Table of Contents

P R O C E E D I N G S

DR. BURKE: Can we get started? I just want first thank everybody for being here. We are going to have, I hope, a pretty intense discussion and your participation is tremendously valued. What I think would be most useful is for us to just start very quickly going around the table and each of us saying who we are and where we come from.

(Introductions)

Review of Working Group Goals

DR. BURKE: Again, thank you all very much for being here. I want to just make a couple of remarks about how we would like the day to go and where we would like to be at the end of the day. Most importantly, what we are heading for is hopefully the ability to report tomorrow morning to SACGT a preferred approach to test classification. I anticipate that in the course of this conversation we may also, in addition to some idea of a scheme, be identifying points that are issues that need to be discussed by SACGT.

The intention of the structure we have put for the meeting today is really just to make sure that we address certain areas of discussion that we know are important to discuss, that we know have come up in SACGT meetings previously, but this is intended to be an open committee discussion during the course of the day. We are not really going to have formal presentations. We are going to discuss topics.

So what we have laid out are the topics that we think are essential to get where we need to get to, starting with what kind of

criteria ought to be considered in test classification, looking then at some strategies that have already been proposed for test classification, making sure that we set aside a chunk of time to talk about social issues and how those fit into a test classification scheme. Then allowing time for procedural and other issues.

Criteria to be Considered in FDA Review of New Genetic Tests

Before I open the floor to get started, I do want to remind you that we have laid out some assumptions, a little bit more detailed discussion of goals in a document called "Goals and Framework of the Working Group Meeting." That is in Tab 3 of your book. In particular I think it is probably the place that we should start. I would like to lay out the topic of what happens after a test is reviewed. That is, what are we looking for as a result of the review process that we are going to be talking about. I should say that we are not going to be talking about the review process, we are going to be talking about test classification. But I think as we talk about test classification we have to think about what happens once a test is classified as either "routine" or some form of high scrutiny, acknowledging that we may have a binary classification, either routine or high scrutiny or we may have different categories of high scrutiny. But as we think about that there are potential different outcomes of being a high scrutiny test or of undergoing review.

The simple one is a test is reviewed for whether or not it is commercially available. So that is the simple, straightforward issue and I think if our test classification is simply to classify tests as needing more pro forma or much more detailed review before

they become commercially available, that has certain implications for how we classify tests. We have listed in this document, in Tab 3, three other issues that might also be outcomes of review. One is labeling requirements, that is, that there might be certain tests that have requirements for labeling that hopefully would make clear to consumers and health care providers some of the caveats or uncertainties about the tests. Another would be counseling requirements, the concept that perhaps certain tests should be released only with the expectation that they are provided in the setting of counseling requirements. If that were an outcome of review, clearly there would have to be some attention to how you document that. Then probably most important, because it has come up a lot in our discussion, the possibility that it may be appropriate for some genetic tests to be released for commercial use at a fairly early stage in their development when inherently there are uncertainties about some of the properties of the test, yet some commercial value, and where outcome of review and outcome of commercial development should insure ongoing data collection.

The concept that I am putting before you at this point is to what extent should these, or potentially other outcomes of review, be considered now because they will determine our approach to test classification. So I am going to open the floor. I think we need to have some discussion about that, but I certainly want people to bring up other immediate issues that we should consider as well.

DR. VOELKERDING: Could I just ask for a point of clarification? When you use the term released for commercial use, is this inclusive of both home brew tests and kit manufacturer’s kits?

DR. BURKE: I think at this point certainly you are bringing up issues that we need to discuss, but we have, the SACGT has determined that all genetic tests should have a pre-market review. So when we are talking about test classification at this point, and please anybody correct me if their understanding of our previous discussions is different, I think we are talking about all tests. Classification might include whether they are home brew or some other form of test.

DR. WATSON: I think it will, because a kit will have been FDA approved, reviewed, and cleared.

DR. WINN-DEEN: I think there is also some black clarity in this labeling requirement because typically when you talk about labeling requirement, you are talking about a kit in a box that has labeling requirements from the FDA. You are now using that term in terms of really what the report form says on it. Is that what you mean by labeling requirements?

DR. BURKE: Your point is well taken that we want to be sure we don't use language in different ways. So yes, let's try and come up with some other kind of language. It could be a test report. Probably that is what it would be. The issue had come up. I know it has come up in discussion before that some tests should be released only with clear information about their inherent limitations. So I think you have just operationalized it. That is what we would mean by that, that when a test result is reported there is some additional information that comes with the test report.

DR. CHARACHE: I think as we think of labeling requirements I would like to expand it. It is not just the report. It is the kind of information that is in the package insert of a kit, so it includes

information of when you should order the kit and what the evidence is that it measures what you think it does.

DR. BURKE: Thanks. Michele?

DR. CAGGANA: A lot of the laboratories that I deal with have a pre-test information sheet that they make available and I think a lot of times they do cover that in there as well, including the limitations.

DR. BURKE: So that would be an example of what Pat was referring to.

DR. NOLL: I think this is a very important area and I think it is something that the community would embrace without any problems. There are other things that we are going to talk about today that will be controversial, but this is something that we can all agree on is extremely valuable and I think it gets at most of the things we are concerned about. It is really quality information, accurate information. I would simply like to point out that in the existing quality assurance program that ACMG and CAP does run, in those checklists and in the commentary of the companies, and what are in fact the requirements that a laboratories must have in order to be accredited, there are specific requirements that for every test the indications for that test be spelled out. It is a requirement, so-called Phase II. Phase II in ACMG and CAP is bad, a reversal of hospital organization.

Secondly, there are extensive questions and then commentaries to explain the questions that have to do with reports. And the reports not only say that you have to say what you have found, but those reports have to provide sufficient information, and the

language goes along the lines that a general practitioner or in some cases a patient, can understand what the report means. And it means giving information about the test results seen in the context of certain ethnic groups or certain indications for testing. It talks about low frequencies. All of that stuff is supposed to be on a report. So I just wanted to remind you or bring to your attention that there already is a framework that recognizes these things that is out there and is working. It may be working imperfectly, but I will bring this up again because I think it is something that we might look to, at least initially, and build on.

DR. BURKE: And perhaps given the context of our discussion, I think there are two things we might need to discuss. One is, are there different classes of tests that require a different, qualitatively different kind of educational materials associated with them. Is that important part of test classification. And in the context of pre-market review, that is, the concept that there will be pre-market review, are we really talking perhaps about a requirement for manufacturers to provide what kind of educational materials they would make available as part of the test.

DR. NOLL: If I could respond to that? The approach we have taken so far is that all tests must have this information across the board. The issue of the classification that I think we are going to be talking about, it seems to me where that really is relevant was when you talk about things like is informed consent necessary for all tests? And is it important or essential or desirable that a person have formal genetic counseling depending on the categories? But in terms of strictly the laboratory’s perspective of doing quality tests, doing it

on the right people and then providing full information about what those test results mean – that applies to all tests. And it may be this big for an uncomplicated test and it may be that big for a complicated test. But I think in the future we don't just need to look at our reports for that information. The report could very well be an abbreviated summary of the interpretation and have a web site that gives full information, something like GeneClinics or something like that.

DR. BURKE: Other comments?

DR. WATSON: There are two places where I think the major problems hit. A lot of it is educational, clearly, which is a difficult thing to regulate. I think there is the stage, and it is all defined differently – I think one of the things that makes this difficult is, for instance, in New York State there is a clear category of investigational types of testing and then there is accepted standard of care kinds of testing. I think a lot of the issues we are talking about, certainly those that have driven people to say we need to regulate genetics more, is that translational step. My I look at it, even though I don't agree with everything that happened with the move of BRCA 1 testing, for instance, I think it probably met most of the requirements. It was done under IRB approval. I don't see which steps were missing there. They may not have been done well was the problem. That is a very difficult issue to get a handle on, that translational step where it is defined as not knowing what the tests do. It is where you learn what the test does. And that is exceedingly difficult to control. It is going to come out one way or the other and what you want to do is be able to monitor it to see where you end up and measure

it by some standard or not. That is a very difficult stage to deal with.

At the other end are those tests where that information may that a very long time to accrue and those, I think, are places where you look for public agencies to get involved in data collection to help that information come together more quickly and more broadly in order to be able to get that information collected rapidly.

MS. BOLDT: I guess my concern is the off label use of different genetic tests and what is currently in place in terms of off label use other medications at this point and is that enforced? If we talk about this, does that matter if people are going to be using off label anyway?

MR. HILLBACK: I guess I would like to ask how we organize at the beginning. There are a lot of people here who have been part of a lot of discussions on this topic through the SACGT. Others have not. The fundamental question is, what are we trying to accomplish by first of all increasing oversight by government, and then, if we are clear on that, we can start asking the question how do you classify tests into those groups that need high oversight? I think, unless we are very clear up front, we will end up wandering over this map quite a bit without really getting to the point and we get in this complexity of just going off on one topic, whether it is soft label use -- not that that is not a good topic. I wonder if there is a way to get back to what is the problem that we are trying to solve and make sure that everyone in the room has a chance to get involved in that discussion first before we go off and talk about the next steps.

