comparing is far from perfect, with a

sensitivity of only 75 percent. But it's one way of having some assurance that the immune reactivity measured by a test at least is present in persons who clearly have tuberculosis, but acknowledging that that is a somewhat different situation than trying to diagnosis latent tuberculosis.

Another way to evaluate is to try to look at specificity in persons who have a condition which may well result in a false positive skin test and to see if it appears that these tests give you great specificity. And so a common population to involve there, or two common populations to involve are persons who've received BCG. So what's the rate of positivity with a new test in persons who have been exposed to BCG and have a very, very low epidemiologic likelihood to having been exposed to TB? That's a good comparison. And a second one is persons have a non-tuberculous mycobacterial infection.

So you look to see how does this test for immune reactivity for tuberculosis perform in these two groups that evaluate specificity, those who have been exposed to BCG and have a very low likelihood of having ever been exposed to TB, and secondly, in persons who clearly have a non-tuberculous mycobacterial infection. None of those are terribly satisfactory.

The measure I would promote as the best measure of clinical truth are the limited number of studies that look at the predictive value of tests for the development of active tuberculosis because that's what we're trying to get at with a diagnostic for latent tuberculosis. Does this test

predict a higher likelihood of this person going on in the relatively near future to develop active tuberculosis? Because that's the patient population in which we want to intervene with therapy.

And so that, I think, is the best way of evaluating an immunologically based test for tuberculosis, are the very limited number of studies that allow you to see the predictive value of a test for the development of tuberculosis within the next two to three years, for example.

Does that get at your question?

DR. BEAVIS: It does. And if I can do a quick follow-up to the study, if I'm understanding it correctly. And, again, I'm thinking of a U.S.-based population where studies would go through an IRB board.

You know, I sit on our IRB, and I'm wondering, if a study came to us and said we want to test people using TST and IGRAs and we're not going to give them the INH that we would normally give them to see if they develop TB in the next few years, what are your thoughts on that?

DR. BURMAN: Yeah, I think all of these studies are challenging to do, and that's an illustration of one of the challenges. There are other ways to go about it. One could imagine populations who -- in whom not both -- both tests aren't done and yet you consider them comparable populations and look for the risk of progression of active tuberculosis.

Other studies that have been published looked at those who refused. So it's not patients in whom you said we're going to ignore your

positive test; it's they chose to ignore their positive test. So there was a large study, for example, from Europe that looked at those patients who had both conventional testing and an IGRA done and refused to act on either one of them. And so in that you had a limited opportunity to look at the relative risk of progressing to active tuberculosis.

But I agree with you that these are tricky questions to evaluate in prospective studies in a whole series of ways, both scientifically and ethically.

DR. CALIENDO: Go ahead, Phyllis.

DR. DELLA-LATTA: I just -- excuse me. Yeah.

DR. GITTERMAN: Oh, if I can just embellish just very quickly. You know, I was talking about precedent. One of the reasons -- and in your background, I believe, is a copy of both the QuantiFERON and the T-SPOT package inserts.

One of the reasons that the intended use actually does not state latent TB is because, you know, as Bill's already said, it's almost impossible -- and then Dr. Gordin could probably comment -- the recent rifampin treatment for latent tuberculosis. What was it, 8,000 patients with 22 cases over -- or actually probably half of the audience in the back is more familiar with it than I am, but it's a huge number of patients over almost a decade that ultimately yielded 22 cases to do so.

And if you look again through the -- even though the package

insert for both tests that are cleared for indirect use do not say latent TB in the intended use. In fact, when you look at the studies, it's -- you know, the study, in fact, the study says here's a study of patients at high risk for latent TB, to do so.

But, again, the heart of this issue really gets back to the discussion, and the Committee would say it may be very difficult to write a special controls guidance for exactly the reason, because, you know, it's part of the heart of the issue.

DR. BURMAN: Thank you very much.

DR. CALIENDO: Okay. So Phyllis is going to have the last question because I want to make sure that we have adequate time for our public comment session. We will have time during our Panel discussion to ask other questions and bring people back up.

So go ahead, Phyllis.

DR. DELLA-LATTA: Phyllis Della-Latta again.

Getting away from this sort of nebulous -- I think it's both of you. Getting away from the nebulous time tests, the skin test versus the IGRAs, and going back to what we believe is more straightforward with the NAAT, as far as our gold standard is concerned, it really is not straightforward. We do know from some of the work that you had presented before, like in terms of the gold standard of culture being flawed and having false positives, I just want to throw out a statistic in New York City, where 24

percent of tuberculosis cases are culture negative.

So unfortunately, in a study, if you're only using one sputum specimen, for example, which happens to be a saliva and not a sputum, then you will end up with these very poor rates.

So as Dr. Castro had stated, tuberculosis is a clinical call, and I think the problem that we've done in evaluating many of these assays is to think that a diagnosis is going to hang on that test result. That is not true. The NAAT tests and some of these others are an extremely important piece of the diagnostic puzzle, but not the only piece. So let's not try to make them 100 percent perfect.

And the problem with some of the data out of Peru and elsewhere is that when they come up with these statistics of positive and negative predictive values and sensitivities and so on, they're not based on what I use. I insist on smear negatives to have three sputum specimens per patient, and my results are based on patient data, not specimen data. So I often can get one positive out of two negatives. And if they cannot give me three sputum specimens that are smear negative and they say she's not producing sputum, it's almost laughable. Then they have to go in for a bronch.

So it's a very complex algorithm and no one delves into how one -- how many specimens were taken per patient and lots of questions that I think are important when we bring up the specific controls that we want to

put into maybe a 510(k) and that we do have specimens in mind.

DR. GITTERMAN: Can I beg to differ with you in a very nice way? In fact, we obsess about these things, and just very quickly and then I'll turn over the mike to Dr. Burman and perhaps Dr. O'Brien, who's in the audience, might want to comment. We obsess about this, and it is a fundamental question you're asking, and let me explain very simply.

Let's say you were to say, as one -- and this I do not want to preempt in any way, shape, or form Committee discussion. But let's say, for a situation like this, you want to demand an absolute level of performance. You said we could expect from the studies we've seen, we want 99 percent specificity. Well, then the devil in the details is enormous.

As we mentioned before and anybody will tell you, you know, there's going to be culture negative specimens, which, for whatever reason, they are culture negative, but they're smear positive, and for every way -- I'm sorry. Let's say NAAT test positive. Excuse me. Even perhaps smear positive as well for every -- clearly have pulmonary tuberculosis. How do you do that?

If you look at some of the very recent large published studies for evaluation of NAAT test devices, if you look at a recent multicenter study or a multinational study, there were 13 very significant exclusion criteria. Anyone who was smear positive and culture negative was excluded. A fair number of patients with pulmonary tuberculosis who were both culture negative and assay negative were excluded.

What you're saying is absolutely right. If you hold yourself to an absolute standard, you have to be incredibly specific about how you -- I'm going to use this in a colloquial way, slice and dice. But writing up your protocol in advance so you're not resolving even again -- and there's people that are far more expert in this -- up front -- and this is a very sophisticated area, how you resolve discordant samples, how you -- the only thing I disagree with you on everything you said is that people aren't thinking about it. We think about it all the time. It is a very, very significant area.

One thing I will say, and this gets back to Dr. Beavis' question before, there's a difference between efficacy and effectiveness. Some of the major issues in the context of clinical care usually, you know, should be done. A sponsor, you know, it would be well against their best interest to not do a study as well as possible so a study will minimize as much cross-contamination and other possibilities. But the question you've raised is a tremendous issue.

And perhaps, I'm sure, Dr. Burman -- but when you look at the method section of any of these studies and, you know, the specificity is 99 percent, or even how they describe positives in, again, the recent *Lancet* paper, the name of which -- the subject of which I won't describe or I'll get slapped, the positives on some sites, the sensitivity was lower because on some sites they required solid and liquid media to be positive. It's a multinational site. Some were just liquid media. Some were just solid media.

Some used some assays which weren't even cleared, which haven't even been cleared by FDA. Blah, blah, blah, blah.

The point is this is a really difficult and important area, and there's no doubt either, one, we would have very, very specific guidance or, two, we would have general guidelines and we would make sure that the sponsors had their protocols written in advance, so the analysis plan, what was analyzed, what the point estimates were with the lower bound was all addressed because a special controls guidance, science evolves and it may be the case that a sponsor could justify things as they change. But certainly we would be very -- we would address this.

DR. DELLA-LATTA: No, no, the we -- please, the we was not addressed to the FDA. It was supposed to be sort of generic and general. And what I wanted to really bring out is the fact that people are only testing one specimen per patient, and this is a terrible disparity, that's all I'm saying in general, and coming up with statistics that I think are flawed.

DR. BURMAN: Yeah, I'd just say that you're right, that using culture as the comparator has its problems, and I think that can be dealt with, as Steve said, by carefully laying out in advance a plan of adjudicating those in which there's a discrepancy and not just assuming that culture is absolutely right. I think obviously that can't be done afterwards, but if it's done carefully beforehand, I think that can be a revealing form analysis, and I'd encourage it.

DR. CALIENDO: Okay, thanks to everybody for their

participation.

We're now going to proceed to the Open Public Hearing portion of the meeting. Public attendees are given an opportunity to address the panel, present data, information, or views relevant to the meeting agenda.

Ms. Craig will now read the Open Public Hearing Disclosure Process Statement.

MS. CRAIG: 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, FDA believes that it is important to understand the context of an individual's presentation.

For this reason, the FDA encourages you, the Open Public Hearing 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 group that may be affected by the topic of this meeting. For example, this financial information may include a company's or a group's payment of your travel, lodging, or 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.

I will now go over the process to ensure a smooth transition from one speaker to the next.

You will have five minutes for your remarks. When you begin to speak, the green light will appear. A yellow light will appear when you have one minute remaining. At the end of the five minutes, a red light will appear and the microphone will go off. Your presentation should be concluded. Since we have a number of speakers, it is very important to adhere to the five-minute time limit. As each speaker concludes their remarks, Ms. AnnMarie Williams will guide the next speaker to the podium.

The Panel will be given an opportunity to ask questions of the public presenters at the conclusion of the Open Public Hearing. If recognized by the Chair, please approach the podium to answer questions.

I would like to remind public observers at this meeting that while this meeting is open for public observation, public attendees may not participate except at the specific request of the Panel Chair.

Dr. Caliendo.

DR. CALIENDO: Okay, our first speaker will be Dr. Melissa Miller. Please come forward to the microphone. We ask that you speak clearly to allow the transcriptionist to provide an accurate transcription of the proceedings.

DR. MILLER: Great, thank you for the opportunity to speak

today. I'm here on behalf of the American Society for Microbiology. I am Melissa Miller. I have no relevant financial conflicts of interest.

The society represents over 38,000 microbiologists, including professionals and public health and clinical microbiology and immunology. And this was a consensus statement written by the committee on laboratory practices.

The ASM strongly supports the reclassification of molecular diagnostic devices for the detection of *Mycobacterium tuberculosis* from PMA approval to 510(k) clearance. Support for the reclassification is based upon the following: the existence of a predicate device that is FDA approved for the molecular detection of TB. There's only currently one in vitro diagnostic product commercially available for the TB detection. Although this product was approved through the PMA process, it requires a significant amount of molecular expertise to perform to ensure accurate results.

Although the risks to both the patient and to the public of an inaccurate test may be perceived as higher for TB, and certainly for MDR and XDR-TB, a rapid diagnostic test for TB would be used as a screening test. As a screening test, it would augment and not replace smear or mycobacterial culture.

It is the position of the ASM that the benefits of early detection of TB outweigh the risk associated with a false negative result with a test that has specificity and positive predictive value. Therefore, it should not be

categorized as a Class III device requiring PMA.

Consistent with the WHO endorsement of the Cepheid GeneXpert MTB/RIF assay as a researcher's only assay, the ASM also recognizes there is a critical need for molecular diagnostic devices for TB with the following attributes: easy to use, provides random access testing, and has high accuracy for detecting TB and both smear negative and smear positive respiratory specimens with high positive predictive values.

The ASM recommends that the FDA consider a 510(k) submission for all new molecular TB assays intended for the purpose of detecting TB from smear positive and smear negative respiratory specimens. We propose that the following minimal requirements for assay performance be considered for this mode of submission: functional equivalence to the Gen-Probe MTD assay, and then relative to culture, as we discussed the issues with that, sensitivity for smear positive respiratory specimens, at least 90 percent, sensitivity for smear negative respiratory specimens, at least 60 percent, an inhibition rate of less than five percent and not inhibited by a moderate quantity of blood, and a specificity of at least 99 percent.

It should also be emphasized that performance characteristics mentioned above should not be the only consideration for a molecular diagnostic device for TB detection.

The ASM supports a 510(k) submission pathway for devices that are also closed systems and random access. These attributes allow

laboratories without significant molecular expertise the opportunity to perform rapid molecular diagnostic tests for TB.

In addition, many larger laboratories in the U.S. continue to send TB molecular testing to reference labs because of the costs associated with having to do batch testing due to small numbers of samples. Allowing laboratories of all sizes and expertise access to TB molecular testing is critical to decrease the time to result of TB detection from direct respiratory specimens.

Although the need to quickly determine rifampin resistance is of critical clinical and public health importance, at this time the ASM does not support the Class II classification of a molecular rifampin resistance detection device.

A significant amount of clinical data will need to be obtained to determine the positive and negative predictive values in areas of low prevalence in the U.S. We realize there is a preexisting FDA-cleared closed platform multiplex assay that detects both the organism and a resistance marker such as the MRSA staph aureus test by Cepheid, but we assert that the risk to the patient and public health is much greater with the missed diagnosis of MDR-TB.

In summary, the ASM is fully supportive of reclassifying molecular diagnostic devices aimed at the detection of TB from primary respiratory samples to 510(k) clearance. Increased access to FDA-cleared

rapid molecular diagnostic devices for TB has the potential to greatly impact patient care and reduce transmission potential throughout the U.S.

Thank you for the opportunity to comment.

DR. CALIENDO: Thank you, Melissa.

Our next speaker is Dr. Fred Tenover.

DR. TENOVER: My name is Dr. Fred Tenover, and I am the Executive Director for Scientific Affairs at Cepheid, which is a molecular diagnostics company in Sunnyvale, California. In terms of full disclosure, I'm also a consulting Professor of Pathology at Stanford Medical School and Adjunct Professor of Epidemiology in the Rollins School of Public Health at Emory University. I also spent 18 years at CDC before joining Cepheid, so it's fair to say my public health hat is still half on.

I thank the FDA for the opportunity to provide comments today.

While the number of new tuberculosis infections reported in the United States is near or at all-time lows, the number of times a tuberculosis appears in the differential diagnosis of patients with lower respiratory tract disease in the United States is substantial. Many of these patients spend up to five days in costly isolation rooms in hospitals, awaiting the results of two or three acid-fast smears, which, if negative, allow them to be released into the general hospital wards. These patients may still have tuberculosis even though they are considered to be noninfectious.

Relying on the results of acid-fast smears, both for diagnosis and for infection control decisions, is archaic. The test isn't sensitive and in many hospitals is no longer performed on site, further delaying the availability of results. Better diagnostic tests are clearly warranted.

To improve the detection of patients with tuberculosis, the Centers for Disease Control and Prevention has mandated that each patient suspected of having tuberculosis, i.e., those with compatible clinical signs and symptoms, should be tested using a nucleic acid amplification test at least once. However, at present, laboratories in the United States have only one FDA-approved option for performing such testing. So the recommendation, while important, has not been widely implemented.

Reclassification of diagnostic tests for tuberculosis would increase the number of options available to laboratories to provide rapid and accurate testing on site. This would not only improve adoption of CDC recommendations, but help to optimize the use of hospital isolation rooms, which are always in short supply.

The availability of rapid tuberculosis results provided by molecular amplification methods should contribute not only to better patient management by ensuring that patients are placed on effective medications early, but will likely improve the cost effectiveness of managing tuberculosis patients in the hospital as well.

The issue of reclassifying the detection of rifampin resistance

simultaneously should also be considered. Concern about detecting rifampin resistance in low-prevalence settings such as the United States, where the rate of a true positive rifampin-resistant *M. tuberculosis* infection approximates the rate of false positive test results, is clearly justified. However, this concern must be balanced with the increased morbidity and mortality associated with delaying the detection of a multidrug-resistant or extensively drug-resistant case of tuberculosis.

Not only does the individual patient risk prolonged disease or possible death due to ineffective treatment, particularly with XDR infections, but there is the public health concern of continuing spread of multidrug-resistant organisms from inadequately treated cases to anyone with whom the infected patient has contact.

One only needs to recall the dramatic spread of multidrug-resistant strains of *M. tuberculosis* in the early 1990s in Los Angeles, Miami, and especially the streets and prisons of New York City to understand why such information on resistance can be critical, not just to the individual patient, but to the overall health of the nation.

The detection of multidrug resistance by standard drug susceptibility testing often was a postmortem diagnosis in the early 1990s. It should not be so in this century in the United States. The fact that MDR and XDR strains of tuberculosis are rare in the United States now does not mean that they will stay that way in the next decade. The source of the next major

outbreak of MDR-TB may already be on an airplane on his or her way to the United States from Southeast Asia, Africa, or even Eastern Europe. To assume such strains won't spread in the United States will probably guarantee that they will.

The combination of tuberculosis detection and presumptive detection of rifampin resistance would be immensely valuable for tracking the spread of such important strains, especially when coupled with the rapid confirmatory testing by DNA sequence analysis provided by public health laboratories such as the CDC.

There are very good biological reasons why one-third of the world's population is still infected with tuberculosis even after four decades of trying to control and eliminate this disease. We may have sequenced the entire genome of *M. tuberculosis*, but we still don't fully understand how the organism spreads and causes disease.

The low rates of tuberculosis that we experience now in the United States are an anomaly when compared with disease rates around the world. This may be because clinicians often include tuberculosis in their differential diagnosis of respiratory tract infections, because we have effective drugs for the diagnosis to confirm --

DR. CALIENDO: Thank you, Dr. Tenover.

Our next speaker is Dr. James Pearson.

DR. PEARSON: Good morning. Thank you for the opportunity

for the Association of Public Health Laboratories to present to your Panel. We are the PHLs that Dr. Castro referred to in his talk. APHL represents the government laboratories that detect and monitor public health threats, including MTB, across the United States. That includes state, territorial, and local public health laboratories. We've already talked through the delays that happen with culture, and I'll kind of cut to the chase.

My disclosure. APHL paid for my travel to come from Richmond to here.

The APHL recommends the reclassification of molecular diagnostics for rapid detection of MTB complex, and the detection of mutations which confer antibiotic resistance to MTB complex, from Class III to Class II devices. And the failure to reclassify will continue to limit the availability of those approved methods for diagnosis of MTB, and it'll result in an increased risk for TB patients not getting appropriate therapy and an increased risk to the public health if those devices are not -- if the manufacturers of those devices are not offered the opportunity to present under Class II instead of under Class III.

The studies that I'm pointing out are actually studies that you're very familiar with. These are done in a public health laboratory in California. Reduction of turnaround time. In this particular study, the diagnosis came at least one week earlier in a study that was done in 2005 using nucleic acid amplification, one of the molecular assays. In 2010, there's

a rapid susceptibility testing study that was done by the Mayo Clinic and another one done in Switzerland that pointed to a change from 26 days to detecting resistance to six days. A rapid test is something that can provide information for the clinician.

The molecular assays are really intended to be a reliable screening method, but do not supplant the use of traditional culture. The laboratories' networks need the organisms to be able to be provide the genotyping and subsequently picking up clusters of infection, as was mentioned in Dr. Castro's presentation. We would like to see culture contamination part of any kind of licensure or product insert on those specimens.

Several of the states have reportable diseases requirements that the first isolate of any TB be sent to the public health laboratory for confirmation and for genotyping, but that's not uniform across the U.S. So getting an isolate to the public health laboratory system is critical.

There are some barriers. They've been clearly addressed, and I really don't need to go into any more detail on the technology restrictions that are in place.

There are some substantial cost savings, you know, with the more rapid diagnosis of tuberculosis, or better information to the clinician that the patient may have a TB and then a resistant TB. Prompt treatment leads to fewer complications, reduction in hospitalizations, and of course

restricts the isolation rooms, certainly a reduced need for contact investigations with more rapid diagnosis, and then a decreased likelihood of transmission.

I do have a couple of more things that I want to mention. While we do definitely need better diagnostic tools for tuberculosis, the delay is really inappropriate.

There is also a need for strong educational programs for the laboratorians using these molecular diagnostics, for the epidemiologists who are trying to interpret the results coming out of the laboratory, and then for the clinicians who are going to be utilizing new tools. And those educational steps really do have to be put in place.

And then a plug. But the public health laboratory system will be the ones that would bear the brunt to do most of this testing and the funding is not there. Thank you.

DR. CALIENDO: Okay, thank you.

Our final scheduled speaker is Ms. Calleja.

MS. CALLEJA: Good morning, I'm Khatereh Calleja. I'm Vice President of Technology and Regulatory Affairs at AdvaMed, and I'm here today representing AdvaMedDx. I have no financial conflicts of interest.

AdvaMedDx member companies produce advanced IVD tests that facilitate evidence-based medicine, improve quality of patient care, enable early detection of disease, and reduce overall healthcare costs.

Functioning as an association within AdvaMed, AdvaMedDx is the only multifaceted organization that deals exclusively with issues facing IVD companies both in the U.S. and abroad.

AdvaMedDx appreciates FDA's holding of this Panel meeting. AdvaMedDx supports consideration of reclassification of molecular diagnostics for the identification of *M. tuberculosis* and detection of genetic mutations that confer antibiotic resistance. We believe that these innovative technologies play an important role in supporting diagnosis of TB and drug-resistant TB in the U.S. and globally, and we commend FDA's efforts today to hear from the scientific community and consider reclassification.

TB is a significant global public health issue. According to 2009 WHO data, almost one-third of the world's population is infected with TB. Approximately 1.7 million people died from TB in 2009. The prevalence of TB in the U.S. is lower, with more than 11,500 TB cases reported, with approximately -- however, the rate of TB in foreign-born persons in the U.S. is approximately 11 times higher than U.S.-born persons, with approximately 59 percent of those cases having occurred in foreign-born persons.

Despite successful TB infection control processes in the U.S., given today's population mobility and the rate of world travel, the U.S. could still be susceptible to TB outbreaks. The emergence of drug-resistant TB also poses a significant threat to TB control and highlights the need to stop TB at earlier stages of infection. Diagnosis of drug-resistant TB is generally difficult

and can take up two to six weeks.