DR. BURKE: Yes, if I could just make a comment before Pat,

I think that you are exactly right, Elliot, and I think that is why we need to discuss what happens as a result of review, because to make it very specific to what you just said, if all we are talking about is making sure that some tests do not come to commercial use prematurely, because we think that is the problem, then that is one kind of discussion. Whereas is what we are really concerned about is that certain tests are going to and perhaps should come to commercial use at a time when there are still uncertainties about them, which I think is your point about the BRCA 1, 2, then an outcome of review may be to make that apparent or increase our efforts to make that apparent in one way or another. Another offshoot of that may be to unsure that the data we ultimately want is collected.

MR. HILLBACK: May I just follow on that? I think one of the points we have made a number of times, and it is a phrase that I have worn out its welcome maybe, is tell them what we know and what we don't know. One of the outcomes of scrutiny may be to make sure that whoever is performing the test – I don't like to use manufacturer because I think that leads people back to big companies that manufacture versus lots of labs that do the test – but it may be that what we want in terms of scrutiny is that there is a checklist that has to be met that each laboratory that is going to perform the test can prove with more depth than is now required by CLIA and now required by the other processes, that they have their thing together. That they have their act together and they have done all that. There are other options, but I think that is what we need to debate. What the outcome is.

DR. CHARACHE: I think along with what Elliot is saying, in

order to clarify our nomenclature, two things I would like to comment on. The first one is, when a test is investigational, when it is patient care, it is a commercial test regardless of who has manufactured it. Whether it is a company that has made a kit or a small lab that has a PCR for a single analyte, but the definition of an investigational test is that the information remains within the laboratory. A definition of a patient care test, and this is according to CLIA, the federal definition of CLIA, is that the information is provided to a patient or a patient's family or a health care provider. Now before then you can use a test to provide information. You have to have a certain amount of information. This group can then say for this test you don't need lot of information. You simply need to know what you know and what you don't know. But it becomes then a patient care test on which you are still acquiring data. As a patient care test is has specific requirements. So I think we have to be clear and sharp in what we mean when we say it is investigational. It may be both investigational and patient care, or it may be strictly investigations, where information is being collected by a laboratory but is not being dispensed for patient care use.

I think the second comment has to do with when a test needs increased scrutiny. Here I would like to suggest we have two reasons that are quite separate and need different people to make decisions on them as to when tests need increased scrutiny. The first is that, which is now in place, when a test is complex to perform, such as a PCR or a new investigational thing, then this is considered in need of higher scrutiny by the FDA at the present time, simply as a laboratory test. That decision is probably best made by the people who do it now,

the laboratory people who understand the complexity of the procedure itself.

The second reason for increased scrutiny has to do with whether you need informed consent, whether this is a test of predicted value but no therapeutic value, et cetera. And this is a separate type of scrutiny. So I think we have to keep those two ideas totally separate because those decisions would be made by a different group of experts.

DR. LEWIS: I just want to comment on the issue of informed consent. I think we get a little bit confused between the difference of having people having to sign a piece of paper, and having people have information that they need to make a decision. I would argue that every test we do and every procedure we do requires informed consent on the part of the person who is undergoing the procedure. I would like to see us look at that at two different levels. One is when we talk about things needing informed consent, I worry about the fact that we are going to say that some things don't need informed consent, when I think that everything needs informed consent on the part of the patient. Therefore, we need to separate out the difference between some kind of written form, which to me signing a piece of paper is not informed consent. So I would like to see us pay attention to that a little bit and not presume that there are some things that happen that don't need consent.

DR. BURKE: Could I just ask, for clarification on that point, if we acknowledge that a written informed consent can be very pro forma and not very useful, it seems to me there is still a distinction between tests that we might be comfortable coming under a

kind of global informed consent, that is, the kind of consent that patients give for routine medical care, versus the kind of test that might require detailed personal discussion about that test implications.

DR. LEWIS: One of the things I worry about is that we do thing as routine standard of care, we do a bunch on routine screening tests, and somebody who doesn't know that they have had a triple screening all of a sudden gets a phone call that says your screening test was abnormal and they didn't even know they had undergone a screening test. So I would like to see us move more in the direction of making sure that people know what is happening to the specimens they are giving. Sometimes that global informed consent gets a little bit too global when it gets to the level of the individual. That is my concern, but I agree with you.

MR. HILLBACK: I would like to follow on Pat's point because I think it was a good point for us at least to divide the world a little bit. I think what you were saying, and I will try to rephrase it, is that we are back to separating are we going to increase levels of scrutiny based on the complexity of the analytic part of the process, versus the clinical part of the process. It seems to me we could get two sets of increased scrutiny, one based on how hard it is to actually do the tests accurately to have high analytical validity, and then how complex is the information that comes out of the test from a clinical point of view.

It might be useful to separate those completely rather than try to get into 11 or 12 dimensions, which some of us have admitted flunking college courses beyond three dimensions. I think it is a

useful separation to try and separate those two because I think a lot of the issue that most of the group has wanted to talk about is more the outcome complexity, the knowledge complexity. How do we manage this complex knowledge we are getting from a test and how do we relate it back to the patient in a useful way and is there a useful way and is there something they can do with it? Rather than focus in on do you have to have a mass spec versus a more capable laboratory.

DR. WATSON: I disagree a little bit with Pat's interpretation of the investigational aspect of testing. I actually like the way New York State approached it. I only presented the two pieces which are the service side. There is actually a third piece which is called research. My sense is that CLIA distinguishes getting results from not getting results as the dividing line. Research in New York is where you're collecting population-based information that establishes a relationship of the gene to the disease and says this is scientifically valid. I think then you move into having to talk to people and get their specimens. Frankly at the front end, that is where the best informed consent is now, is getting specimens and telling people there will be no results provided on that.

At the next stage, once you establish that relationship, and you move into the clinical investigation of what is the magnitude of the relationship, what are the performance characteristics of this test, that is where you are acquiring knowledge and that is where you are providing a service of all that I know and don't know attached to it. At some stage, usually at the payer level, somebody says okay, compared to everything else I can do to answer that kind of question for this patient, is this test better than that test and is it worth

doing. That is when it moves to the standard of care level where the payers all acknowledge it and start covering it for specific intended uses. So it is that investigational service stage that you don't want to get overly encumbered by things that you might link to utility at the accepted for the standard of care level.

DR. KHOURY: Actually, as of this morning, I must say that I am a bit more optimistic than what I was last night reading only the documents. I know this is a very complex area that we have been trying to solve for so many years now. SACGT has made such a big push in that direction, but I think we have the elements of a solution emerging. I don't know why I am saying this early on, but what is clear in my mind – I have a few points I would like to emphasize. The difference between the analytic side and the complexity of the lab test, per se, and then the clinical usefulness, the clinical utility, clinical validity, and these live in different worlds right now. Maybe we are trying to force them too much to go together in one paradigm. That is one comment I would like to make.

The other comments is with relation to telling people what we know and what we don't know. Right now we have tales of this and I really must say with a straight face that if I am a consumer that tries to get genetic tests by one lab versus another, the information I am getting is not standardized. People try to sneak in information that may not be interpretable. Granted that data will always have to be collected. Even for a perfect test there is always new information that is coming. But what we know at this given point in time may be given to me in a very different fashion if I went to place A versus place B and with more or less emphasis on the value of that information

or lack of value.

So even with informed consent, even with all the good stuff, there is no standardization of the actual framework of what we tell people, what are the parameters of the test. That information, granted that it is incomplete, needs to be standardized. We need to create a framework or a template by which certain information is reported and then go collect the data. And the data will have to be collected primarily for genetic tests that are given to healthy people for predictive purposes where there is incomplete penetrance. This is really sort of the high end of that scrutiny. I mean, telling people they have a APO-E4 allele and predicting Alzheimer's disease three years down the road or BRCA 1 for cancer.

So I would like us to separate those issues and try to work towards a framework that makes sense, that is not too cumbersome, and the reason I was pessimistic, I was looking at all the dimensions of the classifications that the various good people around the table here have put forward. We can get back to that in a few minutes, but it is just too complex and it has to be something that makes sense, that immediately hits you between the eyes, that separates the clinical from the laboratory complexities of the issues. Then make sure that information flows in an ideal way, whether you get the test in Oregon or you get it in Georgia or Maryland. It doesn't really matter.

DR. CHARACHE: Returning to Mike's comments about when a test is investigational versus patient care, there is no conflict, Mike. What I am saying is that the first step, which is research and doing your statistical data and so on, is determining clinical validity. That this test does mean something. You are not going to

offer it to a patient unless you know that it means something.