Nucleic acid amplification tests can provide results within 24 to 28 hours. Recently, the WHO endorsed the use of NAAT in the diagnosis of TB, and the CDC recommended that NAAT tests be performed on at least one respiratory specimen from each patient for signs and symptoms of suspected TB. These tests provide rapid results, while providing improvements in TB treatment. Such benefits include better infection control, opportunities for earlier, more directed therapy, more effective public health interventions, and greater opportunities for preventing additional drug resistance development.

AdvaMedDx has long supported a risk-based approach for the regulation of all clinical diagnostic tests. We believe such an approach will promote the public health, while supporting timely access to safe and effective diagnostics. Such a flexible approach takes into consideration the novelty of the anolyte, the clinical intended use, the novelty of the technology platform, and the experience of the user. It also considers the risk of illness, missed diagnosis, or delay if the test is not available, along with other mitigating factors, including scientific information, special controls, and consensus standards.

It should be noted that if a special controls document is contemplated to support reclassification, we are more than willing to provide input and assistance with development of such guidance.

In considering these tests, we note that association with identification of *M. tuberculosis* has been known. Tuberculosis has been known for decades. We also note that earlier methods for detecting TB are not subject to premarket approval applications. The relationship between certain genetic mutations and drug resistance has also been noted in considerable medical literature. Furthermore, NAATs for infectious organisms have been widely commercially available for over 20 years. FDA has cleared or approved a number of amplification tests in microbiology.

In closing, increased access to rapid tests will have a significant impact on the management of tests for TB. We applaud efforts to support timely access to new diagnostic technologies that will advance the public health, promote timely care and better targeted treatment. We appreciate the Panel's holding this meeting and thank you for the opportunity to comment. Thank you.

DR. CALIENDO: Thank you.

I'd now like to open up the public comment. If there's anyone in the audience who has not already spoken but would like to make a comment, you can come forward. I'll remind you that you're limited to five minutes. Please introduce yourself.

DR. CATANZARO: My name Antonino Catanzaro. I'm from the University of California, San Diego. I also have worked with Cellestis; I'm on the board of directors, the manufacturer of QuantiFERON.

And I just wanted to comment about what seems to me a catch-22 that is facing the community as a whole. Specifically, Dr. Castro has addressed the fact that, to eliminate tuberculosis in the U.S., we must make a very concerted effort to diagnose and treat latent TB before it becomes active TB.

Furthermore, as was pointed out by one of the Panel members, in order to do studies in the U.S., we require IRB approvals and basically those are not -- it's not possible to construct an appropriately controlled trial in the U.S. And I was particularly concerned with some of the comments that Dr. Gitterman made regarding the requirement for some studies to be done in the U.S. I just don't see how we can ever do those kinds of studies under the limitations, and the sticky problems that Dr. Burman pointed out in constructing such trials in the U.S.

I also wanted to address the comment that Freddie Poole made, that once a device was under a PMA, it's always under a PMA. And so any revision to the recommendations or requirements for usage would require a repeat PMA, and that seems to me to be a real obstacle to moving forward on this issue.

And the only other comment I want to make is that we're resorting to using -- to just looking at studies again, that Dr. Burman had referred to, where it's been recommended that the prophylaxis be used and it's not been used. This is not the best methodology for scientific studies.

However, the progression studies, all of which have been done outside the U.S., have demonstrated the usefulness of QuantiFERON results in identifying patients who are likely to go on to develop active TB, as well as the fact that people who have a negative QuantiFERON test are unlikely to develop active TB.

So I think that we need to make some accommodation for bringing these studies to inform the usage of these tests in the U.S., and I think it's a real catch-22 that is addressed in part by this reclassification. But I think, if I understand what folks have been saying correctly, it's not an item on the table to correct this imbalance, this concept about once a PMA, it's always a PMA, and I just wanted to introduce that concept. Thank you.

DR. CALIENDO: Thank you.

Is there anyone else in the audience that would like speak that has not already been to the podium?

(No response.)

DR. CALIENDO: Okay. Questions? Does the Panel have any questions for any of the speakers?

DR. TOWNS: Michael Towns, Industry.

Perhaps someone from the FDA would like to comment on the last comment about once a PMA, always a PMA because it was my understanding that once a reclassification does occur and it goes down to a Class II from Class III with special controls, that that is not the case, that it

remains a PMA forever.

DR. HOJVAT: That's correct. Sally Hojvat.

We did down-classify hepatitis A about four or five years ago, so they no longer have to send in annual reports and any changes are dealt with, so it's a 510(k).

And I just wanted to clarify, too. If you're adding a new specimen type, we don't usually go through an entirely new PMA. A supplement isn't sort of starting over again.

DR. CALIENDO: And if that would be the case -- Angie Caliendo. Sally, if that would be the case, if it was a 510(k) with special controls, it would be a supplement but not a full resubmission if they went to a different specimen type?

DR. HOJVAT: I'd have to check that with Freddie about a 510(k), but usually when you're adding something new. But it is as we pointed out, changing the intended use. So they have to come in with something, if in the intended use they state the specimen type, which most of them will in the microbiology area. But usually it's a separate study that's been done, even for a PMA.

For hepatitis, for example, adding on a claim for using neonatal specimens for hepatitis B, it's a concrete study that's allowed, and then that's sent in as a supplement. But it's not the full nine yards with, you know, the whole section on manufacturing because that's obviously not going to affect

just adding a different specimen type.

So it's often on a one-by-one basis, which we'll go through with the manufacturer. But it's not a full PMA for a change in specimen.

DR. RAND: Can I jump in? Ken Rand.

This is really a question of clarification from FDA. If the reclassification route is accepted and done, but somebody, a manufacturer, comes in with a completely new technology with an assay for TB, let's say, but they're using deep sequencing, which has never been approved for the clinical laboratory, would that be a Class III or a Class II?

DR. HOJVAT: Well, if it's something, you know, really different, it'll come in and we haven't seen it before, it's obviously going to take that yes, it's a Class III. Then we'll discuss any additional risks that could be there, and it could potentially be de novo down to a Class II.

I mean, we have been doing that, we have been discussing just what you said. What do we do with sequences? Because they're obviously out there on the horizon. So we have been internally discussing how would we regulate sequencing.

DR. RAND: So this wouldn't really preclude you from reverting, if you want to call it that, to a Class III if it seemed warranted?

DR. HOJVAT: Again, it's looking at are additional new risks brought in by that new technology?

DR. CALIENDO: Lizzie.

DR. HARRELL: Lizzie Harrell, Duke University.

For Dr. Miller and the ASM recommendations. I was just curious to know, for the nucleic acid amplification for resistance, was the reluctance to move -- bring that one in at a Class II because of the infrequent resistance in the United States or because it was a new test, a new test that we really haven't been using a lot?

DR. MILLER: It was due to the low positive predictive value in the U.S.

DR. HARRELL: Okay, thank you.

DR. CALIENDO: Melissa, I'm sorry. Are you familiar with data that's available for the use of the tests in the U.S.?

DR. MILLER: Only clinical laboratory data, and they really looked at the TB portion of it, not large numbers of the rifampin portion.

DR. CALIENDO: Okay, thanks.

DR. MILLER: Yeah.

DR. CALIENDO: Any other questions for Melissa while she's up there? Go ahead.

DR. DELLA-LATTA: Yeah, hi, Melissa.

I was just wondering how you would feel if INHA and CAT G had been added because INH is obviously the most resistance, monoresistance that we see here in the states?

DR. MILLER: It would be the same issue, although certainly the

sensitivity is less in terms of what we know for molecular in detecting INH. And now, like we haven't discussed this specifically in the ASM committee, so I'm now I'm speaking as Melissa Miller. But in terms of the sensitivity of looking at those genes for INH, it's not as sensitive as the rifampin-determining region that we know of, so it wouldn't change my thought at all.

DR. DELLA-LATTA: Yeah, you need more SNPs.

DR. MILLER: Yeah.

DR. DELLA-LATTA: I can see that the problem of how many mutations and how many you add is really critical here.

DR. MILLER: Right.

DR. DELLA-LATTA: So it would be very difficult, for example, with PZA --

DR. MILLER: Absolutely.

DR. DELLA-LATTA: -- for obvious reasons. Okay.

DR. HENRICKSON: Kelly Henrickson. Another question for you.

But the negative predictive value is great. And so, clinically, why can't we use an algorithm and why can't we use it as a negative predictor and then send it on to the CDC? I mean, clinically, you can handle this.

DR. MILLER: Sure. But, again, speaking as Melissa Miller, just to be clear, that needs to be part of the special considerations that we think about in bringing these tests forward because not every laboratory that may be able to do a very easy NAAT test is going to have these algorithms and the,

you know, capability of the interaction with their clinicians to make these things happen.

DR. TOWNS: Michael Towns, Industry.

Not specifically addressed to you, but I think you included this terminology, as well as another presenter in the public discussion, that I want to emphasize, that a couple of people have mentioned and I believe you did mention the use of direct molecular tests for ID'ing MTB as a screening method. And I want to make the statement that these are not screening methods. These are diagnostic methods. Because when you're talking about screening tests, you're applying that to an asymptomatic patient population and that is not how this test -- these tests would be intended to be used. They would be used for a diagnostic purpose, not for screening. Thank you.

DR. MILLER: Just to clarify, I see your point, certainly. Perhaps what we intended more is not as a standalone test.

DR. CALIENDO: Yes.

DR. GORDIN: Fred Gordin from the VA.

I didn't know if Dr. Castro wanted to respond. I kind of took exception to Dr. Tenover's statement here that the CDC has mandated that each patient suspected of having TB basically have a nucleic acid amplification. I mean, we have the paper right here in front of us, one of their guidelines, they're not mandates, and it says it should be done for whom the test result would alter case management or TB control activities,

which is far from mandating it for every patient suspected of having TB. So I don't know, Dr. Castro, if you're interested in responding; otherwise we'll just leave it at my comment.

DR. CASTRO: We'll leave it at your comment.

DR. GORDIN: Okay.

(Laughter.)

DR. CALIENDO: Any other questions?

(No response.)

DR. CALIENDO: Okay, great. So we'll now break for lunch. I want to remind Panel members, please do not discuss the meeting topic during lunch amongst yourselves or with any other member of the audience. We're going to reconvene in this room at 1:10.

Please take any personal belongings you may want with you at this time. This room will be secured by the FDA staff during the lunch break. You will not be allowed back into the room until we reconvene at 1:10. Thank you.

(Whereupon, a lunch recess was taken.)

A F T E R N O O N S E S S I O N

(1:15 p.m.)

DR. CALIENDO: Okay. So welcome back. I hope you had a lovely lunch.

At this time we're going to focus our discussion on the FDA questions. Copies of these questions are in your folders. So I want to remind the Panel members, whenever you speak, each time you speak, to please give your name so that the transcriber has that information.

So Dr. Gitterman is going to start by reading the first question for us.

DR. GITTERMAN: Thank you. I'll start with the preamble to all three questions, then read the first one.

The reclassification process for down-classifying a device from Class lll to Class ll depends on being able to mitigate the risks associated with the use of the diagnostic device such that there is reasonable assurance of safe and effective use. Similar considerations exist for deciding whether a new device is appropriate for the de novo regulatory pathway, that is, whether Class ll designation is appropriate at the time an application is filed for a new device with a new intended use. As described earlier, in both instances FDA is mandated to publish a special controls guidance that outlines the steps that a sponsor should follow for developing a new diagnostic device for the specified intended use. In this context, please

discuss the following:

Question 1: For nucleic acid amplification assays that detect *Mycobacterium tuberculosis* complex directly from respiratory samples:

a) Please discuss the risks associated with inaccurate results for *Mycobacterium tuberculosis* complex detection (that is, false positive and false negative results).

b) Please discuss the minimum device performance standards (that is, the sensitivity and specificity) that should be recommended for these tests. In your discussion, please comment on the appropriate reference method for clinical studies and the role of both prospective and archived/banked patient samples in demonstrating device performance.

c) Based on the above considerations and the earlier presentations, please discuss if sufficient risk mitigation is possible for FDA to initiate the reclassification process from Class III to Class II devices for this use through drafting of a Special Controls Guidance. In your discussion, please note if there are any specific special controls in addition to those previously mentioned that should be considered for these devices.

Thank you.

DR. CALIENDO: So I'd like to hear from everyone from the Panel, but I think it's best to just start with people making comments. And then if we get to the point -- we have about three hours to answer all three of these questions. So if we get to the point where we're getting short on time for any of the questions and people haven't had a chance to speak, we'll go around the table and give everybody an opportunity to speak.

I think it might be more efficient to clump (a) and (b) together, so maybe we could discuss those two together initially and then we'll move on to (c). And then from there we'll move on to the second question.

I know that, Karen, you had a question that you wanted clarified.

DR. CARROLL: Sure. Karen Carroll, Johns Hopkins.

I just wanted to clarify. With respect to the first item, nucleic acid amplification tests, many of us have had well over 15 years experience with these assays. Do you want us to answer these questions, (a) and (b) in particular, in the context of our experience and our knowledge of risk mitigation by following the CDC recommendations and the package insert of the manufacturer, or do you want us to speak in absolute terms? Just Number 1 in particular.

DR. GITTERMAN: All of the above. I'm not being facetious. It obviously should reflect your experience. I mean, I am hard pressed to state,

you know, what an excellent Panel we have, and the reason, of course, people have been invited is because of your particular expertise. So it really is all of the above, all the considerations you think is relevant directly for this question.

Does that answer it?

DR. CARROLL: Yes.

DR. GITTERMAN: Thank you.

DR. CARROLL: Thank you.

DR. CALIENDO: Okay. So we're open for discussion.

DR. WILSON: Mike Wilson from Denver Health.

In terms of the risks associated with inaccurate test results, false positive and false negative, I think we all know what those are. False positive people get treated who don't need treatment. False negative, you know, there's the risk of transmission from untreated patients.

But I think, irrespective of what our recommendation is today, it's important to remember that those conditions exist today. I mean, we have imperfect tests. We're using both culture smears, the nucleic acid amplification test.

So whether we reclassify these devices or not, it isn't going to change very much what's currently going on. So it's difficult to say what the outcome would be if this reclassification occurred because, to some extent, we probably aren't going to change what's the existing situation. So in a way,

answering that question is a difficult one.

And the same thing in terms of the sensitivity and specificity, the performance characteristics. I mean, we're currently using very imperfect tests already. We're actually not suggesting that we would eliminate those tests. This would be an additional test. Nobody's suggesting that we would ever eliminate cultures or traditional susceptibility testing.

So in a way, you can't -- there isn't a pre- and post-scenario. It's not like this is a -- we're discussing tests for a brand-new disease that nobody's been dealing with for a long time. The context for this question is a little bit difficult to handle because we're really not talking about changing much in terms of the risk to the patients.

DR. CALIENDO: Angie Caliendo.

So, Steve, I wonder if you can clarify something. So if I can rephrase what you said, you know, we're so used to comparing or assessing the perfect test and assessing the outcome versus a perfect test, or this test versus a perfect test. To me, the risk isn't just comparison to a perfect test, which we're never going to have, but I think it's also a consideration of if this test is not available.

So it's no test available and what's the risk to the public or the individual patient if you don't have the availability of the test at all? Is that fair game to include in this? Because I think that's what you're kind of getting at in some ways.

DR. GITTERMAN: Yeah. Let me perhaps clarify this slightly. But your point, Dr. Caliendo, I'm going to ask Dr. Hojvat to address.

What it really means is a test has risks and benefits associated with it. But focusing on the risks, of course, there's a risk, exactly as Dr. Wilson stated, of false positives and false negatives. Those have associated risks with that.

I could see a committee saying there are tremendous risks. You know, let's say it was a life-threatening risk of a false positive. Well, people are going to view a test like that very, very differently than they would from their professional background and the experience they'd had with other tests, to say perhaps it's less of a medical risk. So when it talks about the risks in a case like this, it's really specific to the tests and the way the tests would likely be used.

Now, your perspective, of course -- you know, again, let me defer before I comment. Dr. Hojvat perhaps would want to address the latter half of your question.

DR. HOJVAT: You brought up availability, and that's what the crux of this is, and you've heard it from other people, that people would tend to send in a 510(k) rather than a PMA. So availability of new tests is, I think, what's exactly what you're saying, that if we lowered the bar and made them 510(k)s, we'd have more tests coming in here for us to review.

DR. CALIENDO: Angie Caliendo.

And it's okay to consider that in our discussions?

DR. HOJVAT: You know, we haven't gone to the -- I mean, that is the basics of this. It's come up, it's in the report that we gave you, so we did give you that information.

DR. CALIENDO: Okay. Because, in many ways, that's how I see it with some of this, is there's a risk of an imperfect test. There's a risk of not having anything available and to balance this.

DR. HENRICKSON: Kelly Henrickson, MCW.

In pediatrics, as has been mentioned, these direct PCR tests on direct specimens are not going to -- the risk of false positive is not going to be as great because we don't have materials that are going to be positive very often. So another way to say that is that symptomatic children with TB often, again, do not produce sputum, do not produce mycobacterium in quantities that have up to now at least been detectable frequently. So we frequently have children who diagnose clinically to have TB and have no positive results.

And so I think that it's critical that we include in these standards, if we do this, I think they need to do testing in children. Okay. So this is a plea for them to -- especially children under five. I think that I would suggest that when we discuss (b), about standards, that we ask for them to show that these assays work on the specimens we use in children and that they have reasonable sensitivities and specificities.

So if we use, like what was recommended, the sensitivity of 90

percent for AFB positive smears and 60 percent for smear negative, well, most are going to be smear negative or, you know, not even tested in children. I think that we should require that they do tests in children and show that they have sensitivities of 60 percent in children and the same 99 percent specificity that one of the speakers recommended. And so I would put that plea out there.

I think that we also would all agree that really the gold standard that has been brought up multiple times for TB is that we want to have an assay that predicts whether you get TB in your lifetime, right? So if you are infected with TB, most of the time you won't immediately have disease, and then if you're under one, you have a 40-percent chance of developing disease quickly, if you're under five, maybe a 20, 25-percent chance, and by the time you're over five, it's down to 5 or 10 percent over years, as we all know.

Really the perfect gold standard is whether we could follow people long enough that we can say, you know, this test predicts that you will never get -- you know, you're latently infected but you'll never get disease. And that really is the gold standard. Now, that's ridiculous and we'll never be able to do that, but that's really what the gold standard should be, is predicting whether in your lifetime you get sick from TB.

DR. CALIENDO: Right, but I think you're talking about the IGRA assays --

DR. HENRICKSON: Right.

DR. CALIENDO: -- and we're not quite there yet. So we just want to stay where we are, is in direct NAATs.

DR. HENRICKSON: The reason I'm bringing it up is because, in pediatrics, we have to use indirect tests more often than in adult medicine. Okay. So in my career and my experience with pediatric TB, the direct tests have never been as useful as the PPDs and the indirect tests. And so we have to use those in patients who are sick, who we are trying to actively diagnosis active TB in. And so that's how -- so those tests are the tests we use when a child presents with pulmonary TB. We have to use indirect tests, and so I think those tests have to be part of any algorithm that's used.

So I would suggest that, I mean, if we're talking about algorithms, we have to use direct testing and indirect testing in pediatrics, at least, if we're going to come up with some kind of algorithm with children who have pulmonary TB or non-pulmonary TB.

DR. DELLA-LATTA: Yes, I absolutely agree with you. One point I do want to bring up, and that's in terms of these special control guidelines that we would have to consider with NAAT testing.

DR. CALIENDO: Okay, now we're on (a) and (b) right now, so we want a summarization here.

DR. DELLA-LATTA: We're on (a) and (b).

DR. CALIENDO: But go ahead, go ahead.

DR. DELLA-LATTA: Yes, I understand that. What I'm saying really is that there is that need, and because of that we're going to have different types of specimens. So for us to have a sensitivity and specificity breakpoint overall is something that we really cannot have. If there's 60 percent, I believe it was brought up as a sensitivity for smear negatives, that the ASM -- Melissa had brought up -- I think it's kind of low. If you're only considering one specimen per patient, it may be fine, but I don't think we should. So I think the number of specimens should be considered with the type of specimen and the type of tuberculosis we're thinking of.

On the other hand, if we're thinking of pediatrics with gastrics, in my lab, in my experience, we have 60 percent sensitivity with gastrics. And with the technology, that's with the MTD that I'm using, I think we have to evaluate sometimes what we already have and expand on that and have a different specialty control, in other words, different sensitivities and specificities. A very high specificity, yes. On lymph nodes that we do, again, because you do see extrapulmonary disease, it's 50 percent. But you know what? Fifty percent, that's diagnosing 50 percent of cases, is better than nothing.

DR. CALIENDO: Phyllis, give us an idea. I get the idea that you want -- you have different performance characteristics for different assessment types.

DR. DELLA-LATTA: Exactly.

DR. CALIENDO: So let's start in your mind with a respiratory specimen.

DR. DELLA-LATTA: Yes.

DR. CALIENDO: What would you think is acceptable performance criteria? And then let's go to --

DR. DELLA-LATTA: Yes.

DR. CALIENDO: -- a non-respiratory specimen.

DR. DELLA-LATTA: Fine. Respiratory, in my lab, is 85 to 90 percent, if not more. That's minimum.

DR. CALIENDO: On one specimen or three?

DR. DELLA-LATTA: No, I will never do one specimen.

DR. CALIENDO: Okay. So on your group of three.

DR. DELLA-LATTA: So three smear negatives, it's at least three. It's upon consult with pulmonary or ID. So it's an extremely at-risk patient. And my sensitivity has been over 90 percent.

DR. CALIENDO: Over 90 percent.

DR. DELLA-LATTA: Yes.

DR. CALIENDO: Three specimens smear positive.

DR. DELLA-LATTA: Three specimens smear positive. I will not accept one specimen. If I get one specimen, I tell them I will not do the test.

DR. CALIENDO: Okay.

DR. DELLA-LATTA: I speak with pulmonologist and the ID. I

don't think we're utilizing the tests properly.

DR. CALIENDO: Okay, smear negative?

DR. DELLA-LATTA: So that's smear negative. Smear positive is 100 percent. I never have a problem. I never have a problem.

DR. CALIENDO: I'm sorry, you said smear positive.

DR. DELLA-LATTA: Oh, I'm so sorry.

DR. CALIENDO: That's smear negative.

DR. DELLA-LATTA: That's smear negative.

DR. CALIENDO: Okay.

DR. DELLA-LATTA: Amplification test, yes.

DR. CALIENDO: Okay.

DR. DELLA-LATTA: SNPs.

DR. CALIENDO: Now, what would you say if this was a non-respiratory specimen?