What you are describing is investigational, whether you are prepared at this point to give information to a family. And you say I don't know exactly what the penetrance is but we can tell you if you have this gene or not and we can tell you something about what it means.

That is at the point in which it is both patient care and investigational. But you don't offer it for patient care unless you know its clinical validity. You don't know all about it. You don't know its clinical utility. That is what is going to be developed over time as you get more experience and more data. I don't think there is a conflict. I think you are describing what a test is used for. The information is provided to a patient or a patient's physician and therefore it is patient care. Now you are investigating it and you are saying I don't know all about this yet, but this is what I know and I can therefore use it.

In terms of the issue of separating the text complexity from the issue of the social and medical implications of the test, I think Joann Boughman has once again captured what I was trying to say in a beautiful diagram, which you will shortly see. We are saying, however, that all tests need to have established clinical validity before you offer them for patient care. That is a given across the board.

DR. NUSSBAUM: I find the distinction between analytical and clinical very helpful, having gone through CLIA approval for testing that I am doing which is both clinical and investigational at the same time. It is not one versus the other. It can be one and the other.

I thought that the issues that CLIA brought up as part of the analytical aspects were extremely valid and needed to be addressed and were addressed appropriately, but did not at all address what I think should be the focus of this group, and that is, once you know you can do the test properly, no matter how complex it is, what does the test mean? In particular, and the problem that Elliott raised is I think the main problem we should be addressing here. That is, how do we make sure we protect the public from widespread use of testing that is not well understood and whose implications are not well understood and therefore causing harm?

DR. BURKE: Elliott, you are next, but if I could just interject a comment before. What I am hearing in the last several comments is that it may be useful to get away from terms like investigational that have been used in different ways, and just clarify at this point that we are talking, I think, about three categories of tests in a very global sense. One of them is tests we don't know enough about to use in patient care. In some sense, they are not ready yet to go through this whole evaluation process.

Then the two other categories are tests we know enough to use in patient care, those we have enough certainty that there is clinical validity, that we can identify clinical situations where they have a place, yet we know there are many questions about those tests and we know that those tests will only be answered over time. So I would see that as how do you then responsibly disseminate those tests and make them available.

Then finally, tests that we know most of what we want to know about them, taking into account Muin's point that you never know

everything. Yet there is a category of tests where we are pretty satisfied that we have the picture and there would certainly be issues of labeling and education, but we kind of know what we need to know.

MR. HILLBACK: That is a great way of looking at it. Unfortunately, those of us who have spent a lot of time here, and you are one of them, and not just with this committee, if I tried to categorize right now what percentage of genetic disease would be in each of those three categories, we would have a lot in the first category where we don't know enough. We would have a huge amount in the middle category and we would have almost nothing in the last category. Sort of an answer to Mike, I am not sure when we ever get out of the investigational stage, the way he defines it, in genetics. We are talking about an awful long time, and this is the struggle we have had for years. We had it with the task force before the Secretary's Advisory Committee. How do we deal with that increasing knowledge we get over the next ten years after we first start using this clinically?

So I think to use those three categories is fine, but I think the problem we will still come back to is when is the transition from you should not use it at all clinically to clinical – that is a crucial transition point. Then how do you manage the increasing knowledge during the time it is in that middle category so that there is some standardization, so that there is openness, transparency - whatever buzz word you want to use – back to tell them what you know and what you don't know, and there is some way to moderate that and make sure it is happening? Because I think that is the category, the transition from A to B, and once it is in B, how do we make sure there

is some standard that everyone who is going to do a BRCA test has to live up to? Or the one person doing a BRCA test has to live to if there is only one lab doing it?

So I think that sort of gels what the issue is we have been trying to fight with for all these years. It has been a long time.

DR. BURKE: Let's have Joann show is her transparency.

DR. BOUGHMAN: For those of you who are not on the SACGT, bear with me. I am a dean, so that means if I can't draw it really simply, I am not going to be able to understand it. When Pat Charache and I were walking into the room this morning, she was talking about this parallel kind of process and so I sat down an started to draw. So bear with the quickness of this. But the way Pat was talking, I saw this as in fact two sets of parallel processes going on. One would be the laboratory test itself side of the issue, not unlike what is done now through current FDA processes. There are internal processes. There are the kinds of question that are asked and all of these things are done in conjunction with the other federal regulations.

In fact, if the test, in a laboratory sense, has a degree of complexity or higher scrutiny, then in fact what we need to develop, of what the FDA needs to take a lead in developing, is a process that could be internal, that could require external review, as in panel review, and as I say, as currently derived, the current process or one that might in fact have certain streamlined elements and/or differentiation in certain aspects on the genetic side.

The more complex issues, the shift that Mike was talking about, and Elliott just a moment ago, would be medical and social issues. You would once again have a set of criteria, yet to be

determined, on whether it would fall into yes or no. If it did, then I call this, for lack of a better determination process coordinated by the FDA, but this is one of the things we have talked about previously in the SACGT, where the CDC or other agencies might be involved, if in fact data collection were elements of this process, but we have determined that FDA ought to be involved.

In fact there might be an internal or an external process here. I am calling internal to be not necessarily inside the FDA, but appropriate multi-agency issues at this point.

DR. KOENIG: Joann, other people's eyes are better than mine, but if you could just read your boxes to me?

DR. BOUGHMAN: Sure. Over here I just say is the test complex according to – and I say the current FDA categories we have talked about, high scrutiny, low scrutiny and we keep coming back – there might be some differences in genetic versus other kinds of tests. But those criteria would be concerned and we have a no and a yes category here. Over here is, does the test have significant medical or social implications and that is, once again, many of the issues that we listed earlier this morning – the issues of predictive versus diagnostic. We have the frequency issues, the clinical utility versus validity, we have frequency, we have whether a treatment is available, we have all those issues that would have to be overlaid.

Once again, if I could not figure out how the process went, how you would go through the process – of course you could have a very simple no-no. Then I jump down to the most complex. But of course I think what Elliott was just saying is that there might be several tests that, from the laboratory side, would in fact go through what is now a

routine, even streamlined process with the FDA, but in fact would immediately be considered high complexity on the medical social side, and possibly need not only the multi-agency but external review.

I didn't say at the beginning that I had the pleasure of chairing the FDA genetics panel. I am a liaison here so I have a vested interest in what we would pose for a new ad hoc panel to look at over here.

As the discussion was going along, labeling issues could be addressed in any and all of these stages - internal FDA standards, advisory panel standards, from either side of the process. Data collection requirements – for example, I am thinking if you are talking about something where the laboratory process is fairly straightforward, but we really don't know about the application to broad populations, that in fact puts it over in this category so that in fact there would need to be the right kind of review and suggestions to make sure that labeling and follow-up data collection continued to be included. I just threw in another one, the limitations on the test, who can perform it, what kinds of qualifications, some of the CLIA complications, accreditation, certification requirements and so on – the laboratory parts of that could be handled on the laboratory side in those steps. The medical interpretation and other complications and issues would have to be handled on the right side of this diagram.

Pat just asked me to try and put down on paper what I heard some people saying. I don't know whether this is useful, but one of the things I think we have to do as we go through this is this. I do multidimensional genetic modeling, but I cannot draw in the six dimensions and I cannot begin to think in the six dimensions, which is

what happens when we start thinking about all the issues and dimensions of the test. But the process has to be a stepwise process.

So my suggestion is that somehow we come up with at least a starter grid. Then we can start saying which place the issues belong and who it is that should be making the determination about those issues, because in fact the initial steps are going to have to be done by a broader group with appropriate input.

DR. WATSON: Well, there is a whole lot of information there from a lot of places. I agree with some, I disagree with some. I think I disagree with Elliott to some extent in his presuming that there are very few things we know an awful lot about. I think I have probably signed out on over 60,000 cases in the last 15 years or so and I think we know an awful lot about the vast majority of the cases I have signed out of my laboratory. I don't think it is an all or none kind of issue.

I think it is specific to the person sitting in front of you as to whether you know an awful lot for their particular situation or not. I think that is the inherent problem here. I will throw one of my major biases on the table now, which is I don't think you can regulate clinical practice. The practice of medicine has exemptions in virtually every regulatory body. It has exemptions under FDA. It has exemptions everywhere, to practice medicine. I don't think you can realistically control that step outside of one place, and I hate to be crass and commercial, but I think our health care system is crass and commercial right now, and I think reimbursement is probably the place where you are actually able to control to some extent what is done and what is not done by controlling what is paid for and what is not paid

for.

DR. BURKE: So the accountants have to understand this.

DR. WATSON: No, I think the SACGT has to understand that reimbursement is a critical piece of control now and that our billing and reimbursements systems are not built to be able to understand why somebody is doing something. Nowhere in ICD coding do you see something that says I am testing this person with these CPT codes, analytically or clinically, for this intended use, predictive or diagnostic. ICD is just loaded with diagnostic, but it does not deal with genetics systematically in a way that you can actually say, okay, they are doing this test for this reason and I am not going to pay them because I don't think it is justified. That is sort of my bias on the clinical side, that I have not seen examples to date of any ability to regulate that side of the practice. I think the analytical side, through the CAP programs, have found a number of ways and these programs have been enhanced over time. And we are adding some new tiers to them already.

Under CLIA, all genetic testing is high complexity, as defined under CLIA. So I would hate to start by saying that virtually everything we do, there is never going to be a no. It is always starting with yes because it is defined as high complexity testing under CLIA.

Now I think from a laboratorian's perspective, we can look at, within our high complexity testing, at what is straightforward, and that is an assay for a target of known, a direct target of known clinical significance where the interpretation is straightforward. I think that is pretty clear. And that is where you may move to

laboratories that don't have the same general background as a laboratory that might be doing genomic scanning, looking for variations in sequence and interpreting those as whether they are pathological or not. That is a much higher level of practice.

Within the CAP programs, we have already reached general agreements where, for laboratories doing what we consider to be those more complex areas of testing, they will be inspected by people board certified and trained in that area of testing because it is unlikely that an inspector with less skill can actually tell if a lab has met the validity standards for an assay of that type.

On the analytical side I think you can build in tiers of enhanced inspection of laboratories. The tests will be straightforward probably. You ability to tell whether the person actually does it well and has validated appropriately may only be done by somebody very much experienced in that kind of testing, at least for transitional period in which you develop the guidance standards by which somebody less skilled can go into the lab and say, okay, here is a measure that I can use to tell if this SSCV test was done in an appropriate way or not, or if sequencing was done in an appropriate way. What we lack are the standards by which people at general levels of medical technology backgrounds can do these things. So we are enhancing those analytical programs to bring more skilled people to the front line of looking at the laboratory and its practices, to add that dimension to the more high complexity areas. I think we can address many of the analytical areas from a laboratory practice perspective. The clinical area is still the hard side and until you realistically look at the pieces of that, that you can actually control something with, you can talk

scrutiny models until you drop. But ultimately they have to fit into the models around which you can actually regulate, oversee, control, or educate. If it is mostly educational, when you have a population with a range of educability that is enormous, and some will educate and some won't, and most will have some level of understanding.

DR. CAGGANA: Along the lines of what Mike was saying, in New York we do try to go through a process by which the test themselves are reviewed. So we have a system where certain tests by certain New York State laboratories are regulated by us. From that point, if a patient needs a test that is not covered by tests we have already looked at, they have to submit a form to us and we review it, and in a lot of cases we end up having conversations with insurance people, with physicians, with Medicaid, in order to get the patient the test that they need, or sort of say do they really need this test. Each case in handled on a case-by-case basis. So that is sort of our outlet to bring new tests or rare tests that are not regulated or there is not a population base from which to get any sort of clinical data.

The other part of the process is almost exactly what is outlined in the four bullets here. We review the test to see whether or not it should be released. Is there a publication history associated with it? Is it being used in the field? What kind of information can they give us to assure that that test have some validity and some use?

Under labeling requirements we review the reports to make sure that the proper information, at least the up-to-date information available is present and is given the report so that a physician or anyone else reading that report would have a good handle on what they

should tell the patient from those results.

We also review the physician information, the advertisement flyers and those sorts of paperwork that is available. And we have found errors and corrected those and we have that mechanism. In New York, what Judy had said, we require all genetic tests get counseling and they do it by a broad range. In some people, some of the academic labs have a very good genetic counseling program that is associated on a patient by patient basis, obviously, whereas others in some of the industry may have checkoffs or have a consent form that is on the back of the requisition forms and should be signed.

The problem, as Mike said, too, is that we cannot regulate the physician. A physician may order a test. The patient may go and request a test, so that always comes up as a difficult one to deal with. Also we ask laboratories to collect data. We don't always get that back and I think that is the major thrust of what needs some improvement. You have done 10,000 tests for this disease – how many people had the disease? How many people did not? What were the bad outcomes, the good outcomes? Those kinds of questions is where we really need to beef up the process because most of our tests do lie in the B level of the three categories. Usually they don't come to us if they are still in a research setting. We have a lot of tests that are clinical validity and then a lot of the old tests, the fragile-X, CF tests, my bias is that we know enough about those that we can use the information for the patient.

DR. COLLINS: I guess I am a little confused about the laboratory arm of Joann's diagram and I just wanted to ask for clarification. I had thought that most of what we would be talking

about today would be the clinical side and how do you establish validity and utility and how important is it to know a certain amount about that with the tests that we are most concerned about by whatever algorithm we settle on. I assumed, and believe me I have sat around tables like this for years now, and I am a little worried that I missed something along the way here, that most of the oversight, as far as the laboratory analytic validity and the issues about is the test complex or not and is it being performed by a laboratory that is going to give an answer that you can trust, is really under the purview of CLIA and therefore the beefing up of CLIAC with its new genetic advisory committee was supposed to be taking care of this issue.

Now is there a real possibility that FDA's oversight here would duplicate or even be discordant with what CLIA would be doing in that circumstance? Because that would trouble me.

DR. BOUGHMAN: I can try to respond. Steve Gutman jump in here to clarify this. Remember the question that we were really posing this morning was marketability? We have backed up a little bit from the place we actually started this morning. We were talking about releasing the tests or what does "okay" mean for that step, which in fact does include not just CLIA but the FDA process, when we are either talking about a kit or reagents or other processes, that analytes that are being used out there by a huge number of people. In fact I think there is no reason, and would certainly hope we are not talking about discordance between FDA and CLIA, but where yes, that decision is made at least on the marketability question. I may be the one who is missing the whole point, Francis.

DR. GUTMAN: I actually would be disheartened to think there

would be discordance because the basic principles of analytical characterization that would apply under the CLIA regs and under the FDA regs I would hope would either be identical or near identical. There are immense differences in the nature of the review processes applied with the CLIA program and the FDA program. The CLIA program is more laboratory and survey and process oriented, and the FDA is more a product-oriented program. So I think there is potential for administrative overlap or for some dissonance if the programs weren't played off of each other. I can assure you that the FDA has a whole variety of interesting models that could be applied or frankly not applied to this universe of tests. And we frankly could choose the models according to what kinds of recommendations you provide or what kinds of classification hierarchy you develop. We certainly are not working to produce either duplication of effort or inappropriate or unnecessary work.

DR. COLLINS: Let me just follow up with a question. In the situation where we are talking about a home brew, then, for instance, where the laboratory is being inspected by CLIA's system, and you are not talking about sending out kits, where it seems to me it is more in FDA's best interest to see that kits are not so complex, that the average laboratory that has been certified in a general way might not be able to do this, but in fact you are talking about a home brew, where you have a direct opportunity for that kind of CLIA analysis. Would FDA in that situation tend to rely on the CLIA system to take care of the analytical validity part of this as opposed to duplicating that enterprise themselves?

DR. GUTMAN: That certainly is an option, yes.

DR.COLLINS: I would be very reassured to hear that.

DR. BURKE: I believe we have actually stated before that we would, to the extent there is pre-market review which CLIA does not provide, that we would expect that it would be entirely concordant with what CLIA would then require in its ongoing regulatory oversight. So I think that language has already come up in discussion.

MR. HILLBACK: I totally agree with Francis and I know that drives Francis crazy. I think that this conversation though, to me, is very useful, because I think there are three points of regulation that come out of this. I understand Mike's point that he signed out 60,000 tests, but it is not 60,000 different genes. It is related to a few different genes and I think that is the point I was trying to make in terms of how many genes are going to be in each category, or tests for how many genes. But I think we have three different points we are trying to get at and I think we ought to try to separate them. One is can the lab perform the test accurately? That is the analytic validity question. I think CLIA covers that and is getting better and better every day and when SACGT makes its recommendations, we can make sure we get CLIA the resources they need and their partners, such as CAP and others.

I think there are then two other key issues. One, to me, is are we trying to regulate the inflection point when something moves from, no, we are not ready to tell a doctor or patient anything about what we know, to the point where we say we are now willing to start telling them what we know at this moment. If that is what we are trying to regulate, then I think that is where some people, and I am going to ask Francis to comment because he is the one who actually made

the proposal to SACGT.

But I think that was one of the crucial points. That is a point, maybe a point in time, that makes something relatively more straightforward to manage and to regulate.

The third area, which is this long period of time, anywhere from a minute to a hundred years, where we are evolving the knowledge about the test. I have always been worried that this is not easily managed by an organization like FDA whose focus is on points of time when they have a finite amount of data that they review. What we talked about very early on, in some of our early meetings and the working group that we had, was is there a way to develop process, is there a way to develop a template or a set of rules about what you have to be able to talk about. I know Muin has been working on this related to a couple of particular diseases.