DR. DELLA-LATTA: Non-respiratory specimens, some don't work at all in my hands. And unfortunately it's pleural. It's full of inhibitors by the MTD. It's peritoneal. CSF is iffy, but we've diagnosed TB meningitis. So when it's positive, it's positive. When it's negative, it's not believable. But you know what? I'd rather pick up those cases of meningitis when I can. And this is upon consult. So it's not that we're, you know, abusing resources and doing it on everybody.

I think it's extremely important, when we're thinking of any

tests with the limitations that they're going to have, to optimize it and you optimize it by using various criteria that bring clinicians into the picture and the quality of specimens. Sometimes you will get one sputum specimen, so-called sputum specimen, right? And when I ask for more, I'm told the patient cannot produce any more sputum. So what is that telling me? So when I speak to the pulmonologist, the chest X-ray, the patient is not giving productive sputum. So if they really considering TB, I get a bronch or I get induced sputum. And this way, the quality of the specimen is better and my numbers are very, very high.

So I would suggest that we have to consider the performance characteristics based on the specimen, which is, of course, in consideration of extrapulmonary and pulmonary TB. We have not done that in the past.

DR. GINSBERG: Dr. Caliendo, just for clarification either from you or from the Agency. Question 1 and 2, maybe 3 as well, certainly -- no, sorry. Questions 1 and 2 relate directly to respiratory specimens as written. Should we be limiting our discussion to that, or do you want us to answer these questions for multiple kinds of specimens?

DR. HENRICKSON: Is it respiratory specimens or pulmonary TB? I know, but I mean, shouldn't we be considering pulmonary TB? Because, again, respiratory specimens aren't useful in pediatrics, or haven't been so far.

DR. GITTERMAN: That's a very good distinction, actually.

Pulmonary TB would be more correct. I would say our perspective is this is a tremendous amount to bite off in one meeting in the afternoon, and we actually very consciously tried to restrict it to pulmonary TB at this point because it is such a difficult issue, and as we all recognize, extrapulmonary is -- if I am quoting you correctly, Dr. Latta, you had just mentioned that any case, in essence, that had extrapulmonary, even though you did describe off-label uses, which I should note, which I'm sure you're not advocating at an FDA meeting.

(Laughter.)

DR. CALIENDO: I'm sure she's validating.

DR. GITTERMAN: No, no, I'm joking, I'm joking.

DR. DELLA-LATTA: I'm sure I am.

(Laughter.)

DR. GITTERMAN: But no, we thought that this would be a tremendous amount and there's a lot of variation. So, again, if -- and I speak for Dr. Caliendo -- perhaps if we could resolve the pulmonary issue, we'll have accomplished a tremendous amount.

DR. HAMMERSCHLAG: Maggie Hammerschlag.

Again just to clarify, we use the gastric aspirates. Circuit's a bad word, but it's basically, in this case, not a pulmonary specimen that is being used to diagnose pulmonary disease, the theory being these kids are swallowing the stuff because they don't expectorate. But when we're getting

involved in handling children, we usually use epidemiologic exposures as a really important context. You know, if you've got a child with active TB, there's an adult around or a converter with active tuberculosis. So we may have a much lower sensitivity in terms of starting them on therapy, presumptively because of the problems of the higher risks for disease.

But the other point about multiple tests is that these young children are also the ones who, even with active TB, are likely to be negative by tuberculin tests and negative by IGRA. So it becomes, I think, being able to have a direct test that you could use on that gastric aspirate faster would be a real achievement. It would be really helpful.

But I don't know if we have sufficient data at this point to talk about it. I would tend to believe, if we had a positive, I would trust it, probably, by a direct -- by amplification, but I don't know how I would interpret a negative.

DR. GORDIN: This may be back in the clarification group again. So the risk/benefit of the tests would be partly -- it was stated a few minutes ago. This test would never be done alone. We'd always be doing cultures. But actually, hopefully, that it isn't the case and that actually we'll have some technology that, you know, is just it.

So for the discussion today, in terms of risk/benefit, are we assuming this is out there and the labeling will say continue to do whatever, whatever? Or is this going to be viewed potentially by the FDA, in terms of

indication, as the -- you know, with clinical, the diagnostic test? In which case, the risk/benefits change markedly.

DR. GITTERMAN: Absolutely. We would listen to the discussion, Dr. Gordin. In the special controls guidance, as I mentioned before, we would have recommendations on labeling, and the labeling -- and again, I'm assuming everybody recalls that there's a section of intended use, and based on some of what we hear today, we would -- you know, conceivably, one might put in the recommendations for intended use, is that this is an aid in the diagnosis of pulmonary TB, you know, pulmonary tuberculosis. Culturing in all cases should be done and you could embellish that, whether you say it's for genotypic, et cetera, et cetera.

But, again, if that is the feeling, you know, based on, as people discussed with the likely sensitivity and specificity, that it should be recommended in the special controls guidance that the intended use for this device always recommends it be done in conjunction with a -- you know, with culture, then we would take those -- you know, we would certainly take those recommendations strongly.

I would also mention, too, which actually -- well, I won't mention it. Thank you.

(Laughter.)

DR. GORDIN: And then also it specifically says, you know, nucleic acid amplification assays. Obviously somebody could come along with

some other rapid molecular diagnostic that's not the same technique. So is this whole discussion specifically about kind of the technology that you've described here, or is this whole discussion about rapid diagnostics and products that might come along in the next few years or whatever?

DR. GITTERMAN: That's a fair question. Let me answer it generally and perhaps Dr. Hojvat might want to add.

In a sense, specifically, what we're really talking about are molecular diagnostics at this point and a special control, depending again -- I do not want to be presumptive, but if there were, you know, recommendations to, in fact, approach it, that's Class II and a special controls guidance would be written, it would be relatively specific to molecular diagnostics. The type of analytical studies and the recommendations would be there.

If someone then came along with a device for the same intended use, depending on the perspective of the technology, whether it posed different risks to health such that special controls guidance was no longer applicable, then that might be considered separately as a Class III, or there's other options as well. But, in general, what you'd be describing would be special controls for molecular diagnostics. But, again, as we often say in other contexts, the guidance in there is applicable for -- you know, is broader than just for that narrow indication.

So the answer is yes and no, but it's mostly the -- the special

guidance that were written would likely be mostly specific to a molecular test, a molecular NAAT assay.

DR. JANOSKY: Janine Janosky. Questions and comments for both Dr. Gitterman and Dr. Castro.

You know, we heard information today talking about clinical medicine and clinical practice and the use of these diagnostic tests in that context, and then the implication of the use of those diagnostic tests for public health. So I applaud the two different perspectives. That was actually a very unique and innovative way to look at what our decisions might impact and how they might impact.

But I'd like us to return back to the questions that you have for us for Number 1(a) and (b). When we're talking about safety and effectiveness within this context, are we being asked as a Panel to respond to safety and effectiveness from a clinical medicine and clinical practice perspective, absent public health, or clinical medicine and clinical practice in the context of public health? Because, given your answer, my response to 1(a) and 1(b) would be quite different.

DR. GITTERMAN: I think you're asking it correctly. Mostly we mean to refer specifically for the test used in an individual. What would the risks to your health be from a positive and a negative test? But I think you're approaching it exactly correctly, there are broader risks, in the same way there's risks to the laboratory worker from possible exposure.

So we mean mostly the individual risks, as Dr. Wilson started off, of a positive for that implication for that individual. However, you're asking exactly the right question. This is a type of device where there are implications beyond the individual and that question has to be asked. But the focus should mostly be the individual. But I will allow Ken to address it. And it's a very good question. Thank you.

DR. CASTRO: The only thing I would add is I see that as a continuum rather than mutually exclusive categories. You know, as I pointed out, the ability to accurately and rapidly identify people with tuberculosis and get them on treatment does wonders for the individual as well as for the community. And I think we were invited to bring the public health perspective so that it's also informing the considerations by this Panel, rather than, as it usually tends to be, absent from the discussions.

DR. CALIENDO: So Janine, with that information, do you want to give us your thoughts?

DR. JANOSKY: Sure, absolutely. You know, I'm stuck here with a quandary -- and this is Janine Janosky again. I'm stuck here with a quandary because I'm thinking, you know, summing over numerous studies, if we're looking at sensitivity and specificity for these diagnostic tests, if we use them against what we consider to be our epitome of what is typically sensitivity and specificity values for diagnostic tests, one would argue that they fall short. But I'm stuck saying, but what's the alternative? And if I say, okay,

these will be fine, then if we come in with another 510(k) application, I'm also concerned that the next product coming for substantial equivalence will be left to that standard.

Given all of that, you know, we always have to weigh things and always have to think of things, you know, my concern is not moving forward and not reclassifying these as type II or class level II would probably disadvantage the use from a public health perspective.

DR. CALIENDO: Sorry, this is Angie Caliendo.

Do you have any sense of what performance characteristics you would recommend to the FDA?

DR. JANOSKY: Yeah, absolutely. Janine Janosky again.

You know, I would like to see the values that we typically see for diagnostic tests, in the 80s and in the 90s. I do understand that the current tests are not meeting that level or do not seem to be meeting that level. But, again, in the 80s and the 90s.

DR. RAND: Ken Rand.

I mean, the sensitivities they're quoting here are all, you know, way up there, and yeah, maybe in smear negative they're not as high as you'd like. But I mean, you're not saying that the current level of both molecular -- they're not as good as we would like them, but if we could make them better, we would. So I'm not sure, you know, what exactly you're asking of the manufacturers, given that, you know, it's an imperfect world to begin with.

DR. CALIENDO: Ken, why don't you comment on what you think is appropriate sensitivity and specificity.

DR. RAND: In a word, they've got to be, in my view, at least comparable to what we have. Hopefully better. If they're not comparable, how many are you going to sell? You know, I wouldn't buy it for my lab.

DR. HARRELL: Lizzie Harrell, Duke University.

I wanted to go back to something that Dr. Pearson commented on. When he made his presentation he talked about the need for education, and that gets into what Dr. Wilson said when he talked about the context of these tests. We're not going to be doing this as standalone tests. They are part of an algorithm of things we're going to do for the patient suspected of having tuberculosis.

And because of that, I think, if we tried to get the sensitivity and specificity where we'd like to have it, we may end up doing some harm to some of those patients that would be diagnosed earlier if we had these tests. They're not where we want them. Hopefully, as we get through newer generations, they're going to improve on this. But if we keep them at Class III and not move them to Class II, I think it's going to prohibit others coming in and trying to even improve it because they don't want to go that route.

So my comment is that, as Dr. Wilson mentioned, what were the comments about the false positive and false negative, how it's going to affect patients, I agree with that, but I don't have a particular number that I

can put down as far as sensitivity and specificity. But I would say that I think, if we try to get it into the 90s, we may end up defeating the purpose we're here for today.

DR. CALIENDO: This is Angie Caliendo.

Lizzie, do you have any sense of the number of specimens that you would recommend and what that would do for your thoughts on sensitivity?

DR. HARRELL: Not right off. I'll have to think about that. Give me a minute. Thanks.

DR. HAMMERSCHLAG: May I? I think, you know, you have to -- it's very obvious. The thing that was surprising to me is that the tests are not more sensitive and it may have something to do particularly with tuberculosis and possibly, you know, as opposed to what we have, let's say, for other NAATs that are out there, specifically chlamydia and gonorrhea, where the performance parameters are extremely high, very well defined and, you know, like they're well over 90 percent.

I mean, the CDC has recommendations for what they will expect, but I don't think that would even be currently achievable. But then you're dealing obviously with gonorrhea and chlamydia with a different set of public health and social and then sometimes legal implications. So I think we have what we have and it may offer definitely an improvement over culture under certain circumstances. So I think that it can't be a one size fits all. We

have to take it into account of what we're dealing with.

But I think one problem is that, you know, we're in Brooklyn and we're dealing with a relatively high-risk population because we have a lot of immigrants, a lot of Haitian immigrants, and for us it always gets a little antenna going about when we're considering. The concern is when you start getting into other populations where the prevalence is going to be much lower, then we're going to run into, maybe, the issues with the specificity and the negative predictive values.

So there may need to be more education as to when it's appropriate, certainly thinking of epidemiologic exposures, to educate people when testing is appropriate and when things should be sent on for further testing, to central laboratories, because the other issue is what's also going to happen.

And there are a lot of places where this stuff is being -- many hospitals, et cetera, are totally outsourcing their microbiology services to commercial laboratories, which have different standards and certainly a different set of criteria. For what's important to them, I think, you know, accuracy is secondary. They're concerned about throughput. And so then we may be in a whole set of other problems.

DR. DELLA-LATTA: I just want to bring out that point once more, in terms of optimizing the tests that we do have now and assuming that what's out there in the literature, and even the package insert, often is

based on one sputum specimen. How can we address this? Because we're all back to the quality of the specimen. And I think we ought to make some claim, and if people don't have enough data, maybe there isn't data out there -- I know I have tons -- that really optimize the sensitivity. I mean, it bumps it up a good 20 percent if you have three specimens and you get three specimens. It's recommended that you have three for culture. So if you have three for culture, you get three to do the amplification test, the NAAT. It doesn't matter if it's one per day.

We don't do that early morning nonsense that's about 50 years old. So whenever a person expectorates, that's the specimen. Sometimes you can get it in two days or one day and get three specimens. So why can't we just apply what we have for culture, a number of specimens, to that NAAT? And if you just really think about it logically, one out of those three has to be pretty good. So I made the comment before, but I don't know if it can't be addressed or won't be addressed or people don't wish to address it.

DR. CALIENDO: No. I mean, I think we're hearing you.

DR. DELLA-LATTA: Okay.

DR. HAMILTON: So I want to get back to a couple of things. One, the emphasis here, I think, should be on, you know, making high-quality diagnostics available. So I know a lot of your discussion has to do with a patient who is in a hospital, in a hospital that has a lot of resources, has somebody who can do bronchoscopy, has somebody who's going to make

sure somebody submits three specimens. But if you're talking about even rural North Carolina, there's no pulmonologist. There's a hospital that doesn't have any kind of microbacteriology expertise at all.

And so we're talking about someone who might come in to an emergency room or urgent care and it's either -- first of all, somebody has to even think about it, which is the big problem, and this will not solve that, but someone has to think about it, order the test, and then if that test has to trundle off to the state lab or some other place and it's five to seven days and the person is traveling, you know, there are those problems. So if we have a test that has similar test characteristics to what we have out there, but it's really easy and it's right there and it's something that can be done rapidly, I think that will be tremendously helpful.

Or think about situations such as the people who are coming in to jails and prisons, which are two very different situations, but let's say jails, where people are only there for maybe 24 to 48 hours. But before putting them in a holding cell with 100 other people, they're coughing, et cetera, there's some clinical reason and you want to check it out.

So, anyway, my point is just that I think, you know, we can talk about the ideal, but then there's really the public health piece that I think needs to be really considered with these tests.

The only other thing I wanted to say is, while I absolutely believe that we can mitigate these problems by the controls we can come up

with, you know, there are a number of risks associated with inaccurate tests. So 1(a) is supposed to be, have we thought about all of the risks associated with this?

So in terms of false positive tests, you know, we've talked about the fact of unnecessary treatment and the contact investigations, et cetera, but another problem with false positive tests is that pursuing an alternative diagnosis stops. So if you have a patient who comes in and you have a positive smear and you don't know if it's TB or not, everybody just kind of sits there and treats them for TB, keeps them in isolation, and meanwhile their other process is, you know, the cancer that has allowed them to now show their MAC is not being diagnosed. So that's another thing.

If you were in another country, another problem with false positives is that a supposed diagnosis of TB equals HIV, either in terms of stigma, so there are stigma-related issues, not as much in the U.S. but certainly in other countries. In terms of the problem with false negative tests, of course, if somebody is HIV, if you have a false negative test, then you're not going to have that signal, that trigger, to start doing things like testing for HIV, getting started on antiretroviral therapy, et cetera.

So just to flesh out a little bit the issues with false positive and false negative. But absolutely these can all be mitigated with the controls, I believe.

DR. GINSBERG: Ann Ginsberg.

So I just wanted to get back to this idea of sort of not letting the perfect be the enemy of the good here. And I'm particularly concerned about the sensitivity issue in TB and how we're defining it because, you know, I'm not a laboratorian and I run clinical trials and certainly, in the context of clinical trials, it's not even clear that liquid culture, for example, is the gold standard, as opposed to solid culture, which is less sensitive.

But there are clearly well-documented patients, not infrequently, who have absolutely established diagnosis of TB at the beginning and then are followed serially for the next one to two years as they go through treatment and then off treatment to see whether they relapse. And in that context, they're not infrequently patients who, on liquid culture, become positive again, long after treatment. No clinical symptoms and not a positive solid culture.

So are those patients really positive or not? Probably not. Most clinicians would not treat that patient. You continue to follow them and see if they become clinically symptomatic again, and many of them never do.

So, you know, if we try to define sensitivity as, you know, well over 90 percent against a liquid culture standard, for example, we may be perpetuating, in essence, a mistake we're already making. And, you know, where it's hard to define clinical truth, which really should be the gold standard in TB, I think we have to be really careful in how we think about

sensitivity.

DR. CALIENDO: Karen, why don't you give us your thoughts?

DR. CARROLL: So I mean, I think I agree with the comment that it's really hard to expect these tests to be super-performers when our existing laboratory gold standards really are suboptimal. That said, I think this is an opportunity to invite better tests with better sensitivity. So I do like trying to establish some sort of threshold compared to our imperfect gold standard.

So for smear positives, where I think we have the opportunity to make the greatest impact in terms of treatment, prevention of transmission to other people, and infection control, I would really like to see a sensitivity above 90 percent, and I think that's achievable with the existing technologies.

I think the smear negative is a lot more difficult because, I think, in many cases we're dealing with paucibacillary disease that may be below the limit of detection of existing molecular tests. And also, as someone on the other side pointed out, we're in a situation in the United States where many people are sending their specimens off site. We're not looking well at transit issues. We're not looking at what happens to these organisms when they sit out in a summer day like today, or actually like yesterday. What's really happening with the ability to recover paucibacillary infections?

So I think it's more complicated than that, and then trying to

establish thresholds for their performance just in the context of comparing them to our existing laboratory gold standards.

DR. CALIENDO: This is Angie Caliendo.

Karen, let me ask you to clarify. When you say 90 percent for smear positive, let's get back to the point that Phyllis is making. Do you see that off one specimen? Do you see that off multiple specimens? This would be smear positive.

DR. CARROLL: Certainly in our case -- and I think I would echo some of what Phyllis said -- if we have a patient in with a very high index of suspicion who's smear positive and the first test comes back negative, we will send a second specimen and we will have much more confidence. So this would be 90 percent in the context of that patient, not per specimen.

DR. CALIENDO: JoEllen, go ahead.

MS. DeLUCA: JoEllen DeLuca, Patient Representative.

I think when I hear -- I'm not in your world, in your labs, I don't go to your labs. I've got a lot of mouse labs and tissue labs for cancer. But I think the goal here is to get the United States back to no TB or TB is gone. We don't need TB sans; we don't need TB testing.

So in that regard, when I hear about HIV, a lot of people have HIV testing at fairs. You see that at health fairs and things, they set up little booths and things to do that free for the public. Gonorrhea, I think, people start -- sexually transmitted diseases often go to the doctor's office and free

clinics. TB is something nobody wants to say they have or think they have and often don't even feel they have. So there's really no interface for a lot of the people that we're trying to reach and need to reach more.

So increasing the number of patients seen, these are people that aren't -- we've got a whole group in North Carolina, South Carolina, where people never have been to a doctor, never have been to a doctor, never have been to a healthcare professional or a nurse. Parish nurses are wonderful, and it's increasing effectively. I think the more tests we have that are easier to use, I think, the faster we're going to gain momentum on this disease.

DR. CALIENDO: Go ahead, Mike.

DR. TOWNS: Michael Towns, Industry Rep.

I'd just like to start out by saying that I concur with the comments made by the AdvaMed representative here, stating that industry is in favor of the FDA considering reclassification of these devices from Class III to Class II with special controls because we feel that the special controls will provide the appropriate level for risk mitigation.

Now, as we all know, you know, a diagnostic test really should be used in the appropriate patient population. So with these molecular direct detection tests for MTB, you need to be using the appropriate patient population, those that have at least a moderate to high suspicion of active pulmonary TB. You don't want to be using them on patients that have very

low risk of having MTB because even if you have a very high specificity, even in that situation you may be dealing with a significant number of false positives as well.

And you need to make sure -- you know, clinicians, when they order these assays, they need to have an idea of what they're going to do with the test result before they get it back. So you need to know, if it's positive or if it's negative, how are you going to use that result in helping you managing your patient, especially with regard to treatment decisions? Are you going to treat or are you not going to treat?

Now, with that, we're talking about false positive assays. The one that's been approved that's on the market has very high specificity. And then in the smear positive population with a positive test, you know, I feel pretty confident that that truly reflects that the patient has active pulmonary TB. And so that's not as much of a concern to me on the false positive side.

And I think that when you look at on the false negative, that it's probably going to be unlikely in someone that's smear positive, that you're really going to get a false negative as long as you're able to assume that there's not inhibitors present in the specimen.

Now, with regard to performance characteristics of sensitivity and specificity, I think, when you look at the performance of the one that has been approved by the FDA and is on the market, it has very high sensitivity in the smear positive population and less so in the smear negative population.

But I think if you look at it as a whole, lumping the smear positive and smear negatives together, I think the target should be something around a sensitivity of greater than or equal to 90 percent. Now, certainly I think that's much higher in a smear positive population, but it's going to fall down to maybe in the 60-70 percent range in those that are smear negative.

But still, you're actually detecting, you know, the majority of patients who are actively infected with active pulmonary TB, even if that's the case in the smear negative population when you can't intervene appropriately with regard to therapy and isolation procedures, et cetera. So I think that the bar can be lower on the smear negative population and you'd still be impactful clinically.

Now, with regard to specificity, you know, I think that we already know that the specificity is very, very high, which is very good and gives you confidence in the answer if it's positive. I don't know if we should set the bar, if you will, at 99 percent per se because sometimes I know that when you look at the performance and you're looking at the point estimate being 99 percent, that you may end up having a lower 95 percent confidence interval that's down to 95 percent, and my recommendation would be that you look at specificity as being at least 95 percent or greater. And ultimately these may be, as we've already seen by the performance of the one that's been approved, that they are, you know, in the 99-percent range with regard to specificity. So I would be okay with that.