But I think if we can separate our discussion into these three pieces, put the analytical part in our backpacks and say we really don't need to spend much more time on that, and then focus on these two issues of the inflection point and how do you manage a process that is iterative, if I can use my other favorite word, that might last ten years, and how do you deal with Steve's worst fear that he has to manage that. I don't know how even New York manages the, okay, you've launched this test, we have done our pre-market review – how do we know six months late that you updated your knowledge base? Because there are new papers, new knowledge about the disease.

So we can focus on those two things and I would like for a minute, if I can, to get it back to Francis because, Francis, you raised the issue of increased oversight. I am not sure where you would

focus and whether you would agree with my classification and where you would focus us there.

DR. COLLINS: I don't think it was just me that raised the issue. I think it was a strong consensus of this whole committee that increased oversight was necessary as was unanimously endorsed, by the way, in which we handled the most recent revision of the recommendations.

MR. HILLBACK: No, you made the actual proposal.

DR. NOLL: I would like to take up on a comment that Muin made because Muin was showing a flicker of optimism and we badly needed a flicker of optimism. What I recalled that he said was that he was concerned – first, he is optimistic – and then he commented about the fact that when one looks to various laboratories across the country for information, along with the test results they put out, that there is a wide variety and it is often inadequate, what is actually there.

I think this is very important because I rather think we can get a long way towards where we all want to go just by having full, accurate and complete information. And that can be conveyed by a report.

I particularly want to point this out because I think that having a template that you can give to reporting laboratories, saying that these are the essential elements of the report, that you should, you must, deal with, I think we can do that and I want to point that out. You have four professional organizations represented here and every one of us, I know, will have absolutely no problem with that. The whole membership would be in total accord saying that this is a good thing. It is also something that we can work on now, start

working on it today.

Now I don't think this is totally without some problems because it requires resources, the information needs to be there. But a report, as I said before, can refer to a web site such as, and I use as an example GeneClinics, where there is full information, a synopsis and then full information.

Secondly, we are going to have to face the problem in our reports now and down the road is that, as we move more and more to a paperless record, and it is all electronic, these reports have to get into that mechanism and that becomes a little problematical. We are dealing with documentation of informed consent and dealing with complex reports that get into various issues. I think this is something that we could all rally around and agree to, which would be wonderful, and move this whole thing forward just by working on our reports.

DR. BURKE: Pat is next, but I would just like to interject because of what you just said, and what Michele said earlier, and other discussion, and that is, if we agreed that there is a test result reporting standard that we could create, that it is a very important issue that addresses a lot of concerns, it would not address the concern of data that does not exist which I think has come up. That is, that some tests will be appropriate for clinical use in certain defined circumstances, yet come surrounded with questions marks. I just sort of mark that as an issue that is not fully resolved by test result reporting standards.

DR. CHARACHE: I would also like to come back to the fact that – and I think the genetic work group at CLIA emphasized this as well – that there is pre-analytical steps here in addition to the

report form that would have to be addressed with the labeling. Such as, does a laboratory have the responsibility to decline to do a test that they know is inappropriate for a given ethnic group or a given genetic situation. What is the responsibility to teach the physician who orders the test when it is appropriate and when it is not? Then what is the responsibility of the laboratory to decline to do a test if it is not appropriate? I think this is a very key issue that was raised and I want to be sure we don't forget about it because it is a very sensitive issue. This group would be very helpful there.

DR. NOLL: I think it is important to point out that we do have some success and we have taken steps along the way and I think it is important to point those things out. But if you look at the ACMG-CAP checklist for this kind of thing, this is a very simple question. Are there written criteria for questioning or rejecting clinically inappropriate test requests. That is pretty general, but this is what the commentary says. This is what guides inspectors and what guides laboratories on what that actually means.

Many of the recently discovered disease genes subject to molecular testing are extremely complex in heterogeneity, penetrance, specificity – especially with regard to pre-symptomatic testing application of these test to patients not carefully screened and counseled can be meaningless or damaging. For certain tests, only those patients with strict family history are eligible. There are also many ethical considerations, such as the policy of not offering predictive genetic tests to children unless there is a viable clinical intervention to be initiated prior to adulthood. Because primary care physicians may not be conversant with these matters, it is sometimes

left to the molecular diagnostic laboratory to provide consultation. The laboratory therefore should have established guidelines for the rejection or questioning of test requests felt to be inappropriate on clinical or ethical grounds. Reference to the policies promulgated by professional organizations and government agencies and/or consultation with a medical ethicist may be helpful on a case by case basis.

That is already there and I am not saying that that gets the attention it should, but it is there. The important thing, it has been accepted by all of us. We have bought into all this already.

DR. CHARACHE: The second point I wanted to make has been elegantly illustrated by Walter Noll and by Michael Watson. The ACMG and CAP are two of the 22 groups who have deemed status through CLIAC to review laboratories. Not all of the 20 other groups who will review laboratories apply similar standards. Certainly HCFA does not have such a checklist for the labs they review. So one of the points I would like to emphasize is that CLIA, through these deemed status and what you are just hearing through Walter's very elegant summary, some very thoughtful work, is that one of my thoughts would be that it would be helpful if the SACGT, I think working from the documents that are already prepared and they are wonderful documents, through the ACMG and the CAP, come up with such a checklist that all of the agencies that review under CLIA, all 22 of them, would have to apply to genetic test laboratories.

That does not exist right now. We have two very thoughtful groups and even then we have to address the issue of the availability of people who do the reviews who can apply these checklists in an appropriate manner.

The second point I would make is that there is a lot of very thoughtful work that has been done in terms of how these groups should be reviewed but is not being used by the majority of laboratories which are reviewed. About a third are reviewed by CAP, about a third by the JCAHO and the whole group by the other labs. So I think there is an opportunity here.

The final point I would make is that I would just emphasize that the FDA and their review process, for the things they review, look at more than just the analytical validity. They do demand clinical validity before they will approve the test. So we should not think of it as just analytical on the part of the FDA. They also do demand clinical validity, as does CLIA by law.

MS. UHLMANN: As I have been listening to the discussion this morning, what I tried to do in my mind was to think about existing tests. For example, just how this would be put in place is, you take a test like Huntington's disease. One gene, one gene change, whether you do the test predictively, whether you do it prenatally, whether you do it diagnostically, it is a very accurate, simple test to perform. Versus something like HNPCC where we have more than four genes and about a 60 percent detection rate.

So as I look at this I almost envision it is like a drug insert, where you have uses – the use of the medication for the actual condition and you have use of the medication during pregnancy, different stipulations that are listed. So I would almost see for something like Huntington's disease, for example, that if the test is used predictably we come up with a standard definition that would be on any genetic test. If this test is used predictively, informed consent is critical and would list those different components. Or if the test

was for a condition that has low penetrance, that that would be stipulated. This is a test for a condition that has low penetrance and by low penetrance we mean the following. In other words we would have some broad categorizations that could be developed that would just be if the test is used predictively, if it is a test for low penetrance because otherwise if we take a look at each genetic test one by one, it is way too complex. I think you have to somehow make it so it is a very simple and broad classification that can just be used. If you come up with just some standards that if something is used predictively, then X, Y, Z needs to be in place. So it is more almost stipulations for ordering that test.

DR. WINN-DEEN: I just wanted to take the point that Mike Watson raised about half an hour ago, which is using HCFA or the creation of reimbursement codes or reimbursement by private insurance agencies as a gatekeeper for whether a test is proved in sort of general accepted clinical practice.

I think one of the holes in the system right now is that even a test that has gone through the FDA process, their sort of highest level of scrutiny, when it comes out the other end with an FDA approval rating, it is still not automatically reimbursed. And we need to tie that last bit of facilitation of these tests getting out into the market so that when a group of experts believes that a test is ready for that kind of widespread use, so it is approved by one government agency on efficacy and analytical validity, and yet the other side of the government is still not accepting that for patient care, I think it would be really an important thing for SACGT to try and come up with some rationalization that when a test passes some level of scrutiny, it also is reimbursed. So maybe when it moves into

this final category that it is reimbursable or if it is a kit test, when it gets FDA approval that it is reimbursable, and not create yet another hurdle that the test has to go through so that a patient can actually make use of it.

MR. HILLBACK: I have to say that I am scared to death of asking people who spend money to add another level of regulation to genetics. We have enough things going on. The problem we have and the problem with those payers, and Genzyme does a lot of tests and we get paid for most of them, but they take a lot of homework to explain why there is a clinical utility. But that clinical utility is constantly changing. We are constantly going back to them and saying we know more now than we did yesterday. I just don't think they are the appropriate parties to start asking to do that. I don't disagree. It is frustrating when you cannot get something paid for that is helping the patient. But I think we ought to separate that from this regulatory process and make sure we are understanding from a clinical benefit basis how we regulate. Then we will have to figure out how to deal with the reimbursement folks separately.