Now, one other comment. The last comment I'd like to make is I don't think that with down-classification, or reclassification, I should say, with special controls, I don't think we're lowering the bar compared to what we've done through the PMA process because, basically, the clinical evaluation of these devices, the clinical trials, are not that different when you look at whether they're Class II with special controls or Class III, roughly the same. There are very small differences. I mean, the PMA, you're going to get BIMO inspection of your clinical trial site.

But typically, when we design clinical trials that may be Class II with special controls or a PMA, there's not a significant difference in terms of the design of those types of studies. And if they're well conducted, you're going to get performance data that's going to give you a good feeling about the overall performance of that device as it relates to sensitivity and specificity.

So I don't want us to think that we're actually lowering the bar by doing a reclassification down to Class II with special controls because I think the special controls, considering cultures as well as the susceptibility testing, are going to be mitigation factors that are going to be appropriate to mitigate the risks that are associated with these devices. Thank you.

DR. DUERHRING: Gary Duerhring, Consumer Rep.

The only thing that -- I am not from your world, either. I'm actually from diagnostic imaging, is where I put in a lot of years. But

diagnostic tools are diagnostic tools. We can't forget that it's the clinician who has to make the final determination, and they're the ones making the diagnosis, not a test. Those are just indicators for the physician to actually build up a diagnosis. It's a comprehensive act of all things taken into consideration for the individual.

But as the Consumer Rep, I want to let you guys know you've scared me. I am not riding the subway or the Metro.

(Laughter.)

DR. DUERHRING: But I didn't realize that the standard was, or the tests that were out there were so weak. And I don't believe that by changing the grade from a III to a II is going to weaken the tests that are already available. But it may give an opportunity to improve, to find something that will actually contribute information to the diagnostician. And as a consumer and as a member of the public, I would really like to see that happen.

DR. BEAVIS: I haven't commented, that's right. I do favor the reclassification of these nucleic acid amplification tests from a III to the II. And just a couple of thoughts.

I see them to be used in conjunction with culture. I think that they should meet the performance standards of the existing nucleic acid amplification tests that are out there. And I would also be in favor considering different thresholds for sensitivity, depending on whether a

specimen is smear negative or smear positive.

The reasons for that have been alluded to by Dr. Hammerschlag and others. Especially in the paucibacillary cases, you'll get a positive nucleic acid amplification test, depending on which aliquot sample you test. And, again, the reason for that, these are hydrophobic organisms, paucibacillary. They may or may not be in that very small aliquot that's used for nucleic acid amplification testing. So I do favor the reclassification from a III to a II.

DR. CALIENDO: So, Steve, I'm going to summarize for you what I'm hearing and then we'll move on to (c). And if I miss anything, please jump in.

So as far as risk, I think what we've identified for a false positive is the risk of unneeded treatment and toxicity, time spent on contact investigations, and stopping to pursue an alternative diagnosis, which I think is a very good addition. If I understand correctly, the risk is less in peds for a false positive because you don't see positives very often.

DR. HENRICKSON: You know, I wouldn't want that to be said.

DR. CALIENDO: Okay.

DR. HENRICKSON: I think I'd stick with -- I'm much more worried about false positives than false negatives. So I actually think that the specificity needs to stay very, very high.

DR. CALIENDO: Okay.

DR. HENRICKSON: But I'm willing to give on the sensitivity.

DR. CALIENDO: Okay. The risk of a false negative is transmission from an untreated individual, delaying treatment in more advance disease, and missing an opportunity to trigger someone to think about an HIV diagnosis.

Now, as far as performance characteristics, first, is it fair to say that the Panel concurs with the reclassifying down from III to II? Is everybody comfortable with that? I need to see something affirmative. Okay. So it looks like -- go around the table? Okay. So we'll go around the table and just say your name into the mike and yes or no.

DR. DUERHRING: Gary Duerhring. Yes.

DR. TOWNS: Michael Towns. Yes.

MS. DeLUCA: JoEllen DeLuca. Yes.

DR. CARROLL: Karen Carroll. Yes.

DR. HARRELL: Lizzie Harrell. Yes.

DR. GINSBERG: Ann Ginsberg. Yes.

DR. JANOSKY: Janine Janosky. Yes.

DR. WILSON: Mike Wilson. Yes.

DR. CALIENDO: Angie Caliendo. Yes.

DR. RAND: Ken Rand. Yes.

DR. BEAVIS: Kathleen Beavis. Yes.

DR. GORDIN: Fred Gordin. Yes.

DR. HAMMERSCHLAG: Margaret Hammerschlag. Yes.

DR. REIMER: Larry Reimer. Yes.

DR. DELLA-LATTA: Phyllis Della-Latta. Yes.

DR. HENRICKSON: Kelly Henrickson. Yes.

DR. HAMILTON: Carol Hamilton. Yes.

DR. CALIENDO: So if I can summarize the performance information, I like the comment of not letting perfect be the enemy of good. I think we all understand the limitations of these tests, and it's actually, in my mind, not necessarily a limitation of the test but a factor of the biology of the disease is that you don't always see a lot of organisms.

And so the general consensus seems to be the bar should start at our current performance characteristics. We'd like to see it raise a little if it could. But smear positive looks to be about 90 percent. I like Karen's idea and Phyllis' concept of it's the patient, and you would encourage the design of clinical trials that look at more than one specimen; and that the gold standard isn't just the test or just comparing to culture, but it's truly clinical outcome; and have a clinical assessment of each individual patient by some independent person with the trial.

So that would take into account the opportunity to look a multiple specimens, and I think the value there is, if you can see that it's 99 percent sensitive with three, but 95 with two and 90 with one, you in your own lab can decide how to best manage that with your patient population. Everyone seems to agree, we should have a lower expectation for smear

negative specimens just because of the low burden of organisms.

I would also comment that you should have the same expectations for non-pulmonary specimens. I know the focus is pulmonary, but those are very different diseases with very different organisms. And I think Phyllis is correct. Any positive that you can get in that situation can be very useful clinically. I think we may have to tolerate a lower level of sensitivity in children. And that's what I have.

Does anyone want to add anything to that? Oh, and specificity very high. Everyone seems to be comfortable with the 99-percent specificity.

Michael, I think your comment is important. When you put a 95 percent confidence interval around that, you will drop below 99, and we need to accept that and understand that.

DR. GITTERMAN: Two questions. One, you actually addressed it right at the very end. It would be very useful -- and, again, you've been remarkably efficient -- to have an understanding of the lower bound because we could always go with point estimates. But a critical piece is what would the lower bound be on your estimate of specificity? So if we can just very quickly get a read.

Two other questions. What I heard, and just sort of filtering it in my own way, is what I heard from Dr. Della-Latta and as well as Carol, is that you would think it's important to describe in the labeling and in special controls that data has to be presented for all aspects of use of the test, that

is, you know, what the performance would be, let's say, with one specimen, with two specimens. So what you're advocating is that it clearly be described in the labeling, what the performance is under different scenarios.

DR. CALIENDO: And I would clarify that. I think when you determine the sensitivity, it needs to be on a case basis, but I think the data needs to be presented in the package insert and labeled that way so that a laboratory can go back and look and say this was one positive, this was two, this was three.

DR. GITTERMAN: Exactly.

DR. CALIENDO: Or this was only one, you know, so we have a sense of -- maybe we'll see tests next year where it doesn't matter how many you do on a smear positive; it's always 98-plus percent.

DR. GITTERMAN: Correct.

DR. CALIENDO: And we need to know that, and I think that helps us a great deal.

DR. GITTERMAN: No, we really appreciate that.

DR. CALIENDO: Now, let's talk a little bit. I don't know if we want to go to every station again, but the concept of the lower bound of the specificity. Michael threw out, if you have -- if the goal is 99 and you reach it, you may have a lower bound of 95 percent. How comfortable are we as a group with that lower bound?

DR. HENRICKSON: I'm not comfortable in pediatrics because of

the fact that the sensitivity is going to be so much lower in pediatrics. So then you'll have a low sensitivity and if you have a five-percent -- if you have 95 percent specificity, that's not a very good test.

DR. CALIENDO: Okay. So what would you be comfortable with?

DR. HENRICKSON: But I understand the practical issues of trying to get 99 on all studies. But the current tests have -- you know, are approved by the FDA with 99 percent specificity.

DR. CALIENDO: But the package inserts that we looked at, I don't think, for the NAAT, had the lower bound, at least the data that I saw.

DR. HENRICKSON: Oh, maybe it didn't have it in there.

DR. CALIENDO: So I don't know what the typical lower bound goes to for 99 percent.

DR. HAMMERSCHLAG: Yeah, that gets back to the specimen type.

DR. HENRICKSON: Yes.

DR. CALIENDO: No, but we're talking respiratory here.

DR. HENRICKSON: No, no. But that's critical, Angie. You keep going back to respiratory -- I mean, in pediatrics, you have to test respiratory and gastric aspirates as specimens for the diagnosis of pulmonary disease, or blood or whatever else. I mean, you have to find a specimen that works for diagnosing pulmonary disease. So I think that we need to encourage them to

do studies to try to -- or at least to report to you what some of the specimen types are. So I can think of --

DR. CALIENDO: Right. We can say that the specificity is not driven by specimen type, where the sensitivity is. And I think that's what we're all saying. We want the same level of specificity regardless of the type.

DR. HENRICKSON: But you did slip. You said pulmonary samples.

DR. CALIENDO: Right. Well, I'm only going back to that because that's what Steve --

DR. HENRICKSON: Okay.

DR. CALIENDO: -- had started with. But I think everybody hears the whole pediatric gastric piece.

DR. HAMMERSCHLAG: Well, there are no parameters, performance parameters, available actually for gastric aspirates, I don't think.

DR. CALIENDO: Right. And so we run the risk of do we hang a bar we can't hit because no one knows what that bar is going to look like?

Other comments on 95 percent as an upper, lower -- go ahead, Michael.

DR. TOWNS: Michael Towns, Industry.

I actually have the package insert in front of me for the current approved assay, and in smear positive patients, on specificity, the point estimate was 100 percent, but the lower bound 95 percent confidence

interval was 59 percent. Upper bound was 100. And the smear negative patients' lower bound was 96.1 and the upper bound was 100 percent, although the point estimate in the smear negatives for specificity was 99.3.

That's why I brought up the comment because I didn't want us to be tied to having a specificity requirement of 99 percent if the FDA was going for clearance purposes, assuming it gets reclassified, or we can say clearance as opposed to approval, that they demand that even the lower bound is greater than 99 percent, if we were to say that 99 percent should be the bar at which we have the specificity requirement.

DR. HENRICKSON: You said, for smear negative -- Kelly Henrickson. You said, for smear negative, it was 96.3? Was that what you said? The lower bound.

DR. TOWNS: The point estimate was --

DR. HENRICKSON: No, the lower bound.

DR. TOWNS: -- 99.3. The lower bound was 96.1.

DR. HENRICKSON: Right.

DR. TOWNS: And I'm just saying that if we say 99 percent on the specificity, we don't want to be tied to having that as the requirement of the lower bound for specificity for any respiratory specimen.

DR. HENRICKSON: So I'd propose for smear negative that -- Kelly Henrickson again -- that it be no less than 96.3 -- 95, like you said, the lower bound, the lower.

DR. TOWNS: I mean, the reason the lower bound was so low on the specificity of the smear positives, there were so few of those. There are only seven of those. Seven out of seven were positive for the point estimate of 100 percent, but your lower bound goes down to 59 percent. You simply didn't have enough in that bucket.

DR. HOJVAT: And you're exactly right. I wanted to point that out. The lower bound is dependent on the number of specimens, okay? And seven is pretty small.

(Laughter.)

DR. HOJVAT: Okay. I mean, it may be the availability or whatever, but you know, usually you have to have at least 30 -- statistician -- to get a 99 -- let's say it's 100 percent. Your lower bound, the minimum number would be --

DR. CALIENDO: Go ahead, Janine.

DR. JANOSKY: This is Janine Janosky.

DR. HOJVAT: -- 30 or 50 or something like that.

DR. JANOSKY: This is Janine. Can I jump in here? I've been trying to get attention here for a few seconds.

You know, I'm going to caution us about this conversation, and following up on that point, you know, it's the estimate, it's going to be the sample size, and it's going to be the deviation. So simply by increasing the sample size, you typically could have a closer or a less broad confidence

interval. So I really want us to be cautious about placing that lower bound without being cognizant of, by simply increasing the sample size, you tighten it. So that's a very important point, I think, that really needs to direct our conversation. So I hesitate in us choosing a lower bound on a confidence interval.

DR. CALIENDO: I agree.

DR. JANOSKY: So thank you for letting me follow up on your comment there.

DR. HENRICKSON: If we don't choose a lower bound, then, do you want us to choose the number of specimens they should test?

DR. GITTERMAN: No. I've obviously unleashed a monster, so please let me clarify. These two aspects, it's not a normal distribution, but there's the point estimate and there's the uncertainty that you tolerate. And when you're talking about a lower bound, you're basically saying, if the point estimate is blank, how much uncertainty are you going to accept?

And, again, 95 percent -- let's say you assume your point estimate is .99 and you again assume you won't tolerate a lower bound of .95, that will then drive how many patients you would enroll in your clinical study. If you decide that you want a lower bound with the maximum uncertainty of .02, or 2 percent, and a lower bound of 97 percent, then, in fact, that would drive your -- basically how you designed your clinical trial. What you want to describe is basically what performance you'd like and the

degree of uncertainty to design your clinical trial, rather than the other way around.

You know, really the question was directed to say, how much uncertainty would you be willing to tolerate? I think your initial approach was correct. Would people accept 95 percent as a lower bound? You know, it was really just focused simply on that issue, what would be --

DR. CALIENDO: Right. And I think it's going to be tough for us to be more specific because we're not going to know how many people are going to end up in each bucket when they're done.

DR. GITTERMAN: Well, not to be -- and again, this -- but for specificity and -- sensitivity, but specificity, there shouldn't be many buckets. They should all be smear negative, you know, if they don't have TB, of course. Now, one could argue for HIV positive and negative, but I don't see many -- please, I apologize, I don't see many buckets in specificity.

DR. CALIENDO: I'm just looking at what Michael had said, and there was only seven in the smear positive group, which is why we ended up with such a -- so that's the only part I'm trying to make.

DR. GITTERMAN: I apologize, I missed that. Thank you.

DR. CALIENDO: So Steve, are you comfortable that we've covered (a) and (b)?

DR. REIMER: Can I throw in one more?

DR. CALIENDO: Sure, sure, go ahead.

DR. REIMER: This is a little bit out of context.

DR. CALIENDO: Name, name.

DR. REIMER: Larry Reimer from Utah.

In terms of risks, just to point out a couple of additional things. So in terms of false positives, we haven't said anything about the emotional impact of that on the individual who's labeled as being TB positive, and until they find out that they're not, I think that is an issue for them as an individual, and I think, from the public perspective, I think we need to be aware of that.

And the flip side in terms of false negatives, the one context that I think is important is that in people who are HIV positive, it is definitely true that if they have combined TB and HIV, if they are treated for their HIV before they're treated for TB, there are negative consequences for the outcome of their HIV infection. So I think that is something that probably needs to be in your record.

DR. GITTERMAN: I think that's a very good point, but if I could just translate it. What you're advocating is, again, not suggestion we're taking any action directly, but if we were to write a special controls guidance, you would emphasize the need for doing HIV testing in anyone who is determined to have TB. That's how I would translate what you're saying.

DR. REIMER: Which is already true anyway. But yes, absolutely.

DR. CALIENDO: A good point, thanks. Steve, you're otherwise comfortable with (a) and (b)?

Okay. So then we're going to go to special controls, and I'm going to tell you what I've heard you say so far, at least that I was able to write down, and then we can add onto that. One is we need specific clinical trials for children, using gastric aspirates. Two is this isn't a standalone process, that we need to understand that it's going to be used in conjunction with culture and clinical information. And the other one I heard was education, education, education, both for the laboratory, the clinicians, and the public health setting, of just how the tests should be used. And, you know, there's the package insert that we fear nobody reads. And so how else are we going to go about this whole educational process?

So that's what I've heard said so far, but we haven't really focused on special controls, so let's open it up and hear your thoughts on that.

Go ahead, Karen.

DR. CARROLL: So to a certain extent, I think these already exist. I think they exist in the CDC guidance documents that we've had over the years on appropriate utilization of these tests. And if you do read the package insert in detail, which we were required to do for this Panel, there's a lot of really good information in there, on the appropriate use of this test.

And so my feeling is that in addition to the comments that were

made that probably aren't well addressed, I think if we follow the guidance document as set forth by the CDC, in my experience with both a low-prevalence situation and a situation of moderate prevalence in both scenarios, in the low-prevalence scenario, the smear positives are almost always TB. In the moderate prevalence scenario, the smear positives are almost always not TB. The test has been performing as specified by the package insert. And so that is one of the reasons why I feel comfortable with the recommendation to not downgrade or whatever, reclassify, you know, this test.

So I think we should look to what already exists in terms of the performance of these tests, in terms upon drawing the special guidelines or considerations. I think they're out there.

DR. CALIENDO: Angie Caliendo.

So Karen, how do you feel about postmarket follow-up in light of that? Because there will be a certain amount of information that's reviewed. Do you feel that that would be a special control that would be helpful or not?

DR. CARROLL: I absolutely do think that's the case. And I want to add to Dr. Reimer's comments. I think that in addition to TB in our immigrants, in our HIV population, the other challenging population that we are faced with, where the presentation is often very subtle, is our transplant population and those patients who are getting anti-TNF therapy.

And so I would like to see more data on the performance of these tests in those patients who are likely to present in very subtle fashion. And that could be addressed in a postmarketing type of scenario.

DR. RAND: Angie, could I make a comment? It's actually sort of a general comment on postmarketing testing because this does apply to more than just TB. It really would apply to almost any nucleic acid test.

But going forward, one of the risks of nucleic acid testing is that the organism has mutations that occur in the primer set and in the probes, if it's a TaqMan type or a molecular beacon or some other assay along those lines. And you as an individual lab director in an individual lab have no way of detecting a false negative if that sequence mutation is such that you get a falsely negative result because you don't do two unrelated tests, either two nucleic acid tests or, you know, an antigen test every time you do the nucleic acid test. It's prohibitive. You can't.

So the plea that I would make is that in the postmarket follow-up, there be some provision for postmarketing sequencing to verify that the specific assay isn't missing mutations that occur down the road. And one way that could be done would be looking at a preset number of negative tests from around the country and seeing if any of them turn up positive.

Another way would be -- or part of that would be just sequencing the PCR sequence itself to see if there were any mutations occurring in there.

And a third aspect of that would be the possibility of just, en masse, looking at the cycle thresholds so that if mutations were starting to occur in the regions that you were assaying, you might start by seeing positives that had lower CTs. By comparing this year's samples to last year's, you might see that a drift was occurring.

So I would really encourage the incorporation of this kind of postmarketing special control guidance in our recommendation.

DR. HENRICKSON: Kelly Henrickson.

I agree with that completely, and we've done that for years for viruses, and we need to do that for this in this document.

DR. HAMMERSCHLAG: Hammerschlag.

There actually is a precedence because it's happened with chlamydia trachomatis, but limited to Sweden, where actually a variant developed which had a mutation deletion right at the primer sequences for the Roche AMPLICOR, and it was a cryptic plasmid. It didn't seem to be, but it didn't get out of Sweden. More recently, we're also seeing major problems also with DNA assays and specifically also AMPLICOR for gonorrhea.

The fact in this case, however, is more affecting the specificity because GC will exchange genetic material throughout its life cycle with other *Neisseria* species, so that it would -- the test specificity has gone down the toilet. But GC being so much -- I mean TB is slower growing. I don't know if that's going to quite be the same issue.

DR. DELLA-LATTA: I think it's a good point. It was a point I was going to make in terms of resistance factors. I don't agree with it in terms of TB. I don't see shifting affecting any of this stuff. I truly don't. I've been working for 20 years on this testing and I really don't see a problem there whatsoever. It's just extra work.

However, having said that, when we move on to resistance testing, I have been doing work with the mutations and the SNPs that have to be added in. To read it, you have to have a flexible system. So I guess that'll be next on the agenda.

DR. GORDIN: Fred Gordin.

So just for the record, you had asked about special controls and, I think, named three or four. But whoever gave us the book actually had 10 pages of special controls, which I just want to endorse.

Somebody used the comment here, like we would lower the bar. I really don't want to lower the bar. I might want to make it easier to go through the pathway, but I think, given the low incidence of TB in this country, and we're supposed to be talking mainly about using the device in this country, that we really don't want to again have false positives as well as false negatives again, for all the reasons that were brought up about over-diagnosing people who don't have TB.

So I thought this was very well written, pages 17 to 27, and I would just say I would endorse this whole group of controls.

DR. CALIENDO: Thank you.

DR. BEAVIS: I just want to make a brief comment about risk mitigation. This doesn't have to do with patients or public health so much, but with laboratory technologists.

One of the advantages of the kind of test that we're talking about is they can be done in a point-of-care setting, and I think that the proposed testing is fine in a point-of-care setting with the kinds of safety standards that happen or don't happen in a point-of-care setting. But I feel strongly that it should be limited then to the testing of a specimen.

What I would not want to see happening are concentrated specimens or growth from culture in a liquid medium, then being put through the system or tested in an area that doesn't have the appropriate biosafety. Again, I always picture a vial getting dropped, of concentrated TB and then having that dropped. So for me, I would want to stress that the safety has to be a part of it as well.

DR. CALIENDO: How do others feel about that comment and the monitoring of genetic variability for the detection? We're not talking about the resistance ones yet, just the detection. I'm curious to hear other people's thoughts.

DR. CARROLL: Yes, on the safety and I don't -- I'm with Phyllis. I don't really feel that, for the diagnosis of TB in a postmarketing setting, we need to look at genetic drift. I do feel strongly that it's going to be very

important, though, in the discussion on resistance testing.

DR. HARRELL: Yes, I do echo what was said about safety in the laboratory. We don't want those concentrated specimens being moved around where they could be dropped and a technologist get exposed to the organism.

And as far as the genetic testing, I don't have any experience with that, so I'm not advocating it.

DR. CALIENDO: Do you feel that it's the movement of the specimens, Kathleen, or that it's just that we should stick to primary clinical specimens?

DR. BEAVIS: To me it's the movement. I could see if the instrument for the detection were to be placed in a BSL-3 area or wherever else a laboratory works with concentrated specimens, that it could be used in that setting. I just didn't want tubes with, you know, concentrated MTB being brought up to the emergency room so they could be run through the point-of-care instrument. Unless the emergency room has more safety standards than mine.

DR. HENRICKSON: Kelly Henrickson.