DR. WINN-DEEN: I guess my point was just that at some point, where there is consensus in the scientific community and some blessing by that community, that it ought to have some automatic switch that it is then reimbursable, period. And there is not another big debate and discordance among all the insurers, Medicare being sort of the leader, as to whether that test is reimbursable.

DR. LEWIS: I think what you are getting to is the access issue which is something that we spent a fair chunk of time discussing and I don't think it is unique to genetics. Part of my work is in the area of infertility and I just got the booklet on my own personal

health care insurance and is is very clear to the fact that this is an uncovered expense. Certainly that is an area of health care that has been pretty well established.

So I think we get into part of what we want to do is make it that we have access to this information, that we don't develop into a two- or three-class system where people who have resources can get care and people who don't have resources cannot. But I think that is separate and apart from the issue on the table today and I think that access is something that we spent a fair chunk of time addressing in our report that is not available for all of to share yet at this point, but I think that is a separate issue from the one we are talking about today. But I agree with you that it is an incredibly important piece. It may be I have a different take on it than other people.

DR. BURKE: Actually, I think your point is well taken that the issue of access and therefore remuneration of tests really isn't on our agenda at the moment, but we may want to mark that as an important issue for future discussion.

DR. TUCKSON: I can truncate my remarks. Judy got to it. I think the key thing is that if we do this job well today, that provides so much of the basis of information that medical directors and others who have to make those choices will be able to use to accomplish this task. But I do think it is a second order agenda item.

DR. VOELKERDING: I guess what I am struggling with is this. I think what you are really asking, which I think Elliott raised about this point of inflection, is when is a new genetic test really ready for clinical use. I guess the two examples I think of in terms of very broad applicability in the last half decade was the introduction of Factor V leiden testing and then subsequently hereditary

hemochromatosis testing. I think in both of those instances, I think that we are still in a very long process of data collection about understanding both the use and the validity of leiden testing and for hemochromatosis, and I guess I see as representing many people that do laboratory testing, professionals, it really gets at the heart of practice of medicine. That is, when do we practice a certain aspect of medicine? When do we introduce a new surgical procedure? When do we introduce a new medical management? When do we introduce a new genetic test? I think this is one of the most challenging things. I guess what you are trying to conceptualize is how would we appropriately govern the introduction of new genetic testing and I don't feel we have really hit at some of the models. I guess one model that has been proposed is to have the FDA perform that and our concern, my concern representing laboratory constituencies is will there be enough appropriate, shall we say, other medical practitioner representation in that process? So I am hoping that we can maybe tackle that particular kind of issue. This is an issue in part of the practice of medicine, regulating that.

DR. BURKE: Just as a comment on previous discussion and not wanting to sort of reiterate discussion that has already reached some level of resolution, we have recommended, SACGT has recommended that all tests have some form of pre-market approval, all genetic tests, that it is appropriate to classify them into a hopefully reasonably large group that is routine and has a streamlined process and hopefully a somewhat smaller group that requires higher scrutiny. That the pre-market approval will be under the aegis of FDA, but with appropriate collaboration with, as we put it, other agencies, private sector organizations and public representatives.

So I don't think we want to re-debate that issue, but it does seem to me that a lot of discussion comes around in Mike's phrase, that we cannot regulate medical practice. To me, that gets us back to what we were talking about when we were evolving the concept of what Elliott called the A, B and C – test not ready, test is okay for use but has lots of questions. I think that is where the temptation to regulate medical practice comes in and where we really need to focus on what is doable and what is appropriate in terms of test regulation. Then finally test okay for use and we kind of know how to do it. We are not so worried about those kinds of tests.

So it seems to me we really are talking about that middle category. We have enough information about clinical validity, that we would like to make the test available, but there are still lots of uncertainties about that test. What I think the discussion has circulated around is first of all a lot of interest in figuring out how many of the problems inherent in that kind of test can be solved by appropriate pre-test and post-test information. That is to say – and I think that gets to your point, and Mike's point, of not regulating practice. If there are tests that come into use around which there are a lot of questions, a number of those questions may be dealt with by appropriate pre-test information that lays out for the provider and consumer a variety of uncertainties about that test or caveats about its use. Then when the test result is made available lays out some uncertainties. For example, major concern in a lot of genetic tests is the lack of predicted value when the test is negative. The sort of non-informativeness of the negative test, for example, which is different from a lot of other kinds of tests. But it seems to me, in moving our discussion forward, I am sensing a lot of interest in making

sure there are appropriate pre-test and post-test information provisions, per what Walter said, and it sounds like a lot of good work has already gone forward that could be built upon.

It seems to me a major question for us falls into are there other outcomes of the review process that really inform us again about test classification, being our issue, to help us identify which tests are high scrutiny. The two other items that have been discussed, that we should talk about, are the possibility of some sort of informed consent/counseling requirement being tied to tests. And I think we really have to look critically at whether that is in fact trying to regulate clinical practice.

The other big issue that has received a lot of discussion in previous issues and we have not talked much about is ongoing data collection. I think what I have heard is, yes, we know there needs to be ongoing data collection to resolve the questions about those tests in that category. We have heard from New York State that is not easy to do. So it seems to me we really have to ask is that part of the outcome of this review process?

DR. KHOURY: You have a nice way of summarizing things. I just wanted to continue on my previously stated optimism and focus a bit more on the data collection. I see it as an integral piece of the initial review that would lead – because most tests would fall under B, using Elliott's classification – and in that vein, and given the diagram that Joann put forward on the analytic side versus the clinical side, there are two types of data collection which you heard me talk about before. They are very crucial and both important. One is further data collection that would update the parameters of genetic tests, be they analytic or clinical or utility-wise, I mean, on the

continuum of tests, so that people would be more informed a year down the road or three years down the road, that these parameters will change.

The second kind of data would be collecting data on the genetic testing process, which is sort of what New York and CLIA will try to do. You have to have characteristics of the genetic test in theory as related to the actual knowledge base, but what is really going on in the field? Are people getting informed consent? Are labs doing analytically that they need to do? What are the forms for requirements for reporting how are these being handled in Wisconsin versus Missouri, et cetera. So data is very important and key both as a testing process itself -- pre-analytic, analytic, post-analytic and then improving the parameters of the genetic test, per se. But having said that, the point is we are studying right now is that you cannot collect on everything and the ways we have to deal with that and get perhaps the initial regulatory teeth in it is which ones of these require that additional data collection. Going back to Joann's diagram, on the right-hand side you have this yes/no. Unfortunately most of them also fall under the yes. So similar to what Mike said, the complexity on the left-hand side, unfortunately most of them fall under yes, at least in the initial scheme. But forget about the analytic side. On the clinical social side, most of them right now are under yes. So as we devote further discussion in perhaps the next section, I think we need to try to focus on really which ones we need to have a concerted effort to get the data collection piece on in a very systematic way, actually required by the pre-market approval process.

DR. BURKE: Can I just make sure I am understanding what you

are saying, Muin, that a lot of tests will be in B. That is some reason for clinical use but still lots of question. And we perhaps should start thinking in terms of classification of those tests in classifying those that where various uncertainties that are still there when a test comes to market are well addressed by pre-test and post-test information provision and the energy should be put on standardizing those templates. Whereas there is a subcategory where the uncertainties are of such significance that energy and resources need to be put into ongoing data collection. Is that right?

DR. TUCKSON: With that summary there, Muin, I need to ask the New York people to help me. I understand the collection of the data pre- and post-test are to continue to refine the utility decisions and so forth. But this process of what is going on in the field, whether or not informed consents are being obtained, to me this is a very different kind of data that has to more with compliance with the rules and regulations. It seems, and frightens me, as administratively cumbersome and expensive process. New York, should we be – I mean, this is a whole different kind of game here. Is that something we should be involved in in terms of these kinds of rules of overview and scrutiny of whether people are following the rules or not? And does it cost a lot of money to do something like that?

DR. CAGGANA: The way our system works is we have inspectors that go out every two years and they simply do spot checks. They may ask to pull out five reports or five forms from samples or whatever disease that you offer and then they check that there are consent forms on file, that the report was signed and issued and that is expensive, just to do that. That is every two years. So the cost of that kind of follow-up, we don’t have administrative wherewithal to do that. So

that is a difficult issue to follow up.

With our data collection, we ask that all prenatal samples that have been studied are followed up to make sure the outcome was correct. A lot of laboratories do that just by submitting a card to the patient to have sent back after the birth of the baby. The other issue we check, and we have revised our questionnaire that should be going out soon, I think, is our questionnaire data is based on the number, the types of tests that you do, the number of tests that you have done and then also we have added components for methodologies, how you do the test, and whether you do prenatal, predictive, diagnostic, et cetera. So that data will come in but it will be for the year before. So that is sort of how we handle that, but we don't have a good mechanism to do that kind of follow up.

MR. HILLBACK: Michele, the other question we were trying to talk about a minute ago was also can you follow up, or do you inspect that the lab has changed the way it talks about a test? Because two years later there are a lot more papers out about the disease, there is a lot more information about the disease, whatever disease it is. Is there any way that you follow up that they are telling what they know and what they don't know? You approve the test initially, but do you have any follow-up mechanism to do that?

DR. CAGGANA: We don't have a real mechanism to do that, but it sort of comes to us anecdotally. Sort of getting to the issue of which tests, which high scrutiny and which low scrutiny tests need to be followed up, we have anecdotal evidence also, and a lot of this revolves around the market obviously. But one example was when CCR5 delta 32 tissue came out, a lot of labs applied to us to do that test clinically, and there was very little use of the test. There was no

real clinical utility for the test and so most of our labs dropped it. I think maybe one lab still does it. Most of them have dropped it and so that is sort of how we get a handle on which tests are being used.

MR. HILLBACK: You should sort of clarify what that test is.

DR. CAGGANA: For HIV infection, CCR5 is a gene that deals with how the virus binds to the cell receptor and there is a dilution of 32 bases in that gene. So if a person is homozygous, has two alleles for that gene, they are thought to be less susceptible to HIV infection.

DR. KHOURY: I'd like to comment on Reed's comment. In my mind it is probably easier to have some kind of data collection going on on parameters of genetic testing versus genetic tests, because that is lab-based. You can do ongoing surveys, maybe periodic surveys, but to update the parameters of genetic tests, if you think about clinical validity and utility, especially for low penetrance, or you have to conduct studies that might require years and a cohort set up like following people for some disease or testing interventions.

So I think collecting the data on genetic tests and their usefulness is probably much more expensive than having some kind of system, national or state based, where people are trying to have a handle on the genetic testing as a process. So my bias is that it is more expensive to do the former rather than the latter.

DR. TUCKSON: I just wanted to make sure, given that we really want to delve into and bring to some closure the characteristics that we have in front of us now in terms of determining what should be the characteristics of the classification system and so forth, as opposed to focusing, at least in terms of this session, on whether or not people are complying with the things that are set up. I just want

to make sure that I am okay with you if I separate those out as two very different kinds of things.

DR. BURKE: Walter will be the last one before we break.

DR. NOLL: I wanted to make a comment on costs. The New York program is extremely important as a model as well as how it works. I think we would want to look to it for guidance on a model on a larger scale. Marie did raise the question that things are expensive and I had asked to show how the program runs, how it runs financially. She is going to correct me, but I think what we need to understand, it is a costly program and the genetics part of the New York program is not funded solely on the income from the genetics labs. It is funded from the income from all of the laboratory activities. So what we have is – correct me if I am wrong – is an expensive program with very much hands-on attention by experts that is dealing with this many tests, and is being funded by all those glucoses and BUNs and all the rest, which are this many tests. I think we must remember that, that this is not flying on its own. It is flying on sodiums and glucoses.

DR. BURKE: It is time for our break and although I am not sure we explicitly addressed all the bold points in this first discussion, I think we got to where we needed to get to. After our 15-minute break, we are going to look at proposed strategies for test classifications. I think we have defined this category B as being of interest. We have examples and we will look and see how those work with different kinds of classification schemes. So a 15-minute break.

(Brief recess)

Review of Proposed Strategies for Test Classification

DR. BURKE: As we start the next session we realized that it

would be useful, first of all, to just review how SACGT defined a genetic test at our last session. We actually spent some time on that. So Joann is going to read for us the definition.

DR. BOUGHMAN: Only because those of us who have been at the meetings cannot quite remember because we went through so many iterations and those who are new would benefit as well. In this context and for our purposes, the SACGT has been defining a genetic test as an analysis performed on human DNA, RNA, genes and/or chromosomes to detect heritable or acquired genotypes, mutations, phenotypes or karyotypes that cause or are likely to cause a specific disease or condition.

So in fact we have limited it to human specimens, but not necessarily only the heritable or germ-line mutations. It could be acquired mutations. And the recognition that they need to cause or be likely to cause a specified disease or condition.

DR. NOLL: May I comment on that? The community that I represent would have found it much easier to deal with if you could separate the heritable from the acquired. I imagine that something that would satisfy me and others of us would be when you had your discussions you would say, okay, now we are going to talk about heritable, and then when you have the next discussion you would say now we are going to talk about acquired. I think you may raise all kinds of confusion if you try to talk about them together.

It is also important, at least there are some of us in that community who lose some credibility when you do that because we think then that you are confused.

DR. BURKE: Actually I would like to recapture some of that discussion because we certainly talked a lot about heritable versus

somatic. I know there were some important reasons I think they included that. Some genetic tests might be used in both instances. In other words, it was not necessarily easy to identify genetic tests as solely somatic.

DR. COLLINS: I remember this conversation because it was one we debated considerably and I think there was a concern that when you set out to do a test you don't know whether you are going to find a somatic change or a germ-line change. You cannot cleanly separate these things in some instances. You can in others. I think, to reassure you, Walter, the point here was to have a broad definition of the tests, but then to do what we are about to do, which is to focus on which tests really require additional scrutiny and which ones can basically be considered low risk and not require that kind of oversight. By having a broad definition, we place even more importance on deciding how to categorize tests into those that need attention and those that don't. I think when we get to that discussion, if you are talking about a test like Her2-neu, where you really do not expect to find a germ-line mutation and if you did it would be totally reportable, that somebody had a germ-lime amplification of that locus, then the concern about many of the aspects of this as it relates to heritability is quite different. That is going to factor into the next part of the conversation.

DR. BURKE: And in fact I think we specifically identified that a circumstance where a test was done solely to identify somatic mutation might be an example of a low scrutiny.

DR. COLLINS: We did.

DR. BOUGHMAN: Now that we have talked about some of the nuances in that first sentence of the definition, I need to get into

the second part – a genetic test also is the analysis of human proteins and certain metabolites which are predominantly used to detect heritable or acquired genotypes, mutations or phenotypes.

We get it too broad in one breath and we will be back where we were in this room about a year ago or a year and a half ago.

DR. BURKE: Other comments on the definition of genetic tests? I would like to ask Francis to just review with us the sort of scheme that helps us to see where IRB, CLIA and FDA fit in.

DR. COLLINS: Yes. I found in the earlier part of the discussion, which I think was helpful in getting a lot of issues on the table, that we were sort of converging on a pattern of classification of tests in terms of is it research or is it patient care. But we did not entirely enumerate what the consequences of those were.

And I also though maybe by going to A, B and C implied that we did not have some divisions, and I am going to argue that A actually has an important subdivision that we skipped over. Just for you to shoot at, I put up this table which attempts to say what many of you were saying, but organizes it into this kind of table.

So instead of A, B and C, I am talking about A1, A2 and B and C, because I think the research phase, which is pre-approval by any FDA review, can be divided into research studies where results are not given back, which can still establish information about clinical validity, and studies where the research is given back. And I was a little worried that A2 was almost being lost in the discussion, as though you would always go straight from A1 to B. Clearly a lot of research that we do in attempting to collect clinical validity and even utility information does involve giving results back to patients. We are doing lots of those with NIH funds right now.

Then B, of course, I think we agreed the inflection point is at that transition, between A2 and B. So when you get to B you are approved for patient care, but we all are agreeing in many instances you still need more data collected because you don't know enough about this.

Then ultimately one hopes to get for tests to the point of C, and we can disagree about how many are already there. What I put on this table was an attempt to try to show what kinds of approvals and other parameters are attached at each phase, and many people talked about the importance of having pre- and post-test information provided and it seems like that applies to every stage except the first one, which is essentially anonymous testing. Ditto with consent, although consent if a little different in A2 than it is in B and C.

But then there are also these other oversights. The IRB process is obviously very significant in A1 and A2 and probably also in the part of B where you are trying to collect more data to further refine what you know about utility and validity. CLIA kicks in as soon as the results go back. So CLIA does enter the process – at least I think this is right – perhaps before FDA, if you have an A2 phase. Then the FDA oversight obviously in B and C.

I don't if that is helpful or not. It may be that it is wrong in some ways, but it helped me sort of organize what a lot of people had been saying.

DR. CHARACHE: I have what I think is an important comment that I don't think grossly changes what you are saying, Francis. Could we please make that a plain A, and then make it B1 and B2. It is very clear that by the time you get to what you have under A2, what I want to call B1, that this is patient care and I want to make that major

distinction, when it becomes patient care.

DR. COLLINS: But isn't this just semantics?

DR. CARACHE: No, it is not semantics in terms of the researchers who don't know that they are now doing patient care. So if we could just make that A and then B1 and B2, I would be very happy.

DR. BURKE: Or should we go A, B, C, D?