For those who made the comment about the genetic drift, I'm an RNA virologist. Are you saying that there's data out on this -- RNA target that demonstrate that it hasn't changed for 20 years and that's why we shouldn't worry about it? I mean, if there's data out there, then that should

be presented.

DR. RAND: I don't know of any. I'm with you on the virology angle, and that's where I'm coming from, making the comment. But it is a generally applicable comment, although it may not be specifically applicable to TB.

DR. HENRICKSON: There are many bacterial genes that are very stable and that's why, not being a bacteriologist, I didn't --

DR. DELLA-LATTA: But these generally are, so I do have some information, but not on me and not in my head. But I do know that it's -- I mean, virology is different. Viruses -- things that mutate.

DR. HENRICKSON: RNA is the target, but there was no package insert for this. They didn't give me the genetic background.

DR. CALIENDO: This is Angie Caliendo.

Any other comments about special controls? Phyllis.

DR. DELLA-LATTA: I hate to go back to the specimen again, but that's where it all begins. Can we somehow make sure that they are sediments and not raw sputum? I'm very afraid there are too many problems, and I think there might be companies out there that would consider it very advantageous to say spit in a cup, and then just stick in someplace to test. And I feel we all know about the problems with lack of homogeneity with sputum specimens that have not been processed.

DR. CALIENDO: So Angie Caliendo.

So, Phyllis, I think that's interesting, but I would bet that'll come out in the performance characteristics if they chose to do their trial that way.

DR. DELLA-LATTA: Okay, fine.

DR. CALIENDO: It would be my thought on that.

Carol, I think you had a --

DR. HAMILTON: Yes. Just in reading the documents in preparation, there were a couple of questions around like what kind of clinical studies should be done, and one specific question had to do with, I believe, whether patient outcomes needed to be considered in clinical studies. And while I love getting clinical trials with patient outcomes, I do think that that's a rather prohibitive kind of study to do. And so I guess I would just advocate to hold to the standards. I don't believe the NAA tests that are currently out there looked at patient outcomes.

But one area where I think it would be very important in a postmarketing setting is to look at those individuals who were culture negative but NAAT positive. I think that's a very interesting population, as we have already talked about, that our liquid cultures are not perfect. And so what do we learn from that? So, anyway, I just wanted to address that issue of patient outcomes.

And then the other kinds of things I thought we might want to consider is, in terms of the clinical studies, we need to think about what kind

of population. So is the FDA going to need to ask for studies that are, I mean, easy or potentially getting from low-incidence countries like the U.S. and Western Europe? But what about, you know, Latin America, where -- really most of our cases are coming from Mexico and other parts of Latin America. So what about the test performance in those populations? And are they going to consider allowing or should we encourage getting specimens and a study done in a high TB incidence area? And what about from a high TB and an HIV incidence area? And then, of course, the issue of adult and pediatrics.

So just in terms of the populations that I think need to be considered, I think those are some. And then when we get to resistance testing, of course, then we'll get into the MDR populations.

DR. CALIENDO: Can I ask you to take it one comment further, in stored versus primary specimen?

DR. HAMILTON: Yeah. Carol Hamilton again.

So, for sure, I think it would be ideal to have fresh specimens. Now, for one thing, I would be concerned that using archived specimens is going to potentially make the test look suboptimal. You know, I think, as much as possible, fresh specimens.

I will say that in terms of the process specimens and the sediments, et cetera, I mean, it depends on what the test characteristics are. So I think the one that's been recently approved by WHO, it is being used on non-sedimented specimens, which is real world in many places in the field.

So, anyway, I don't feel like that should be confined.

I would prefer fresh specimens, but in situations where we have very few. So drug resistance. If we're going to try to see how it performs in drug-resistant cases in the U.S., you're going to have use archived specimens. So I think when there's, you know, rare circumstances, then it should be allowed, but it should be a minor proportion.

DR. CALIENDO: Any other thoughts? Phyllis.

DR. DELLA-LATTA: Actually, the archived frozen specimens and sediments work beautifully. In fact, they get rid of inhibitors as well. So I have mine over 20 years.

DR. CALIENDO: Okay, Karen.

DR. CARROLL: I was just going to echo. There's really good literature on that as well. I mean, it's old literature, but it exists.

DR. CALIENDO: But the specimens are old too.

DR. CARROLL: Yeah.

(Laughter.)

DR. CALIENDO: Any other thoughts on special controls?

(No response.)

DR. CALIENDO: So let me add to what we've come up with. So we like the idea of following the CDC guidance. Everyone loves what the FDA has put together so far, so thank you very much for your work on that. There seems to be kind of mixed feelings on we want postmarket studies,

particularly on special patient populations like transplant patients, immunocompromised patients in general.

There probably isn't data, but if you can find data to support genetic variability, whether we need to do it on NAAT, it clearly needs to be done for the resistance testing; but just a detection NAAT, maybe not.

The lab safety needs to be considered. So if you're going to do this on anything other than the primary specimen, they have to have their lab designed to accommodate that.

Sediments, not. Raw specimens is what Phyllis' comment was. We gave you populations for clinical trials: low, medium, high; HIV/TB co-infected; and both adults and pediatrics. And then several Panel members have experience with archived specimens that are quite stable. So in certain circumstances, that might actually be quite helpful to augment.

Have I missed anything? Go ahead, Ann.

DR. GINSBERG: Just to clarify, I think you had said before that we don't really want to specify whether it has to be sediment or a fresh specimen. That's going to come out in the product specifications.

DR. CALIENDO: Right.

DR. GINSBERG: And that's probably --

DR. CALIENDO: So this is Angie Caliendo.

I think the only issue, Ann, is the laboratory safety issues associated with which specimen they're using. So I do agree. The trial could

be designed any way the company wants to design it, but they need to keep their lab safety issues in consideration.

Is that fair, Kathleen?

DR. BEAVIS: Yeah, because preparing the sediments has its own risks and complexities as well, and I think that would lose some of the advantages of the test.

DR. CALIENDO: Okay. So I think we finished Question 1. We have 15 minutes before break. So, Steve, if you can come up and read Question 2.

DR. HOJVAT: Angie --

DR. CALIENDO: Yeah, sorry, Sally.

DR. HOJVAT: -- just a point of order. You know, this is an issue panel. Can we sort of not do the vote thing again?

DR. CALIENDO: You want us to vote again?

DR. HOJVAT: No, no, we're not supposed to vote.

DR. CALIENDO: No, no, no, I want him to do Question 2.

DR. HOJVAT: I know, but I just wanted to make sure. You know, we did a vote.

DR. CALIENDO: Oh, oh, that was --

DR. HOJVAT: I know that was really just for you --

DR. CALIENDO: Right, right, right, right.

DR. HOJVAT: But if we can sort of not --

DR. CALIENDO: So that's really a consensus opinion. That was not a formal vote.

DR. HOJVAT: Consensus opinion.

DR. CALIENDO: Right. Sorry, Sally.

Okay, Steve.

DR. GITTERMAN: For nucleic acid amplification assays that detect genetic mutations associated with antibiotic resistance to *Mycobacterium tuberculosis* complex directly from respiratory specimens:

a) Please discuss the risks associated with inaccurate tests for the detection of genetic mutations associated with antibiotic resistance to *M. tuberculosis* complex.

b) Considering the relative infrequency of antibiotic resistance mutations in different populations, please discuss the appropriate role of prospectively collected samples, archived/banked samples, and spiked samples in developing device performance; in addition, please comment on the minimum device performance standards that should be recommended for these assays.

c) Based on the above considerations and the earlier presentations, please discuss if sufficient risk mitigation is possible for FDA to consider classifying these devices

as Class ll by drafting a Special Controls Guidance through the de novo regulatory pathway. In your discussion, please note if there are any specific special controls in addition to those previously mentioned that should be considered for these devices.

DR. CALIENDO: Thank you, Steve. This Angie Caliendo again.

So why don't we try to get -- we're going to break at 3:00 for a few minutes. It's quarter until.

Why don't we try to get 2(a) done before we break and then we'll come back and do 2(b) and 2(c). So comments on 2(a), which is basically the risks. And now we're on tests that detect genetic mutations associated with resistance.

DR. GORDIN: So this is Fred Gordin.

So speaking as a clinician, I guess I'm more concerned about this just in terms of low rate of TB than the low rate of drug-resistant TB, that the false answers, mainly false positives, potentially, I guess are of great concern to me, that people may not use a drug that is highly effective if they're led to believe the drug is not of value and could delay by many weeks getting on ideal therapy. So I guess I have some more concern here than I did with the earlier discussion, just to throw that out.

DR. HENRICKSON: Kelly Henrickson.

Do you feel better if we put in a special controls some kind of

link to sending it in to the CDC to be actually, you know, genotyped?

DR. GORDIN: Well, it gets back to the whole idea -- I think the world needs the products. We in the U.S. could use them. We probably don't need them as much. And I think Michael, as I'm sure everybody would be, to help move the products along. So it gets back to that whole earlier discussion. We don't need to repeat about what's the difference between a III or a II with lots of extra attachments?

So I'm not saying it necessarily should be a III, but I'm just raising the issue that in the U.S. there's really not very many cases of drug resistance per year. And so if people use this widely, if there are false positives, the predictive value is going to be not nearly what it would be if you were in South Africa or somewhere else.

DR. CALIENDO: Angie Caliendo.

So, Kelly, I see that as a good postmarket management for a number of years, until people are comfortable. I think that's a really good idea.

DR. RAND: Could I clarify? Wouldn't the drug rifampin testing only be done on something that was already known to be TB positive? I mean, you wouldn't do it as a screening test. The drug resistance wouldn't be done as a standalone screen on people who were being -- separately from molecular testing for TB.

DR. CALIENDO: So Ken, just to clarify, you could visualize a test

that did both detection and resistance in the same test so you would get the answers simultaneously?

DR. RAND: Yeah, I would visualize that or that a standalone rifampin test only be done in conjunction with --

DR. CALIENDO: And that might be a good special control, that one test couldn't be done unless the other one was done first.

DR. GORDIN: But even if that's the case, I guess the data we heard this morning from Ken was one percent of isolates in the U.S. are rifampin resistant. So, by definition, even though that population of everybody with TB, we already know one percent. So 99 percent are not going to have it. So I'm just saying the predictive value is going to be a whole lot lower than if the rifampin resistance in the United States were higher.

So, again, I don't totally know what to do with that information. It just makes me think we're going to end up with more false positives because it's one percent of everybody with TB.

DR. HAMILTON: Right you are. I mean, by definition, in a very low-prevalence --

DR. CALIENDO: Name.

DR. HAMILTON: Carol Hamilton.

In a very low-prevalent situation, the majority of what we're going to see are going to be false positives. And so I think that's where we get to our controls, where you have to say, you know, you do this. But we

would probably have a recommendation that says, if that's the only information that you've got and you don't have another reason to suspect drug resistance, you put them on standard four-drug therapy and two additional drugs, while you're waiting on your culture, your sensitivity test from your culture.

But I do want to get back to something that Dr. Rand was saying. So the test that's been approved by the WHO right now, the GeneXpert, in fact, you get the answer, TB, yes or no, and rifampin resistant, yes or no, at that moment. And so yes, there will be the situation where you didn't really suspect rifampin resistance, but you've now got that information.

DR. RAND: Ken Rand.

Yeah, but future tests might not be, you know, conglomerated that way.

DR. REIMER: This question is worded broadly about antibiotic resistance in general. It's not specifically about rifampin. And I think keeping it in those broad terms, it would be very difficult for us to say that we could come up with special controls and things like that that would mitigate the risks that Fred was talking about.

So I would be really reluctant to move very far ahead with this right now, without having something more broadly defined in terms of what all those resistances might be and knowing what kind of mutations to expect, where all of that is going to go in the future.

DR. RAND: My question is, how would Class III protect you from those concerns?

DR. CALIENDO: Angie Caliendo.

Just a comment. I think the CDC presented us some interesting data of what they're doing internally, as far as expanding out genotypic testing. So we also have that, knowing that that system is well underway at the CDC.

DR. GINSBERG: Ann Ginsberg.

I guess I was just going to flip the thinking a little bit. Under the current situation, patients go for a long time now, potentially, on inappropriate treatment if they're actually resistant and there is no rapid testing, right? So to some extent, to me, that's an even bigger concern than using the approach suggested, and I think it was by Dr. Hamilton, that if you get a rifampin or an INH resistance rapidly, that you might put patients temporarily on standard care, plus a couple of second-line drugs, while you get phenotypic testing done.

And I don't think anyone's suggesting using these rapid genetic mutation detection approaches in the absence of also doing phenotypic testing in culture. So I think I'm a little less concerned than some of the others.

DR. CALIENDO: Okay. Other thoughts on risks of false positive and false negatives. Go ahead, Phyllis.

DR. DELLA-LATTA: Firstly, when we're talking about INH and rifampin, both of those, I think our knowledge about the mutations on a primary SNP is pretty good. I think there's not as much variation because, again, we're dealing with -- I'll just talk about New York, people from all over the world. So it's a melting pot.

But still, you have -- I've been working with some of the mutations. However, when it comes to the other drugs, that's another problem. What I would love to see is performing the test directly from sediment, yes, for INH and rifampin. But I'd love to see it from culture. I'd like to be able to have the culture available in case there are any problems.

DR. CALIENDO: Right. I don't disagree, but I think we won't drive what the manufacturer decides is right.

DR. DELLA-LATTA: No, no, no, no. No, I do understand that, but this is something that -- the only time you would get false positives as an issue is if a company decides to use a SNP that's -- it's not relevant. And there are some out there that are not relevant to -- it doesn't compare to the clinical or the phenotypic. It doesn't compare well to clinical or phenotypic correlations, so you would get a false positive. But it's more false negatives that you want to worry about.

DR. CALIENDO: Okay, Karen.

DR. CARROLL: I think I share the trepidation already mentioned about the false positives because this is not a trivial issue. And adding two

additional drugs doesn't sound like a big deal to us, but it's a huge deal for patients. And if we act on information and rob people of the first-line therapies when those would be optimally the best in the case of sensitive tuberculosis, I think it has enormous consequences for the patient.

I also somewhat disagree with the comment that patients would likely go four weeks without getting therapy if they needed it because I think right now the approach in the U.S. is to risk stratify. If a patient has acquired their TB or likely acquired their TB in a country where MDR or XDR is very prevalent, our infectious disease docs treat them as though they may have drug-resistant TB until the results of testing come back.

So I think it gets back to the clinical context, and so I have a lot of trepidation about this.

DR. CALIENDO: This is Angie Caliendo.

So my comment to you is, how is that practice different than what we would do with a false positive result, weighing that against what Ann said, is that resistance testing might help us up front in a significant number of patients?

DR. CARROLL: So I guess if it's the first specimen and physicians decide to stop and we have a patient then who's culture negative and we don't have benefit of the phenotypic testing and the results are false positive. So I'm just thinking --

DR. CALIENDO: So you're talking false positive still?

DR. CARROLL: Yeah. So was that your question to me? Okay. So I guess we would -- if we were to reclassify these, I would have to say that the special controls would be that we not deviate from the practice of getting a minimum of three specimens on patients and that we continue to follow the guidelines, you know, for phenotypic testing.

My concern is more along the lines of this is the first specimen; action's taken; physician feels they have a diagnosis; subsequent specimens aren't obtained; their culture is negative and we don't have benefit of phenotypic tests. I think there are lots of implications there.

And I also feel that, as a Panel member, there's not a lot of data for me to really feel strongly about, you know, where to go with this. I mean, I feel very strongly that we have enough information on nucleic acid amplification tests. But in the United States, where resistance is so low and no data on the performance, I mean, we weren't really presented with a lot of data on the performance of any resistance test, other than the CDC presentation, which really didn't have a lot of data, it's really going to be hard to really support reclassification of these, in my opinion.

DR. CALIENDO: Go ahead.

DR. BEAVIS: No, I second what Dr. Carroll was saying. I know, for example, if we were to get a false positive or a false negative susceptibility result in my lab, I could point to the two or three steps in the process that would explain why we got that false result. And that was one of

my questions about the performance of the rifampin resistance test.

If somebody were to get a false positive, you know, and then you got two additional specimens, would those two additional ones be negative, or is it a characteristic of the patient specimen that's causing a false positive? And just don't know enough about the performance characteristics of the test to be able to weigh in on that.

DR. RAND: Could I offer a comment? Suppose the manufacturer --

DR. CALIENDO: Name.

DR. RAND: I'm sorry. Ken Rand. I shot my mouth so much, I thought you all knew.

Suppose a manufacturer took the trouble to get a PMA and followed the Class III, you know, requirements and got approval for, let's say, a rifampin or an INH resistance nucleic acid test. Would that process change your concerns about the risks of false positives and false negatives?

DR. BEAVIS: I'm familiar with the current testing that's on the market, and I guess that's Class I, not Class II or III, and with how the testing is performed in the laboratory. And I'll be honest, I see this more as a totally different change in technology. It's more of a black box kind of thing. So I wouldn't know how to evaluate a laboratory -- if we turned out a wrong result, I wouldn't --

DR. RAND: But if it was on the market, it was FDA approved as

a Class III device as opposed to a Class II, how would that change your dealing with a problem?

DR. BEAVIS: I think if it were a Class III, there'd be the opportunity for some more postmarketing surveillance to try to go back and troubleshoot what may or may not being going on.

DR. CALIENDO: And if that were -- this is Angie Caliendo. If that were part of special controls, would that help?

DR. BEAVIS: Yes. I just like to understand what's going on with the process, and I always make the assumption that certain things are going to go wrong and I'd like --

DR. CALIENDO: Right.

DR. BEAVIS: -- to be able to understand the process to do it better.

DR. CALIENDO: Right. And I think we have to remember -- Angie Caliendo again -- there will be a clinical trial. You know, there will be data there that the FDA will be assessing.

Someone else. Go ahead, Phyllis.

DR. DELLA-LATTA: I just want to echo that whether it's II, Class II with special -- with those special options and special requirements or controls, I think it would be the same as a III. I really think we can make it such that we wouldn't have a concern there.

And in terms of doing multiple specimens, it really depends

upon whether you have a contamination issue that was in the lab, in terms of performing the test, whether one specimen would be contaminated. But if you did three and it was a primer issue and the wrong primers were there or the primers that were there were unable to detect the TB from my patient from Ecuador, that's a different issue. Then you'd get the same result times three.

So we have to think in terms of somehow proving the importance, the utility and the validity and accuracy of the primers that they're using in the tests. Their target is primers.

DR. CALIENDO: We've talked about false positives. What are your concerns from false negatives?

Go ahead, Mike.

DR. TOWNS: Michael Towns, Industry.

I just wanted to make a clarification to something that Dr. Beavis just mentioned. She mentioned Class I for the AST testing for mycobacteria. They're Class II instead of --

DR. BEAVIS: Okay, thank you.

DR. TOWNS: -- instead of Class I. I just want to make that clarification.

DR. CALIENDO: False negatives. Any thoughts?

DR. BEAVIS: It's Kathleen. This is Kathleen Beavis.

I don't think there really are any because we'd be waiting for

routine susceptibility for that as well. And I know in our patient population, patients are put on three or four drug regimens, depending on their risks.

DR. HAMILTON: This is Carol Hamilton.

I have one more thing I want to say about false positives that can be, I think, handled by the controls is, you know, we should not -- I mean, this kind of gets -- maybe I'm going to say exactly the opposite of what I just said to Dr. Rand a minute ago.

But in terms of thinking about how to interpret that rifampin resistance piece, or whatever resistance, in general, you're right, that the time we're really going to be seeking that information should be in patients where we have a specific reason that we are looking for and we're concerned about drug resistance. So people who have failed therapy, people who are getting retreated for therapy and defaulted from treatment, or people who are coming from countries where there is a high incidence of MDR-TB. So, again, it's the pretest probability.

And so if we have that population, we're going to be, you know, paying a lot more attention than just out of the blue somebody happens to send it and it's -- so I just wanted to say that.

DR. CALIENDO: Angie Caliendo.

It's an interesting comment, Carol, because if we got a rifampin-resistant isolate in our laboratory -- and we see a lot of TB in Atlanta -- we would be shocked because we don't see MDR. So I think you're right,

there is -- what is happening in your area of the country? You know that. You've been culturing TB for years. You've been looking at your phenotypic susceptibility results for years. You have a sense of this is a common occurrence or is this not a common occurrence?

DR. HAMILTON: This is Carol again.

I'd be thinking more on a patient-by-patient basis. But, anyway, same point.

DR. CALIENDO: You still know --

DR. HAMILTON: Absolutely.

DR. CALIENDO: -- how often it comes to your lab.

DR. HENRICKSON: Kelly Henrickson.

So the practical aspects of that would be that you'd have to do two molecular diagnostic tests in your laboratory, the one you use routinely, except for when you get patients where you think they could have drug resistance, and then you run this multiplex test because otherwise you'd be running the multiplex test on all samples and then just not looking at the answer, like sometimes we do.

DR. CALIENDO: Angie Caliendo.

So I think it depends on how the multiplex test is designed. Can you suppress some of it or not? And if you cannot, then I think it goes back to judgment. And this is where every case of TB should be seen by an infectious disease physician or a pulmonologist. And education. As we said before,

education, education, education.

Yes, Phyllis.

DR. DELLA-LATTA: I think false negatives is more of an issue. I think it's a big issue, particularly if it's multiplexed. If it's multiplexed, how are you going to resolve that? I mean, the assay is called a multiplex. It's called an INH rifampin TB test.

DR. CALIENDO: Can you clarify what you're talking about with the false negative or the multiplex?

DR. DELLA-LATTA: I was talking about -- no, if you have a test that is able to detect TB directly from sediment and also do INH and rifampin, that would be ideal, right? If there are, let's say, a high number of false negatives, or even a few false negatives, I mean, false negatives are a problem because, again, in reporting a result, you would report positive for TB, and then what do you report for the other two? I mean, are we going to be using a test that we can't with confidence say no to? You're spending a lot of money on an assay that detects three different -- you know, two drug resistance and one for TB.

DR. CALIENDO: So, Phyllis, help me understand the risk of a false negative. Put the risk in a way that the FDA can put their head around it.

DR. DELLA-LATTA: All right. The risk of a false negative is I'd be using a test for the purpose of detecting the resistance and therefore I would

report it. It's there, I'm doing the test, I report it. So if it's false negative, that's obvious.

DR. CALIENDO: What's the risk of that?

DR. DELLA-LATTA: If it's a false negative, we have a drug-resistant patient that's running around, being treated --

DR. CALIENDO: Untreated.

DR. DELLA-LATTA: -- with primary drugs and transmitting multidrug resistance, unfortunately, and we'll be back into the '90s in the Big Apple.

DR. CALIENDO: Okay. Now, balance that with what Kathleen Beavis said about, without the tests, you would be waiting for that result from the phenotypic testing or the clinician would put them in a moderate or high-risk group and treat accordingly.

DR. DELLA-LATTA: That would only be important to me if there were separate tests and not multiplexed in one test. And that test standing by itself, fine. A test for drug resistance, if it was negative, what am I going to say? I'm spending a lot of money, you know, doing a test, so what do I say? Presumptive negative? Do I say nothing and waste money? I mean, you just can't do that in a laboratory. You have to give a result. If you do the test, you have to give a result.

DR. BEAVIS: I think there's two questions that are getting embedded, and one question is, what do you do with an individual false

negative result on a particular patient? And from where I am right now, I don't know how much effect that would have on a patient given our current situation, all right, because the patients would get treated with three or four drugs, depending.

But Phyllis, you're asking a much more sophisticated question, which is, how does a laboratorian decide whether or not to perform a test based on the test sensitivity and specificity and the prevalence in your area? So not only do you choose to perform it, but what do you choose to report based on the kinds of results you normally get? And I think they are totally different questions from what are the risks of a false positive or a false negative result in a particular patient?

DR. CALIENDO: So let me ask both of you. This is Angie Caliendo. Would you put as a special control that you could -- you would have to be able to choose whether you wanted the resistance component of a multiplex?

DR. RAND: I think you already do. Ken Rand. Sorry. I think physicians have to be able to order tests separately as part of CLIA or something like that. So that complex they're mixing. The combination of the tests have to be available both individually and as a group. And then most hopefully, one has the software to suppress the result that was not ordered, if it was ordered separately.

In our institution we would probably make it detection and

sensitivity as a waterfall test. So, in effect, if it was positive, we could then look at the rifampin result. That's how we would probably handle that.

DR. CALIENDO: This is Angie Caliendo.

Sally, do you want to us to go down this road or not?

DR. HOJVAT: I'm not sure whether you can separate out in some of the tests.

DR. CALIENDO: That's what I'm thinking too.

DR. HOJVAT: Unless we say, you know, we wanted to block anthrax because it's not going to be something. But I think, in multiplex testing, you don't have the option.

DR. HENRICKSON: No, no, that's not true. All of the assays that have been developed in the last 20 years have been multiplex assays, and I can tell you that it's exactly as Dr. Rand said, that most places in our country it's a requirement for physicians to order tests individually or the hospital has to agree to a multiplex panel and vote on it, actually, as the test.

But there's certainly lots of political push for physicians in most places to order tests individually, and then the laboratorian gets to the choice of running them if they have a multiplex test, and if they ordered those five respiratory viruses, they can run that. If they didn't order the sixth coronavirus, then they have to not report that and try to develop methods to not look at it.

We certainly do that in our respiratory virus tests. If someone

doesn't order hMPV and it's in our multiplex, it's blacked out and we don't look at it, unless they come back and insist doing it.

DR. CALIENDO: Angie Caliendo.

Kelly, these are lab-developed tests?

DR. HENRICKSON: These are laboratory-developed tests.

DR. CALIENDO: Right, because if a manufacturer put something out there, we may or may not have the ability to block one component. I think that's what Sally was saying.

DR. HOJVAT: That's what I meant.

DR. CALIENDO: Yes. Okay. So we're going to have to take a break. Anything else on false negatives and false positives?

DR. DELLA-LATTA: I again just want to say one thing. Phyllis Della-Latta again.

What I'm trying to bring out is I think we should decide whether we want parameters of special controls to be either highly sensitivity or highly specific. What is more important in terms of drug resistance? And we discussed this before with the NAAT test in general, for detection, yes/no. But now let's just think about what's most important in terms of these resistance factors, the sensitivity or the specificity, because we're going to have false negatives and false positives either way.

DR. CALIENDO: Okay. So with that, let's take a break. I want to remind Panel members, you're not to chat amongst yourselves or to

anybody in the audience about what we're discussing, and we need to be back in 10 minutes, so that will be 3:20. Thank you.

(Off the record.)

(On the record.)

DR. CALIENDO: Can we please cut the lovely music? Thank you.

Okay. So, Panel, here's the situation. It is 3:20. We have until about five o'clock. Anything we do not complete, we can take home as homework, and I feel that I am too old for homework, so I would like to get through as much of this as we can.

So what I would like to do for about 15 to 20 minutes is talk about performance characteristics for resistance testing and then go around and have everyone give their opinion on how they feel about reclassification. We'll then summarize what we've discussed about special conditions because we've already brought up quite a few important issues and then see if we can't spend at least 30 or 40 minutes on the IGRAs so that we can, as a group, get each other's thoughts on that. I think that's a little bit different than the molecular assays that we've been talking about -- and get as much done for the FDA while we're here as we possibly can.

So let's start with what people's thoughts are on using spiked specimens. Again, resistance testing, prospective specimens versus storage specimens and anything, any comments you have about performance

characteristics of the test.

DR. HARRELL: Angie, Lizzie Harrell.

I don't know a lot about this area, but I would say, knowing that the resistance rate is so low in the United States, if we don't use spiked and/or storage specimens, I don't think we're going to get enough data to make any decisions.

DR. HAMMERSCHLAG: Hammerschlag.

You know, it's low in some areas but, you know, and for -- we offer assistance in Brooklyn because of our Haitian population, is actually pretty high. So I mean, maybe if it's targeted to certain populations, but I still wouldn't see any problem with using archived specimens from other locations. That can certainly be done to use, to validate the performance of the test.

DR. GINSBERG: Ann Ginsberg.

I mean, I do think one potentially really important role of either archived or spiked specimens is being sure that something that's called resistant really is genotypically resistant because the phenotypic methods aren't perfect. So if you don't have strains that have really been genotypically characterized by high sequencing, for example, you're not sure what you're comparing to, so there's an appropriate role for those specimens, to be sure.

DR. TOWNS: Michael Towns, Industry.

I agree with the combination of looking at fresh as well as

stored and spiked because to obviously have enough of an end that you're going to be able to make reliable considerations with regard to performance of these assays, so I agree with that approach.

DR. CALIENDO: Angie Caliendo.

And I would make the comment about specificity. So we spent a lot of time talking about false positives keeping us up at night and, you know, there shouldn't be a lot of difficulty in getting plenty of fresh specimens for specificity testing.

Yes, Phyllis.

DR. DELLA-LATTA: Could you just clarify what you mean by spiked? What would be the matrix that you're spiking into?

DR. CALIENDO: Well, we can discuss, throw that out. I think it would be the matrix that the test was intended to use.

DR. DELLA-LATTA: Good, okay.

DR. CALIENDO: Oh, my God. I'm starting to talk like the FDA.

(Laughter.)

DR. CALIENDO: Any other thoughts on -- let's -- I think specificity, we all think it needs to be as high as it possibly can be.

Go ahead, Michael.

DR. TOWNS: Michael Towns, Industry.

I agree. Given that the prevalence of MDR is so low in the U.S., and I know Dr. Della-Latta made the comment about concerns about false

negative, but given that it's so low, I think the premium on the genetic resistant marker testing should be placed on specificity, and I think it should be very high, akin to what we were saying with direct detection. And so I would put more emphasis on specificity given that we have such a low rate of MDR in the U.S. to prevent unnecessary reporting of false positive results in the U.S.

DR. CALIENDO: This is Angie Caliendo.

Any other thoughts from the committee?

(No response.)

DR. CALIENDO: So let me rephrase that.

How low are you willing to sacrifice sensitivity? Is there any reason to think we can't have both or shouldn't try to get to both?

DR. HENRICKSON: No, we should -- Kelly Henrickson.

Since it's only going to be on positive -- I mean, amplified specimens that have demonstrated to be positive in another -- then it seems logical that it would be unlikely that you couldn't have high sensitivity and high specificity.

DR. DELLA-LATTA: Phyllis Della-Latta.

I have to echo. I think we need to have pretty high standards here with this first assay that's going to be around for detection. I think we have to really set our bar pretty high.

DR. CALIENDO: Angie Caliendo.

Phyllis, what does pretty high mean to you?

DR. DELLA-LATTA: I'm talking in the 90s.

DR. CALIENDO: Okay. Karen.

DR. CARROLL: I'd like to echo what Kelly said. I mean, I think, you know, we know that we're going to be applying this test to people who are already positive, we think, right, for another assay. So I think that changes things and I think we -- I agree with Phyllis. We should strive for sensitivity well above 90 percent.

DR. TOWNS: Michael Towns, Industry.

I would concur with the sensitivity requirement of, you know, greater than or equal to 90 percent, but I think that the specificity needs to be relatively high. But that would be -- my premium would be specificity over sensitivity, but sensitivity should be relatively high as well. But greater than or equal to 90 percent.

DR. CARROLL: Karen Carroll.

I guess we should qualify and say greater than 90 percent for the resistance determinant that the manufacturer claims, not greater than 90 percent for resistance to that drug. I see that as two separate evaluations.

DR. CALIENDO: Yes, Phyllis.

DR. DELLA-LATTA: I would recommend a couple of things, one of them being that when an analysis is being done, when we're considering this particular device, that the comparators be not only phenotypic but

sequencing. I have done that in the past for microarray analysis, and you can get separate results. You'll also find very interesting results about the mutations, which would be interesting moving forward.

And the second point I wanted to make was I think we should be able to use sediments from other countries, particularly countries that are immigrants here, that are foreign-born, that are living here in the United States. I have worked -- just as an example. I know some anecdotes are important in these particular considerations.

And so I have a lot of -- that are multi-drug resistant, but many people are from Mexico, if they are from Asia, we should be able to use them in our laboratory, sediments, and I think this would really help out in being able to determine the performance characteristics of the assays.

DR. CALIENDO: So this is Angie Caliendo.

Ann, you had mentioned genotypic, the importance of sequencing as a comparator, not just phenotypic. I'm seeing some other heads shake forward, so it sounds like that's a performance characteristic that would be quite critical.

DR. BEAVIS: Angie, can I comment about sediments?

DR. CALIENDO: Absolutely.

DR. BEAVIS: I don't know if you're purposely saying sediments versus specimen.

DR. CALIENDO: I am.

DR. BEAVIS: Okay. Because I think it's important that if this test is designed to be run on specimen, that the testing and the performance be established on specimens. I think that by doing sediments you're changing the matrix of the specimen, and I don't know that the performance characteristics would be the same. The viscosity changes often with freezing -- this has been mentioned before -- inhibitors that would be present in a sputum specimen are therefore not present, and I think that's going to change some of the performance characteristics.

DR. DELLA-LATTA: I'm confused. You're saying sediment is what -- not specimen, for specimen?

DR. BEAVIS: The test is designed for specimen, and I think that as many studies as possible should be done on specimen, not on sediment because sediment, to me, is a different --

DR. DELLA-LATTA: But sediment is what we're using on all the NAATs.

DR. BEAVIS: Yes, it is, but --

DR. DELLA-LATTA: And so why don't you want to do that?

DR. CALIENDO: Angie Caliendo.

So I would interpret it as this, if I -- because I don't think we can decide this, the manufacturer will, but I think what you're both saying is if you're going to use sediments, you need data on sediments. If you're going to use primary specimen, you need data on primary specimens.

DR. BEAVIS: Okay.

DR. CALIENDO: Is that a fair --

DR. BEAVIS: Fine.

DR. CALIENDO: Because we won't ultimately decide what the manufacturer decides to bring to the FDA.

DR. WILSON: Actually, Angie, the question is antibiotic resistance in tuberculosis is complexed directly from respiratory specimens. We're not supposed to be addressing the issue of sediments in this.

DR. CALIENDO: I would say that that's probably hair splitting, how the FDA views it. I think what they wanted was something not off of culture, if I understood correctly. Sally, is that correct?

DR. HOJVAT: Yes, that's correct.

DR. CALIENDO: Okay.

So let's summarize performance. So what we have here is patient group needs to be considered, whether they're high, low, or moderate risk for resistance. Everyone -- seems consensus to support the use of spiked and stored specimens, consider using specimens outside -- from countries outside the U.S. where there's a population in the U.S. that has resistance.

Look for areas of high level of resistance. There are pockets in the U.S., and that might be good places to set up studies in addition to places with low prevalence.

We need genotypic sequencing confirmation of anything stored that was chosen to use, but it's also the comparator that needs to be done is both phenotypic and sequencing.

Looking at a sensitivity of at least 90 percent, a specificity of at least 95 percent. And the sensitivity and specificity are for the resistant determinant, not to measure for the drug in general.

And then again, it's specimen specific. If they come, a company chooses to come, with a primary specimen, that's what the clinical trial needs to be designed for versus sediment. They would be viewed as a different, definitely a different specimen.

Have I missed anything?

DR. BEAVIS: Kathleen.

I'm really sorry to bring this up at this point. It's not in our briefing. I didn't read it before I came. I don't know, off the top of my head, what the performance characteristics are of the currently available ones that are out, the Class II ones, the MGIT, the other companies as well that have susceptibility testing. And I would hope that the performance characteristics of this would meet or exceed those. I just don't know if it's 90 or 95, and I didn't think to look it up before I came.

DR. CALIENDO: Okay. I think that's a very fair statement. Does anyone know that off the top of their head?

(No response.)

DR. CALIENDO: Okay.

All right, so what I think we'll do now is go around. We'll start with Gary and just give your thoughts on whether or not you would reclassify, and then we can discuss what those controls, what those special controls would be.

DR. DUERHRING: This is Gary Duerhring, Consumer Rep.

I would consider reclassification.

DR. TOWNS: Michael Towns, Industry.

I'm in favor of reclassification because I think that the special controls that should be in place with the reclassified effort would be very helpful in risk mitigation.

MS. DeLUCA: JoEllen DeLuca.

I would vote -- I would be in favor of reclassification. It feels like it's coming together.

DR. CARROLL: Karen Carroll.

I'd prefer to talk about the special controls before I decide on whether or not to reclassify.

DR. CALIENDO: Angie Caliendo.

Do other people on the panel feel that way? Because we can reverse this conversation.

UNIDENTIFIED SPEAKER: Yeah.

DR. CALIENDO: Okay. So let's do it.

So let me tell you what I have taken from the group so far is you would need postmarketing monitoring of the mutations. Ken gave us some very good approaches to that. You could either sequence specimens that came up as negative in the resistance test that were positive by TB and confirm that they truly don't have a mutation that the assay misses.

Monitoring CT value, I thought was a clever idea, from a population perspective.

So there are a couple ways that they can be done. I think GenBank is another way. We're used to RNA viruses that have a whole lot more genetic drift.

One of the other things that Kelly brought up that I thought was interesting was, as a postmarketing approach, is sending off any positive specimen to the CDC for confirmation or for sequencing or setup somehow in the public health system. I know the person who spoke representing the public health labs talked about the lack of potential resources for that, but I think there is value in that.

Resistance testing not to be done until you've confirmed or at least detected TB by either culture or another molecular, another NAAT test. So then you would go back to the primary specimen or whatever specimen the assay's confirmed for or approved for or cleared for, to do the resistance testing.

You would consider setting up the clinical trials where they

were testing more than one specimen, and I think this is something that Karen brought up, to help mitigate the whole issues with the false positive results. And I think, as we talked with NAAT testing, it would be a very helpful way to design the clinical study. And then you must do phenotypic testing in addition to the sequencing.

Any other special controls that I have not picked up on? Kelly.

DR. HENRICKSON: Kelly Henrickson.

I would recommend a black box or something in the special controls that says only to be used in high prevalence populations. I think, I mean, if you use it in low prevalence populations, you're going to get more false positives, so I think you should restrict the use of it to high -- or you recommend the use in those populations that you think --

DR. RAND: Can I comment on that?

You may not know -- your overall -- you may know your overall population, but you may not know the true population risk of the individual specimen sent to your lab.

DR. HENRICKSON: I think the physicians have to give that to you.

DR. CALIENDO: This is Angie Caliendo.

Another way to look at that is you give this information to the physicians and they put it into their matrix of information. Just another way to frame that.

DR. HENRICKSON: But then you test 99 negatives out of every 100.

DR. CALIENDO: Um-hum.

DR. HAMMERSCHLAG: And I think this is the same -- Maggie Hammerschlag.

The same problem applies to any test in a low-prevalence population. They've got to be aware of the fact that you're going to have those little predictive value problems and it just -- it can be more and then just say interpretation of the results, be aware that, you know, if there's low prevalence. On the other hand, many of the patients, of course, that we deal with, you're going to be dealing with, are going to probably be coming from high-prevalence populations, the individual patient.

DR. DELLA-LATTA: Phyllis Della-Latta.

I really don't think it has a place here, as Maggie had said. That sort of thing is a laboratory director decision in terms of what tests are brought onboard depending upon your patient population, the prevalence of disease, et cetera, et cetera. And if it's not present in any other test, it's not present. We didn't recommend that for the NAAT test when we went through it. I don't think it belongs here.

I think, when we do have positives, in my patient population, I do have to be concerned about whether those are multi-drug resistant or not.

DR. GINSBERG: Yeah. Thank you. Ann Ginsberg.

I was just going to say I think -- I don't think it belongs as a black box warning because I think the relevant question is whether the patient is high risk, not whether the overall population is high risk, and that's really a clinician or a lab director decision.

DR. HENRICKSON: Yeah. I mean, we give these warnings -- Kelly Henrickson.

We give these warnings right now for antigen detection tests, right? We say, you know, we give the warning, saying that if you're using antigen detection tests during a low-prevalence time period, their positive predictive value is very poor, okay. And so it's just that kind of warning we need to tell people. I think there should be stuff in the special controls that warns them that --

DR. CALIENDO: Yeah, I think -- Angie Caliendo.

I think the reaction may have been the term "black box."

DR. HENRICKSON: Excuse me.

DR. CALIENDO: So I'd put that under the education pile, where people really do need to be educated what the results of these tests mean. So I totally understand that.

Karen.

DR. CARROLL: Well, I'm just speaking to Kelly's point. I think -- Karen Carroll.

I think this is going to be specified in the instructions to the

manufacturer with respect to the design of their clinical controls, right? So that language should automatically show up in the performance characteristics of the test as a condition of the clinical trial, and we already talked about the fact that we need fresh specimens in low-prevalence populations as well as spiked samples from populations where we're likely to see those resistance determinants. So I think it will be addressed through the clinical trial.

DR. CALIENDO: So back to Dr. Carroll. After this discussion, anything else you want to say about special controls?

DR. CARROLL: No, I think you've covered the special controls that I feel strongly about.

DR. CALIENDO: Okay. So back to your opinion on, or your sense -- not your vote -- on reclassification.

DR. CARROLL: Yeah, I guess I'm still very concerned about it, but I guess my feeling is if the special controls really solve the mandate as specified in the Class III classification -- and I deliberately pulled this out so I could actually address this.

If our special controls can accomplish what currently the Class III regulation accomplishes, which is that we can get sufficient information in the de novo application to satisfy the safety and effectiveness of this testing, that I would favor that.

DR. HARRELL: Lizzie Harrell.

I agree with what Karen said.

DR. GINSBERG: Yeah, Ann Ginsberg.

I would certainly encourage the Agency to consider reclassifying from Class III to Class II because I think there are special controls that can be put in place, as we've discussed, that can mitigate the risk.

DR. JANOSKY: Janine Janosky.

Support the reclassification to Class II.

DR. WILSON: Mike Wilson.

I think there's no way you couldn't do this because I can't imagine being a manufacturer and being in the position of having one device that had Class II and the other part of the same device had Class III on it. I mean, these tests are meant to be used in conjunction, so if you're going to recommend that the detection be Class II, you almost have to recommend that this also be Class II.

DR. CALIENDO: Angie Caliendo.

I think this is less straightforward for me than NAAT, but I do believe that the special controls that we've put into place, the risk can be mitigated. I think it's very important on the design of this clinical trial, and I think that low-prevalence part of that clinical trial is going to be very important for those of us in the U.S. that are going to be using this on a day-in and day-out basis. But with that special control, I am comfortable with bringing them down to a Class II.

DR. RAND: Ken Rand.

I also favor bringing it from Class III to Class II, but again, it's critical to do the studies in both high prevalence and low prevalence because you need to know what you may be missing in a high prevalence as well as, you know, what you're over-culling in the low prevalence.

DR. BEAVIS: Kathleen Beavis.

I also favor the reclassification from Class III to Class II and hope that in the clinical trials that the comparators will be commercially available methods that we have access to in the United States so that we can see how the proposed test would do with that, you know, obviously other methods used for discrepancy analysis, whether it's sequencing or something else.

DR. GORDIN: Fred Gordin.

Nothing more to add. I would agree.

DR. HAMMERSCHLAG: Maggie Hammerschlag.

I agree.

DR. REIMER: Larry Reimer.

I'm going to be a renegade, and I don't think we're ready yet. I don't think there's enough information at this point, either, in terms of the global construct of resistant mutations in what's going on with TB or that we know enough about it to do this right now. Even though I agree with Mike Wilson about it's hard to do this for one part of the testing and not for the

other.

DR. DELLA-LATTA: I agree this is not that straightforward, but I do agree with the reclassification. I think that we have nothing to lose, it's something we really need, and we'll be gathering data as we do the test. I don't want to discourage testing and tests to be performed, and I think we'll get a lot of knowledge out of this. And I think it will be strengthened and will be just as good as a PMA.

DR. HENRICKSON: Kelly Henrickson.

I think it's safe to reclassify.

DR. HAMILTON: Carol Hamilton.

I agree. I think it should be reclassified to Class II.

DR. CALIENDO: Okay, thank you.

Steve, we're all going to take a deep cleansing breath while you read Question 3.

DR. GITTERMAN: For immunologically based tests such as IGRAs that are intended for the detection of tuberculosis infection by indirect means:

(a) Please discuss the risks associated with inaccurate test results for the detection of tuberculosis infection (that is, false-positive and false-negative results).

(b) Please discuss special controls to mitigate each of these risks, including the following:

(i) Clinical studies that would be appropriate for documenting device performance, and

(ii) Minimum device performance standards (for example, sensitivity and specificity) that should be recommended in guidance; and

(c) Based on the above considerations and the earlier presentations, please discuss if sufficient risk mitigation is possible for FDA to initiate the reclassification process from Class III to Class II devices for this use through drafting a Special Controls Guidance.

DR. CALIENDO: Okay, so why don't we start with the risks associated with an inaccurate test, false positive/false negative?

Go ahead, Maggie.

DR. HAMMERSCHLAG: I think that this obviously, again, is related to in what clinical study are you using it. Mostly, I guess, it's going to be for the diagnosis of possible latent tuberculosis, so is it a question of talking about people who are exposed that you're evaluating in a contact tracing, is it being used in an employment, as an employment screening, like, you know, we -- they're trying to shift away from that.

So most often, I guess most adults and let's say using it at employee health, let's say at Downstate University Hospital in Brooklyn, the probability usually of missing it if somebody isn't sick is pretty low, I think, at

this point, the risks. The ultimate risk will be in the long term, of course, of potential reactivation if they're not, you know, prophylaxed.

You're dealing in children, then, of course, there's a much higher risk of reactivation of disease with air, contact, et cetera, although we tend to be more aggressive, and if we do have a positive contact, well, generally these kids will prophylax regardless, even if they're negative, little kids.

So it's hard -- you know, it's a moving target.

DR. DELLA-LATTA: I think, along those lines, I wanted some clarification as to what we're thinking of in terms of which population. Again, are we talking about employees, are we talking about active disease?

The second point is the comparator or "gold standard" is so poor that I was asking if there were another type of assay that was brought before the FDA that had a similar type of situation, you know, whereby you're comparing one test against something that's dismally poor.

DR. CALIENDO: Sally, can you comment on the IGRAs that are approved because they use that dismally poor gold standard and how the Agency looked at that?

DR. HOJVAT: Long and hard.

Freddie, do you want to make a comment on this? You were more involved with it than I was. It took us a lot of work to get through these two. But there did seem to be a need in terms of, you know, latent TB, so --

DR. CALIENDO: Angie Caliendo.

Is that what you want us to be thinking about, latent TB, or do you want us to be thinking about this as a diagnosis of active TB?

DR. HOJVAT: Well, I think it's just TB infection.

DR. CALIENDO: TB infection.

DR. HOJVAT: Yeah. But the tendency is more towards the latent.

DR. CALIENDO: Latent, okay.

DR. HOJVAT: We did include both, if I recall.

MS. POOLE: Freddie Poole, FDA.

When we reviewed them, we had lots of discussions. We also had an advisory committee meeting and got input from our advisory committee on -- because of that fact, yes, there was no other test but the skin tests that were available at the time. We also looked at the IGRAs in different populations beside high risk, low risk; we did -- there were healthcare workers, they were compromised patients, they were all different types of populations that we looked at, and we did believe that the data that we had was sufficient, and the panel agreed and they recommended approval.

But we did lots of studies and we reviewed lots of studies that were done and accepted what we had as the best possible standard. And the cultures were also done on the patients in high-risk settings, so they didn't

only depend on the skin tests; they also used culture to make a decision of whether or not the tests would be for the sensitivity and specificity of the tests.

DR. GINSBERG: Sorry, Ann Ginsberg.

Could you clarify that, in latently infected individuals, what were they culturing and how were they interpreting --

MS. POOLE: Okay. What they did was skin tests, but it was the people in the high-risk population that cultures were also done on, to help strengthen the -- if the test was positive and that was also like a supplemental test was done on cultures from the people that were in the high-risk population.

DR. HAMILTON: This is Carol Hamilton.

They were looking at -- they used them in patients who were TB culture positive.

MS. POOLE: Um-hum.

DR. HAMILTON: So then they looked at performance of TST and IGRAs in patients who were culture positive as well. And other populations were those who were at high risk epidemiologically but did not have TB.

But just to remember that, you know, we know that a substantial number, 20 percent or more, of patients with active TB will be TST negative, they have active TB, culture positive, but we know that.

This is Carol Hamilton again. I just wanted to say that in terms of the -- I mean, we have the data in our packets, I can't put my hand on it right at this second, but these -- okay, the tests actually, the IGRAs do not perform, they're not particularly more sensitive than TST. I mean, TST is in the 70 percent range of sensitivity. And in patients who have not had BCG and are not living in an area with a lot of MAC, the specificity is, you know, not great but it's in the 50 to 70 percent range. And that's where the IGRAs are more helpful, is that they tend to be more specific. They really don't add much in terms of sensitivity.

DR. GORDIN: Fred Gordin.

So, again, I think the data we saw was roughly 4 percent of the people in the U.S. presumably have latent TB and in some populations, less than 1 percent. So we all know we need a better test, but these tests really do have risks involved, which they wanted us to talk about.

Over-diagnosing, we know from people, even in this 8,000-person research study that the TBTC just finished, people even taking only a 12-week regimen, 20 percent did not complete it. This is in research settings. So the amount of effort going into trying to find people, if they don't have TB, if these tests are false positive, is really wasted as well as the toxicity.

This compared INH to INH in rifapentine. And while there was no -- very little major toxicity, there was a fair amount of minor to moderate toxicity that people dropped out of the study, et cetera. This is a well-done

study.

So I'm just saying that there is real harm in over-diagnosing in a low-incidence population, people with latent TB, so we clearly need much better tools. And, again, in terms of the -- so that's kind of a wishy-washy statement, but I guess I just wanted to point out, again, where there's a low incidence, we really are going to be over-diagnosing and putting people on drugs they don't need, et cetera.

I'm slightly -- I mean, it would be great to have, also, a more sensitive tool as well, although I'm actually a bit more worried about over-diagnosing. I think Joe Bass and others published that, you know, in low-incidence settings, probably well over half the people that have a positive PPD don't have tuberculosis in their body and it, you know, may or may not be the same for some of the IGRAs out there.

MS. DeLUCA: JoEllen DeLuca.

I think one of the things that we're sort of going around the edges of is much better settings for diagnosis. If we get into the realm of testing everybody or going through, particularly going through the workforce, we're also putting people in jeopardy for being denied mortgages on -- possibilities -- being denied Social Security and Medicare disability and maybe even a job, so these are things that have happened with people with HIV/AIDS, and I think we should remember that people's perceptions of disease is far greater than the actual reality.

DR. HAMMERSCHLAG: Hammerschlag.

One thing about these tests, the IGRA as opposed to the skin test, is that it supposedly, it's objective. And one of the concerns driving its adaptation, for instance, by the health, employee health service is the fact that they were tired of trying to chase people down to read the PPD.

And then, you can get false negative PPDs because it isn't done correctly, and there is a subjective component in reading the PPD, and I've seen that get screwed up quite a few times. So that the feeling about this test as well, it's, you know, you get a readout, it's not going to be subject to the same interpretive problems or having to chase down your reader.

I mean, it used to be that, as -- I get PPD tested every year. Used to say that they would let doctors read their own and they got rid of that because a lot of people just -- well, there was obviously concern about truthfulness but, you know, you have to go back there.

So many times, sometimes tests are selected not so much because the performance parameters may be better, but because of the ease in administration and interpretation so that you have to put that into that context, is using it.

Now, but I -- but we have -- I think there are two approved, at least two, probably three, approved tests right now, and the question is, at least from what I remember, also reading the package, the various studies on them, is that they're -- aside from the fact that we do know, Number 1, in

children it's the same issues, is that there doesn't seem to be great reproducibility between the three tests, which brings up a whole set of other issues about that they're not maybe totally equivalent and we don't know exactly why.

DR. TOWNS: Michael Towns.

These tests are actually different from what we've been talking about today in that they're what I would consider sort of hybrid and that they can be used for screening, asymptomatic, for LPBN as well as an aid in diagnostic purposes for patients that may have active pulmonary TB, so there's actually sort of a need to have relatively high sensitivity and specificity. But unfortunately, we don't have a good reference method for LTBI to really judge what the sensitivity of these assays are for that application.

But we do know that as long as they're used for that purpose in the appropriate population, someone that's at least moderate to high risk of having maybe have contracted tuberculosis and need appropriate follow-up, then we would have to have good, you know, good performance.

I think, on the diagnostic part, they're an aid to the diagnosis of patients with active pulmonary TB, but I think the specificity should be appropriately elevated, like at least 95 percent for that particular application, the specificity there.

So I look at them as kind of a hybrid, but I don't look at these

assays as having any incremental risk above what we currently have been using for like a hundred years, which is the skin test, basically, and they do provide, I think, benefits that exceed their risk and that they -- patients that have been vaccinated for BCG and maybe are immune-compromised, they're not affected, these tests results are not affected by those conditions, so I think they do have the benefit there as long as they're used in an appropriate population that's at least moderate to high risk of TB, and I would be, you know, in favor of looking at, you know, reclassifying those.

DR. GINSBERG: I think I want to raise the question that Michael -- the point that Michael just raised about the context of how these would be used because they are significantly more specific than TST. So, you know, you worry about the false positives, but it's actually you'd be doing much better with the IGRAs than with TST. So we have to keep it in perspective.

DR. HAMILTON: Carol Hamilton.

And the points have been made that the IGRAs are more specific than the TSTs and the problem we've had, as you -- as Dr. Burman showed in one of the slides earlier today, you know, the majority of active TB in the U.S. today is in foreign-born individuals, so we would like to have a way, a tool, so that we can try to get more people who come to the U.S. are from high-burdened countries and we can get evidence that they have latent TB infection, we would like to treat them with a preventive therapy regimen.

But right now, they do not trust the results of the tuberculin

skin tests, so they've been told, from home, don't let them put that test in your arm because it'll show positive because you had BCG vaccine and it doesn't mean anything. And so very few people in that situation will accept taking those drugs. And so one hope is that with these more specific tests, that then we can get more of this high-risk population on treatment.

So, anyway, just to make that point. And I absolutely agree with what Michael and with what Ann said about reclassification.

DR. HENRICKSON: Kelly Henrickson.

I agree with all of those last three statements, and I think this whole class has already been demonstrated and approved by the FDA to be safe and effective. And so along with all those comments, I agree that I think that reclassifying from III down to II is an appropriate thing to do.

DR. DELLA-LATTA: Having come from a place that where we do have a lot of patients that have been on BCG and a lot of immune-compromised patients, we obviously have a need for it and for those patient populations. I think that reclassifying it is totally approved.

My concern is with active tuberculosis. I think there are enough good tools out there, and I really don't see a place for this, and my fear is that if somehow we do get a budget, you can do it in those patient populations I just spoke of, that it might be abused out there on the force, and I don't know how one can separate this. I don't know if it can be done.

DR. CALIENDO: Angie Caliendo.

I would have come in with that comment, but won't leave with that thought after listening to Kelly and Margaret talk about the pediatric world. I do believe that that is a place where these tests have a role diagnostically. Should that drive this decision? I don't know, but I think we have to keep that in the back of our minds. They both repeatedly made that point and I don't --

DR. DELLA-LATTA: No, but that is true with pediatrics. Fine.

DR. CALIENDO: Yes.

DR. DELLA-LATTA: But the adults, no.

DR. CALIENDO: Well, that's another --

DR. DELLA-LATTA: That's -- yeah, exactly.

DR. CALIENDO: Larry.

DR. REIMER: Yeah. Larry Reimer, Utah.

I have a perception, and I don't know whether there's any data to support this, that our clinicians believe that these new tests are magically perfect, that a positive result is a right answer; negative result is a right answer in comparison with TST testing.

And I don't know how that should play into any of this, but I'm concerned in terms of talking about false negatives and false positives and the education piece to make sure that if these tests are conveniently available, that we make sure about the education so that clinicians know how to use them and what the implications are of results in both directions.

DR. HENRICKSON: Angie, Kelly Henrickson.

If it makes you feel any better, we have a lot of discussions in our -- group about these assays and there's a lot -- there's no overwhelming sentiment of just accepting them, so --

DR. CALIENDO: So this is Angie Caliendo.

Any more comments on risk, false positive/false negative? So what we have for false positive is -- I'm sorry, what we have for false negative is that you're going to miss somebody with reactivated TB. This seems to be, possibly, a higher concern in children than it would be in adults.

False positives, you're giving drugs that aren't needed and they aren't without their side effects. And then the public perception of this disease and giving someone, potentially, a thought that they have been infected when they, in fact, have not.

Anything else on the risk side of it? Go ahead, Kathleen.

DR. BEAVIS: If I could just echo what Dr. Hammerschlag and Dr. Reimer have said about these tests. I get this in my institution as well, where people say this is obviously a better test because it's objective, and I don't buy that. I don't like just that perception that they have and, you know, I've also run into the non-reproducibility of these tests, and what I don't like is people saying well, this must be a better test because it's objective. Or the TST is so bad. We have an alternative; this is better and this should be used.

The other thing I don't like is that, you know, this test is

designed and ideally it would help us figure out among the BCG-vaccinated population who has latent TB, and we have a lot of healthcare workers in our institution who have been BCG vaccinated. And we have not necessarily seen that clarification where people will be TST positive and BCG -- no, TST positive and IGRA negative. In fact, we've often seen the reverse, and that really puts us in a quandary. So I think part of my hesitation about this area of testing is just a lot of misperceptions, a lot of lack of knowledge. A lot of non-reproducible results are driving the usage of this test.

DR. CALIENDO: So Angie Caliendo.

Kathleen, any thoughts on how you would mitigate that problem? Or not only to Kathleen, but anybody else on the Panel.

Angie Caliendo. I'll just make one comment. You know, it's interesting, at Emory, the perception is it's a better test, I think, based on the whole BCG piece. I don't know if it's perceived that it's a better test because it's an objective -- that's an interesting comment. And it's hard to know what drives people to think a test is better. Sometimes it's just newer is better, newer technology is better, even though it may not be, but I think the way it's carried is that whole specificity issue.

DR. REIMER: Larry Reimer.

And when I made my comment, it had nothing to do with BCG. It was entirely about this is a new test; it's obviously better. It had nothing to do with BCG.

DR. HAMMERSCHLAG: Hammerschlag.

But to go back, you know, the ATS, they've come out with guidelines for interpreting the PPD in the setting of a possible history of prior BCG vaccination. And given the fact that even the efficacy of BCG is all over the place and you don't know if they got a take, the general feeling is that if you get -- or X number of millimeters, it doesn't matter what the history of BCG is.

And certainly, the older you are, the less likely that any positive reaction, if it's going to be over, I think it's 10 or 12 millimeters, is going to be indicative of BCG, so I think that's almost -- that's a straw man; they've overdone that for the most part. That's not what I'm concerned about. But it's like everything else. I think it's a fancy new toy.

And they -- given, sort of, the general lack of understanding of most physicians of the technology, and this is a totally different technology, I mean, they don't understand that, if they get problems understanding the significance of this, it's like with every other test, there needs to be education so that they can understand the limitations, what the indications are, and the limitations of interpretation, and that even if you're in a high, you know, very epidemiologically, in a high-risk situation, frankly, frankly, many times we don't care what the test shows.

DR. HARRELL: Lizzie Harrell.

I was going to ask Dr. Beavis, at your institution does the

microbiology laboratory do the test?

DR. BEAVIS: No. We send it out. We do not perform the test in our institution. We do it as a send-out test.

DR. HARRELL: Okay. Because that's one of the questions that we've had over the years, of looking at the test and trying to incorporate it into our laboratory. It just didn't fit our workflow that well. So I just wondered if others were doing it in-house or sending it out.

DR. HAMMERSCHLAG: It's an interesting point because I'm not sure in ours because it does -- you know, it is an overnight thing. You've got to incubate the cells, and there's always issues there about adequacy probably of sample collection. I mean, there may be other facts, technical aspects involved that will impact on the performance -- performing of PPD that have nothing to do with the actual accuracy of the test.

DR. CARROLL: Karen Carroll.

I just want to get back to your question to us, to add to the implications of a false positive. It's not really just -- not necessarily for the individual patient, just the risk of unnecessary drug toxicity, but other things that happen to that patient in terms of chest X-rays and monitoring with laboratory tests. I mean, I think it really can increase the cost of healthcare unnecessarily.

And getting -- in responding to Lizzie's question, we send this out as well, and if it were to be brought in, it would probably go to the

immunology lab.

DR. HAMILTON: So just again, to remind people that -- I mean, a lot of this discussion has to do with people who are in-patients who are ill, but the major -- a major place of this issue, of course, is in the public health setting, in clinics, and the big problem there is people come in, they get tuberculin skin testing, and then that person has to show back up again two to three days later. It's not that they work there; they have to somehow get there. So right away, you have this huge falloff of people who never show up, so you have no idea what that TST showed, which you all know. And, anyway, so it goes down from there. So I think, in that sense, we have to remember that these tests do offer that advantage.

In terms of controls that might mitigate the issues, of course, the combination ATS-CDC-IDSA guidelines are actually being worked on again to -- that will come out and will come out regularly, that talk about how to interpret these tests, and these tests, in association with tuberculin skin testing and associated with epidemiologic risk, et cetera, and so there will -- there are and will be guidance, you know, that will address that.

And, of course, the whole goal of all this is that it will be easier for companies to, you know, come in with hopefully competitive tests that are even better.

DR. GORDIN: Fred Gordin.

I mean, what Carol said is frequently talked about, that you

don't have to return. I just don't buy it. If somebody with a large bump on their arm won't return, what is the chance they're going to take 12 weeks, 6 months, or 9 months of therapy?

So in terms of the cost of the test, a learned result versus the action that's going to follow, I mean, if we had a better test than the TST, without question, I mean, one of the reasons at least our hospital hasn't switched over is because of the literature, that it seems to be no better, it's more costly, and again, the mindset that if they won't come back for that one visit, you know, what is the chance they're going to complete it.

I don't know that that ties in with this question at all, sitting out here. Again, I still think we get back to real risks of over-diagnosing, but we're probably under-diagnosing, under-treating, and as Ken pointed out, the whole idea to eliminate TB in the country is to go find those millions of people with latent TB and do something, you know, active so yes, we need better tools.

I guess I'm still leaning towards this shouldn't be reclassified because there is no gold standard, and we need, you know, the best data possible, but if, again, the FDA thinks that having it as a Class II with all these caveats built around it, you actually get the best data, then, you know, maybe more companies will come in, et cetera.

DR. CALIENDO: So let's -- this is Angie Caliendo.

Let's spend a few minutes talking about Question (b), which is

how would you design clinical trials to address some of these issues because you're -- Carol's been talking about the low-risk person that's being screened for employment or -- which, our place still does. I mean, many places still do.

So what would these trials look like? We talked about kids. What would the trials look like there? What would the trials look like in immune-compromised patients where we've actually been frustrated with the performance of both PPD and the IGRA? So let's talk, let's give them a little bit of information regarding clinical trial design.

DR. DELLA-LATTA: I just have one question. In terms of specific controls, I don't know if it would be possible, but in terms of reclassifying it, just to get back to what we spoke about before, is it -- and we're concerned very much about false positives, and you obviously don't want to have a false positive IGRA. And if that does happen, you do get a positive reaction.

What is the -- how do you feel about the idea, in terms of the Panel, of maybe having a suggestion that a TST be done on those that are positive or those that are thought maybe to be a false positive in order to sort of have a confirmatory test or -- I don't know. I'm just throwing it out there, particularly for those that are users of the IGRA.

DR. HAMMERSCHLAG: Hammerschlag.

I think, actually, you can get indeterminate results with this test, too.

DR. DELLA-LATTA: Yeah, that's -- yes, yes. Yes.

DR. HAMMERSCHLAG: And I think there's definitely a recommendation that you do a skin test. But there have been studies that have been done in children, actually, that have demonstrated problems. I mean, studies have been done. It may be an inherent problem with just the immune development of children. I mean, theoretically, it's going after the same kind of a principle, and it's not perfect to begin with. I mean, it may be impossible to tweak it or make it any better than it is. And it is what it is, and we have to understand the limitations and how we're using it.

DR. DELLA-LATTA: So confirming the positive wouldn't be a good idea in your --

DR. HAMMERSCHLAG: I don't know in this situation or not because --

DR. CASTRO: Thank you.

Just as a point of information, the existing guidelines discourage the use of one test to confirm the other because all of the tests have shown, at best, about 80 percent concordant, so you're going to have 20 percent discordant cells no matter how you look at it.

Now, the one instance in which the existing guidelines say use either or both tests is if you have a high pretest probability index of suspicion, I think this person has latent infection, and if latently infected is in the road for a bad outcome because they're each having effect and et cetera, then use anything available to you.

That's the only situation which we recommend using both, but the rest of the time we discourage the use of one test to confirm the other because of the available body of work that shows the 20 plus percent discordant cells.

DR. CALIENDO: Thank you, Dr. Castro.

DR. WILSON: Mike Wilson.

In terms of possible clinical studies, some of these are actually under way or being done right now. We are just wrapping up in collaboration with CDC a large multi-center clinical trial looking at latent TB. Many, many thousands of tests. Those data will be forthcoming soon. And there have also been some studies recently looking at immigrant populations in high-risk areas.

Some of the preliminary data, for example, are -- and if you look at children coming in from high-risk areas who get a skin test about -- with a positive skin test, you know, fully 30 percent of the kids do not have TB and pass. We give chest X-rays and get treated and so forth.

So a number of these studies are actually under way and they're being funded through CDC through a series of task orders, so in terms of the clinical trials that need to be done, I actually think a lot of this work either has been done or is being done now.

And to be honest, because of the public health implications of skin testing, IGRA testing, TB control, this probably is not something that can

readily be handled by a manufacturer. It's probably best handled in the public health arena.

DR. CALIENDO: Other comment?

DR. HENRICKSON: Kelly Henrickson.

Angie, you asked for -- this is for future tests that we might be looking at, interferon release. At least that's what the guidance documents and the special controls are for.

So I guess I just -- it's fine if these other postmarketing tests are going to go into how they write the special controls so that future tests can be developed with the same kinds of information and dividing them between what's needed for the 510(k) versus what's needed postmarketing.

But I guess I would suggest there have been some pediatric trials, but not a lot, okay, and so I guess I would still recommend pediatric trials where they run their new test against FDA-approved tests. And I would suggest these trials be in children who have clinical diagnoses of TB and probably expect that they will be taking place in foreign countries. So I guess I would -- because we only had 26 TBs in Wisconsin last year; that's all.

DR. HAMMERSCHLAG: Well, Hammerschlag.

What about the -- here's another issue. What about a child, a young child, who's in contact of an adult with active tuberculosis? That's another extremely high risk at the timing.

DR. HENRICKSON: There's lots of studies that should be done.

I was just recommending one that I think will help define what it means, is if we look, if we have data on children who have, who we think clinically have active TB. And under 1, under 5, you know, different age groups of children so that we can have that data.

DR. REIMER: Larry Reimer.

Except that people with active TB are -- respond immunologically different to mycobacteria than people who have latent TB, so I'm not sure you can translate one into the other.

DR. HENRICKSON: Kelly Henrickson.

I'm not trying to translate. We use this test for diagnosing TB and treating TB, so we need this data in pediatrics.

DR. HAMMERSCHLAG: Well, Hammerschlag.

We don't -- it's an adjunct. I mean, we try to go for -- if we have a child who has, you know, evidence, they're positive by whatever tests we're currently using and we do a chest X-ray, then we try to chase it down. So we do not primarily use it. It's more supportive than anything else, if we think that there's active infection, and we definitely have had more than 26 cases in Brooklyn, which is not even the whole state of New York.

DR. HAMILTON: Carol Hamilton.

So just to remind you that in the Section K of our packet are the updated guidelines for using IGRAs, and it was published in 2010 in the CDC MMWR, and they have a very good review of the different populations that

were used to study the IGRAs, and I do believe there's been very careful thought given to what populations do you want to look at to try to determine sensitivity versus specificity.

And so I guess my, without going into detail about those studies and those populations, because we do not have a gold standard for how to know who has latent TB infection, I think those are the populations that future tests would need to look at to try to determine that.

DR. CALIENDO: Angie Caliendo.

I don't want to put you on the spot, Carol, but without a gold standard, how would you see the design of those studies?

DR. HAMILTON: Well, you're really putting me on the spot.

No, but I mean I think the way the studies have been done, so kind of the one thing that you can do is in people, like we talked about, who have culture-proven TB, you'd like to see a test performed pretty well in that group. But we know that there's something funky immunologically in that group, so -- but anyway, that's kind of the closest you have. You know they've got TB in there, so how do they respond there.

And then you look at, you know, for specificity, you look at populations that have had BCG vaccine but maybe have, you know, lived outside the country where you don't think they actually have active disease, they have no nodules on their X-ray, et cetera. So I think those are the populations you do know ideally then -- I mean, the old studies, the TST

studies years ago, really were done looking at patients where they followed them prospectively for years and years, and then they determined, you know, how much reactivity did they have in their TST and then what percentage of those people went on and got disease. Those are not studies that are likely to be done anymore, but --

DR. CALIENDO: Angie Caliendo.

So let me throw something out. You know, we want a specific test. I think of the employees that we screen at our place would not be -- some of them are pretty low risk. What do you think about screening people with what would be perceived as a very low risk to look for specificity?

DR. HAMILTON: No.

DR. CALIENDO: Why not? Just asking for clarification. I'm not trying to be difficult.

DR. HAMILTON: Well, I do also want to kind of correct -- a minute ago you said -- so when I think in terms of what happens in public health clinics, I do know that in public health clinics, for example, people are referred there for employment screening, teachers, et cetera. That is not what we want public health to be doing. Those are -- that's really a waste of time.

So I think -- I mean, the point that the public health, TB public health community has made for 10 plus years now is targeted testing. And so, you know, you only should be testing people who have an epidemiologic

risk or who have symptoms. And so I really -- now, in a healthcare situation, you know, you're in a situation where those are people who are at risk or they are about to be at risk and so you want to know what their baseline is. And so I think that's -- you know, that's a reasonable population if you really think that they are going to be at risk over time. But I think these studies, to really understand the test characteristics, should be done in, you know, people with fairly defined risk.

DR. HENRICKSON: So you say you aren't even --

DR. HAMILTON: For the specificity?

DR. HENRICKSON: Yeah.

DR. HAMILTON: Not really. I mean, I think for specificity, the big issue for specificity is people who have either been exposed to MAC or people who have been exposed to BCG. I mean, that's where you'd like to know. They don't have TB, but they've been exposed to these two things that might make your test reactive.

Fred, am I -- are you -- do you agree with this? Yeah. So those are big populations that I'd be interested in.

DR. RAND: Could I suggest -- Ken Rand.

One population you might be able to get a handle on would be transplant patients in a high-risk area, if you could test some pre-transplant. And what you could do, you could use a variety of tests.

You could look at the patients who tested negative on your skin

test or, in this case, on your IGRA to see what percentage of them reactivated and got TB because they presumably wouldn't be treated since the test was negative, so they wouldn't be prophylaxed. It might be a population where you could actually get some follow-up data.

DR. DELLA-LATTA: What I need to go back to is our original, I think, objective. We're trying to decide how we would want to reclassify a test that's already FDA approved and for other tests and other assays coming out. And in so doing, maybe what we would want to do is just to summarize the characteristics of the current tests that we find are not up to speed and that are a problem and raise the standards for future tests to make it more specific.

I don't know if we're doing with the sensitivity, it might be too sensitive or it may be more sensitive or just as sensitive as skin testing is. You'd want sensitivity to be higher. Maybe you might have an assay that some company might be coming out for urine, for example. It might be another type of specimen in that invasive -- so in just thinking ahead, how, what parameters, what performance characteristics, what patient population data do we want in the next assay, in the next -- insert for the next assay. I'm thinking that maybe that's why we're here.

DR. CALIENDO: Angie Caliendo.

So we've done a pretty good job of patient populations. We've talked about immunocompromised patients, whether they'd be HIV patients

or transplant. Talked about healthcare screening. We've talked about children. We've talked about using it as a diagnostic test versus a screening test. So I think we've given the FDA some good feel for populations.

What we haven't done is giving them much of a feel for the performance characteristics, so getting back to what Phyllis just said, how specific is good enough, how sensitive do we want to make this, knowing that there is some concern about false positives and the implications of those. So let's talk a little bit about that and give the FDA some feedback on that point.

DR. GINSBERG: So Angie -- Ann Ginsberg.

Given that there are IGRAs out there that are approved -- maybe this is a question for the Agency -- it seems to me the standard has to be substantially equivalent even though we might like better. I don't think, and correct me if I'm wrong, that the Agency could disapprove something that was as good as things they've already approved just because, you know, clinically we'd like something better.

DR. HOJVAT: Not unless something changed in the practice of medicine. There has to be some justification for us to do that, and it can be over time. But I don't think there's enough information on these to sort of say, you know, there's another level there. Is that something I'm hearing from the Panel? So we would not, obviously, let something through that was less than this.

DR. GINSBERG: So, you know, I would propose, you know,

simple-mindedly, that the standards be at least as good as the already approved IGRAs and ideally better, but --

DR. WILSON: Mike Wilson again.

I think that, in follow-up to that, as some of the data come out of CDC studies, we'll have a much better understanding of what the performance characteristics are.

The one thing that you have to avoid with these assays is lumping. I mean, they can be used for a number of different purposes, and I think that one thing is that we may want to have, in terms of special controls is, in terms of intended use, they really can have two broad roles. One is to screen for latent TB infection; the other is to look for more active disease, and they're fundamentally different uses of the same device.

So there might be something to specify clearly in future studies, but also performance characteristics and how are you intending to use this assay because you just can't lump those two things together.

DR. CARROLL: Karen Carroll.

I also think that the performance characteristics of these tests are still evolving. So for me, I think it's very difficult to say what a new -- so what the current performance characteristics of these tests are. I mean, I have talked to colleagues who have attempted to use these tests and find them very difficult to perform and not necessarily as objective as we think they are, depending upon the assay that's used. So this is, I have to say, not

published, anecdotal, but in sort of trying to inform a decision for bringing them into our own institution.

So I'm struggling with the fact that I think the data is inconsistent and we don't know what their performance characteristics really are. And I would like to echo, knowing that there is some postmarketing data coming out that is likely to address at least some of the questions on page 13 of the CDC document, I almost think that we have to wait until these postmarketing studies are available to have a better idea of where to go. I mean, that's what I'm struggling with. I'm not sure I can actually help you with that, personally.

DR. TOWNS: Michael Towns.

Looking at the tests that are currently approved, they actually have, when you look at specificity, which I emphasized as probably being more important with a hybrid test that has both diagnostic purposes and screening purposes. They have point estimates on the specificity between 97 and 99 percent with a lower 95 percent confidence interval at -- around 95 percent.

So I think that is acceptable level of performance, at least, I think, from the industry viewpoint, and so I think that's a favorable thing that future assays should be, you know, judged against if they sort of get around meeting that mark for specificity, which I think we're all concerned about.

DR. HAMMERSCHLAG: Hammerschlag.

The point is what are those for? I mean, is that performance parameters for diagnosing active TB, is it a performance parameter as a screening test to look for latent TB? And then remember, it's an indirect test, actually. So I mean, I think that's just a little too optimistic for what we're really seeing in practice.

DR. CALIENDO: Ann Ginsberg. Sorry.

DR. GINSBERG: Sorry. Just to address Karen's point, which I think is a fair one, but it seems to me one approach to that is actually to call for head-to-head tests, right, so you'd have to show the new test in the same trial head-to-head performs as well as the old test. Today we don't know what those -- how to put a number to --

DR. HAMMERSCHLAG: But it would have to be also for the -- yeah. That would be very important --

DR. GINSBERG: For each intended use.

DR. HAMMERSCHLAG: Because they're different animals, different animals.

DR. REIMER: Larry Reimer.

And sort of so what? So you do a head-to-head comparison, which one is right, which one is wrong. How do you interpret either one of them to know whether you've got an accurate result or not?

DR. HAMMERSCHLAG: You should do three and two out of three?

DR. GINSBERG: But I think the point is we've got approved tests on the market, so the question is do the new tests perform equivalently or better? Now, maybe it's hard to define better. That's going to depend on the intended use. I think those are addressable questions. There are clinical trials going on with designs that make sense, so I don't think that's impossible to do.

DR. CALIENDO: Angie Caliendo.

So my issue is how will I'm thinking about these tests change if it's reclassified as a Class II. So we'll have this large postmarketing study, we still have these tests if the performance is what it is. We don't have a gold standard. That's unlikely to change.

And so what's -- I'm trying to decide what's -- why not declassify or declassify [sic]? Sorry. Why not lower the classification? What data, what could possibly happen, like I go back to what Sally said, until the way we practice changes, is there really that much harm in classifying these down a level?

I'm trying to think, in my mind, what would the trigger -- if this isn't what we're going to do now, what do I need to do it later? And that's what I'm struggling with.

DR. GINSBERG: So I think -- again, Ann Ginsberg.

You know, I think as long as we have the appropriate special controls in there so that there are standards and amount of data being

learned in the process, I don't see any risk, either. I think if you reclassify them as II's and didn't put in any special controls, then the risk is you could have worse --

DR. CALIENDO: So let me put you on the spot and say what special controls?

DR. GINSBERG: I feel like just what Karen just said. I don't know.

DR. CALIENDO: Well, the specific patient populations, I think, is very helpful.

DR. GINSBERG: Yeah.

DR. CALIENDO: What Carol said.

DR. GINSBERG: So, again, I think it comes down to the manufacturer carefully defining the intended uses and the populations in which they'd want to use them and then showing that in those populations, the tests are performing at least as well as the currently available tests.

DR. WILSON: Mike Wilson.

One thing that's been suggested, and we're actually doing, is if you have a patient who is in a low-risk category and they get a positive IGRA, that you repeat the test because amazingly, more often than not, all of the tests will actually be negative. There are a fair number of false positive test results with these assays. It's not very high, but it's not insignificant.

DR. REIMER: Larry Reimer.

But which one of those is the right answer, then? So you have one that's positive and one is negative.

DR. WILSON: It turns out that if you follow the patients through time, as we've done, that a small number of patients will have a positive, but if you test them serially through time, they'll be repeatedly negative. And generally, if you wait a few weeks and you retest and it's negative, then it's a negative.

DR. HENRICKSON: Kelly Henrickson.

Those are all surrounding the cutoff, the indeterminate cutoff zones.

DR. WILSON: And some are actually pretty strongly positives.

DR. HENRICKSON: Yeah, but studies that have looked at this statistically have found a statistical correlation around the indeterminate range, the cutoff range. And so that can be handled by having indeterminates.

DR. TOWNS: Michael Towns.

One of the comments was made about for future assays, that we would want to have performance at least equivalent to what is currently marketed or better. But one thing that we have to remember is that with a reclassification down to a Class II with special controls, you're looking at a performance standard that is substantially equivalent. That would be the FDA decision, substantially equivalent.

If you're saying that new assays should have better performance than what is currently marketed, that's a PMA, so that bumps it back. At least, that's how I understand it, that that would bump it back potentially into a Class III if you were trying to get a claim for even better performance than what currently exists.

Would you like to comment, Dr. Hojvat?

DR. HOJVAT: That's always the issue with the 510(k)s. You know, we've had that with serological assays and then the NAAT assays came in. Obviously, you get discordant specimens; which one is the right one. And you get that if you were trying to raise the bar on and improve something at the 510(k) level, you're right.

Now, we can do that by having -- we can't actually use discordant resolution, but we can determine that the new assay result is better than the other one, and we've been able to do that with the NAAT assays by, you know, other tests going on, sequencing, things like that, to prove that actually the NAAT assay result is correct.

So we can do that. It's not the best way of doing it because discordant resolution is not allowed, in a way. You know, you have to go with the performance against the predicate device. And so that's always been an issue, but we've been able to come up with ways of resolving that.

But you're absolutely right, it causes an issue to try and show equivalence when you're trying to improve an assay because those two just

don't go together.

DR. CALIENDO: Phyllis. And then we're going to go around the table.

DR. DELLA-LATTA: One point I did want to bring up and that's in terms of reclassifying it. The positive that I see is that it will be easier for companies to come out with other assays than going with the PMA, that would be easier. And if they want to be competitive and they want it better, hopefully their performance characteristics will be higher. It would be easier to use, et cetera, et cetera.

So I think if we concentrate on the special conditions and the special controls and do them carefully would be fine, but I was also thinking that if we were to, sort of, target certain populations -- and pediatric is one that's always, I mean, in my heart as well as in my head -- that there are no good diagnostic assays out there.

So that's one place where perhaps we could rethink what the performance characteristics might be with that population and something that the pediatricians could come up with, that they would feel comfortable with because it seems as though the data is all lumped together, and as it was lumped together, it's very hard to tease apart.

But in trying to get the data that we need and to improve a test in an area where the diagnosis is lacking, like in pediatric TB, I think that would be a very helpful thing to have.

DR. CALIENDO: Angie Caliendo.

So for special controls, I think Ann said it best, we'd need well-defined intended uses, well-defined populations, and showing equivalence to the currently approved assays. The interesting concept of repeating low risks that are positive, I think, we'll know more once these postmarketing studies are done and we have that data and when the FDA has that data in front of them.

Any other special controls you want to emphasize? And then we'll go around and get people's thoughts.

DR. RAND: Angie, can I comment?

Does the, well, for want of a better word -- Ken Rand -- enthusiasm for the postmarketing studies that are coming suggest the need for postmarketing studies on the new IGRAs? That might help.

DR. CALIENDO: I think that's a very fair point. That would be a good special control since it's not likely that one study would cover all the populations and intended uses that we've talked about, so -- all right, so we're going to switch it up and we're going start with Carol, just your thoughts on lowering the classification.

DR. HAMILTON: Well, I'm in favor of lowering the classification in doing -- and I agree with the concept of any new tests would need to be substantively equivalent, as we've talked about.

I have some concerns about, although I'd love to get the

postmarketing data, I don't want to put so many controls here that we now, again, squash the enthusiasm for companies coming up. We need a better test than the ones we have, and so I'd love more data, but I'm not sure I would say that that would need to be a requirement.

DR. HENRICKSON: Kelly Henrickson.

I am in favor of lowering the classification.

DR. DELLA-LATTA: Favor of reclassification. Phyllis Della-Latta.

DR. REIMER: Larry Reimer.

I'm still on the other side of this. I'm not in favor yet. Again, I don't think we have standards that we can use to say that we really know all the special controls that need to be in place at this point, and maybe after the studies come out, which I don't hold great enthusiasm, are going to answer all the questions for us on top of what we already have. I would choose not to.

DR. HAMMERSCHLAG: Hammerschlag.

I would go along with reclassification, but I think that we need to understand that these tests may actually just be inherently limited because of the immunology, just the natural part of the disease process. Especially when it comes to children, this has to be probably better defined, and it can't be a one size fits all.

We have to do studies where the -- it has to be evaluated, the performance for specific indications, whether it be diagnosis of latent TB,

screening, or actually being used as an adjunct for diagnosis of active TB. And, of course, currently, I don't think the assays are really not recommended for children under five, anyway, and that may be, again, because of the inherent disease, the response to the disease in children.

DR. GORDIN: This is Fred Gordin.

I would not favor changing. I would agree with Larry. I think that there's -- you know, as we've talked about, there is nothing really to test these against, so I think with that difficulty we need to make especially sure, however the Agency does that, that we're getting "to the clinical truth," as we heard this morning, and not over-treating people, causing harm for them, their family, and the public health system. So I'm not trying to quash development of new tools, but I think this needs to be held to a fairly high standard.

DR. BEAVIS: Kathleen Beavis.

No, I support reclassification but only because I don't think that these devices fall into the category of being life-sustaining or life-supporting. And the devices are supposed to be that, as well as have a substantial importance in preventing impairment of human health, and that's where I was going back and forth and debating. If it weren't for that, I would not support reclassification, but it's hard for me to support reclassification based on the second part of the definition of a Class III device, for that reason.

DR. RAND: Ken Rand.

I would favor reclassification as a Class II because I think, if anything, we need to stimulate some basic research in this area and get a better test, and lowering the -- well, favoring the development of such innovative new testing would be the -- could be the outcome of lowering the classification.

DR. CALIENDO: I find this more difficult because I don't feel like there's a lot of good data. I think what Maggie says may be true, that these are inherently flawed tests. I would favor reclassification, and I'm doing it more on the basis of what Kathleen said, which is the risk. I just don't think it's that high of a risk to keep it as a PMA.

DR. WILSON: Mike Wilson.

We're trying to reduce what is actually a hideously complex subject into a limited discussion here. I think we're doing okay.

I would favor reclassification for the reasons that Dr. Beavis said, but I also think that TB is an enormous public health issue. I think it's naive to think we're going to eradicate TB anytime soon. And it's an enormous issue globally, and I think anything we can to do facilitate development of better assays is important.

DR. JANOSKY: Janine Janosky.

Favor reclassification.

DR. GINSBERG: Ann Ginsberg.

I also support reclassification largely because I don't think

keeping it at the level of needing PMAs mitigates much risk relative to a Class II with special controls.

DR. HARRELL: Lizzie Harrell.

I agree with Ann. I vote to reclassify.

DR. CARROLL: Karen Carroll.

For some of the other reasons that were mentioned, I do not favor reclassification.

MS. DeLUCA: JoEllen DeLuca.

I would favor reclassification. Hopefully, we'll get a broader open market system going.

DR. TOWNS: Michael Towns.

Speaking on behalf of Industry, I'm in favor of reclassification because I don't see these devices as being that high a risk to remain as a Class III requiring a PMA.

But I also don't believe that with the reclassification down into a Class II with special controls, that we should have the onus of burdensome postmarketing studies that should be conducted as part of that effort, so -- but I am in favor. I don't see these devices having any higher risk than what we currently have with regard to TST testing. So for those reasons, I am in favor of reclassification.

DR. DUERHRING: Gary Duerhring, Consumer Rep.

My concern is we do have these studies going on right now

concurrently with this discussion with the CDC. I would really like to hear the outcomes of those studies. It might resolve some of the questions we have as far as to guidelines and intended uses and assist the structuring that the FDA would have to put out in their guidelines and recommendations. But knowing that it's really not a strong risk area, I would support reclassification.

DR. CALIENDO: Steve, do you have any other questions?

(No response.)

DR. CALIENDO: So I'd like to take a minute to thank the Panel. It's been a long day. We've burned some serious brain cells. We really appreciate your participation. You've hung in there, been incredibly educational for me. So thanks to all of you. I very much appreciate all of your effort.

DR. HOJVAT: And I would just like to add to that, too, from the FDA. This has been very good for us to get outside opinions. We can sit in our offices and try and go through the what-ifs, but hearing the actual use of the assays and how you use them with patients is very, very helpful.

So the next steps from here are really for us to go back and consider all your advice and recommendations and then decide if, you know, probably if we do or do not go through with a reclassification petition for each of these three things.

DR. CALIENDO: Sally, just a question. Is the committee notified of your decision?

DR. HOJVAT: Yes. I mean, I think you saw, we may have to come back and if we decide that it's not a straightforward thing, we may have to come back and convene the Panel again. I'm not sure we would need to do that, but, you know, that's -- I think we probably sucked you dry.

(Laughter.)

DR. HOJVAT: But, you know, we have to go back and consider it. It might be homework assignments rather than the whole Panel. So we have to go back and just go through all of this information.

DR. CALIENDO: Ultimately, how is the Panel notified of the decision, I guess is what I'm asking.

(Laughter.)

DR. HOJVAT: We will.

DR. CALIENDO: Yeah.

DR. HOJVAT: We will.

DR. CALIENDO: Okay.

DR. HOJVAT: Obviously, we would.

DR. CALIENDO: Okay.

DR. HOJVAT: And it has to be a sort of closure for all of your efforts, too.

DR. CALIENDO: That's right. That's what I'm thinking.

Go ahead, Phyllis.

DR. DELLA-LATTA: Okay. Just one informational point I'd like

clarified. What issues do you actually vote on and what issues do you ask for just opinion? Just for our opinions?

DR. HOJVAT: You mean if we had another Panel or --

DR. DELLA-LATTA: Yeah. Well, what issues do you actually ask for a vote and what issues --

DR. HOJVAT: It would be -- it should be reclassified. I mean, you sort of did a trial around the first round, although we weren't supposed to do that, but in essence, we've heard, sort of, your opinions.

DR. DELLA-LATTA: Oh, okay. Opinions, okay, fine.

DR. HOJVAT: A list of your opinions.

DR. DELLA-LATTA: Thank you.

DR. RAND: So would there be an official vote at a later time?

DR. HOJVAT: We don't need to if we feel we've got enough information. I mean, that's one of the -- what I understood from what we heard this morning. That depends on where we want to go with this. We don' t have to go back to a panel to reclassify.

We didn't do that for down-classification of HAV, for example. But if we felt there were still issues that needed to be discussed, then we would call the Panel together again. I think that's how this works.

DR. CALIENDO: Okay. Thank you, also, to the audience for hanging in there and giving us moral support.

Remind the Panel members, leave all of your booklets here.

You can take the handouts from people's presentations, but the booklet that the FDA gave us stays here. Thank you very much. Have a safe trip home.

(Whereupon, at 5:00 p.m., the meeting was adjourned.)

C E R T I F I C A T E

This is to certify that the attached proceedings in the matter of:

MICROBIOLOGY MEDICAL DEVICES PANEL

June 29, 2011

Gaithersburg, Maryland

were held as herein appears, and that this is the original transcription thereof for the files of the Food and Drug Administration, Center for Devices and Radiological Health, Medical Devices Advisory Committee.

_____

CATHY BELKA

Official Reporter