DR. CHARACHE: No, because really A2 and B are parts of a continuum.

DR. COLLINS: Okay, I actually started with A, B, C and D and I was sure that would screw everybody up because we already went to A, B and C.

DR. CHARACHE: I just want to make it A and then B1 and B2.

DR. BURKE: Do we have support for that?

DR. CHARACHE: That separates straight research from research plus patient care.

DR. KHOURY: When would the FDA process kick in?

DR. CHARACHE: We can talk about that. I would think that before patient care there should be a checklist that has been met.

DR. BURKE: But it does seem to me that it is a very critical issue. I think the usefulness of this table and maybe we should keep it up and talk about it a bit, is precisely to be sure about when FDA approval kicks in.

DR. CHARACHE: But if you can make that B1 and B2, then it becomes patient care. I want to separate straight research where there is no patient care.

DR. COLLINS: The FDA approval is still in but at B1 and B2.

DR. CHARACHE: That is fine. I am not arguing that. I am just saying when it becomes patient care it becomes B.

DR. KING: There are research tests in which information goes back to the family in CLIA-approved labs that are not patient care.

DR. CHARACHE: If it goes back to the patient, that is the definition of patient care. That defines it as patient care.

MR. HILLBACK: I think Michele and I are over here talking because her comment was A2 is illegal in New York State. If you use New York State as the surrogate for FDA in thought process, you cannot do what is now B1, what was A2, before New York State approves it. If you are going to do patient care, New York State is going to look at it first, no matter whether you call it research or anything else.

DR. COLLINS: So you want another arrow for New York State?

MR. HILLBACK: I guess the question really comes back to the one that Wylie asked, which is where does FDA get in? Is it before any patient care or is it before big patient care, lots of patient care, which would be an interesting question.

DR. COLLINS: I cannot imagine how this could really be a serious discussion. Are you going to say that research is supporting an effort to try to understand clinical utility and validity early on in the development of tests has to already have gone through FDA approval before you can get to that point?

MR. HILLBACK: I thought that was the point you were trying to make five meetings ago.

DR. BURKE: I think this is an extremely important point. I think we should go through and make sure we resolve it.

DR. WATSON: I think there is an important distinction and that is that even in New York State, those two steps are quite different. If you were in what is now B1 in New York State, they would

be saying I have this orphan test or developing test which I know a little bit about and we need to sort out, and you would say okay. That would be very different from the process you would do between B1 and B2 when you are saying okay this is generally accepted.

DR. CAGGANA: So B1 might be like our non-committed lab approval test. But that is the loophole. The law says that if a patient result goes back, if the result goes back to that patient, it is no longer research. I know it is sort of a bad way to say it, but that is how it is. That is sort of the black and white of how you have to think about it.

DR. BURKE: I think, Michele, you may need to explain to us – you mentioned loophole in passing, but it sound like that loophole is there because it has to be. So explain that a little.

DR. CAGGANA: When the law was written, it was written such that it is research anonymous, no data returned. That is research period. The difficulty that occurs is exactly what Mike is saying, where you have an investigator who is doing a study or an NIH-funded study and that individual wants to report results back. The way the law is written in the state that is illegal. That becomes a clinical test.

Sort of the intermediary is the access issue where there is a family that has a rare disease, there is one laboratory doing it, they need the test, so we have to be able to give an approval. But we give the approval with a caveat in the letter that goes out to the ordering physician that states that this is outside our review. We have not reviewed it, we just know there is a lab that offers it. We know there is one lab in the United States that does retinoschesis testing. Any test for that goes to that laboratory.

MR. HILLBACK: This is the point that Pat made. This is the

continuum that, once you are giving out test results, you ave a commercial test. You may not like the fact that research labs do commercial tests, but they do, because you are giving out clinical results. It is a clinical test and it is a little arbitrary to draw the line and say FDA we are not going to let you in until later. I guarantee, Francis, if we draw the line there, there are a lot of commercial labs, like Genzyme or pick anybody else, and they'll say we are doing research on the patient. We have our board-certified MDs and PhDs, so we don't have to go to FDA yet. Let's call it home brew and let's not go to FDA.

DR. BURKE: Are you arguing FDA has to come in at B1?

MR. HILLBACK: That is the question. Where is the crossover point? If you create a loophole, you are going to find that everybody is now going to join that loophole. I am not sure that is what Francis wants. I thought we said before we go to clinical practice, before we give data to patients, we are going to have some pre-market approval.

DR. WATSON: But you presume a loophole is a pejorative term.

MR. HILLBACK: Okay, but if market is giving clinical information to physician to manage patients – maybe that is the wrong definition, but if that is how you define market, in BRCA 1, until people gave out clinical information to patients, nobody cared. As soon as you started giving out clinical information to patients, people got upset. Is that the point that FDA should say it is now okay under high scrutiny rules – you can't give out any information on any patient until you pass that. If it isn't then, you are into somewhere dividing a continuum with an arbitrary line.

DR. BURKE: I know we have a lot of comments and I want to

make sure I get everybody. I've got Ann, Judy, Bob, Paul, Bob – anybody else who wanted to comment?

MS. BOLDT: I guess my only comment is I do see some of this tied to reimbursement at some point and I really liked A1, A2 better than I liked B1, B2. Because I do see some of that having an incentive to go back to get that tied into reimbursement and do agree with Michael Watson that is going to have to happen, in terms of the counseling services and tests.

DR. BURKE: You are then arguing for FDA review between what we now call B1 and B2?

MS. BOLDT: Exactly. But I still like A1, A2.

DR. CHARACHE: You can put the dividing line any way you want.

DR. LEWIS: I just want to note that I think this is a continuum and I would argue that even when you get down to C, there is still some ongoing research going on. I would see that basically what you are looking at, even when you get down to C, I would say you have a proved established, but there is still new knowledge development going on there and I think to start arguing over what we call it becomes a little bit arbitrary because it is a continuum from no results to patient and a lot of research to less research and more results going back to patients – this is a continuum both ways. I think the opportunity we have here today is to figure out what should be and how we should do it so if we spend a lot of time looking at what is – I think Francis has put up a model and we could really look at what is the way we want it to be. I think that is our purpose for being here today but to spend a lot of time – and I hear what Pat is saying about when is it research and when is it patient care – but I would argue

that even basic research is being done to better patient care, even though it is being done totally in the laboratory. I would hope that is the purpose of it all and I would like us to look at this more as a continuum rather than trying to spend a lot of time looking at what category should be what.

DR. MARTIN: Actually I wanted to make the same point. Basically what has been put up here is what is. What we are doing here is not arguing about what category should things be in, but where do we want to be in the future, what do we want to change. So for example, if we want FDA scrutiny to come in at a different point, that is one of the decisions. This table describes what is today. I would say if you were to put New York State up there instead of CLIA, you would be covered under this same table.

DR. WATSON: In fact, it may be just predictive. It might shift your FDA review up if it was a predictive test, then you might want a little more look at that transition inflection point. But for most other tests, I think it is appropriate where it is.

DR. BURKE: So I think we are saying two things. Bob has just said this is what is. Certainly the FDA review we are still deciding about, but to the extent that it has been decided it is here. And we need to have as ongoing discussion through the course of the day whether this is the model we want to stick with and use as the basis for test classification, or whether we want to shift where the arrow is.

DR. FERNHOFF: I think it is important to leave the investigator with some option – who is developing a test once it has gone through IRB and it can give out results – but it seems to me another level, the reimbursement issue, when another lab wants to start

charging for this. So some consideration of that needs to be built into the system.

DR. NUSSBAUM: I was going to say that trying to have the FDA approval at what is called the B1 level is unnecessary and superfluous primarily because of the need for good IRB and consent efforts. You can do research and clinical care at the same time. That is what B1 is. The fact that it is research and is recognized as research, that is why it is necessary for IRB approval and formal consent. I think it would actually have a chilling effect on a lot of work to institute that kind of FDA review at that level.

DR. WILLIAMS: Opening a can of worms, I don't see why FDA approval is not instituted at A because how is this different from introduction of new drugs or investigation of new devices that FDA regulates? Generally the FDA is involved at the A level.

DR. BURKE: DO you have a comment on that, Steve?

DR. GUTMAN: Sure. The FDA makes an odd distinction, at lest in devices, between research and investigation. We actually use the term in a non-congruent way from the rest of the world. For us, research means what Francis or someone said before, you are just sort of fishing around for some general population-based insight into disease or the physiology of a disease process. You actually don't have an intended use. You actually are not ready to look at the precision of a test with sensitivity or specificity because you don't know what it is being used for. You are hunting for a use.

At least for a commercial test, not for a home brew, but for a commercial test, when it actually has a use it becomes investigational and the level of oversight of the agency frankly depends on the nature of the test and the investigation. If it is a

blinded test and the investigation is not reported back, it would be unimaginable for us to get involved. The minute the test is reported back to a patient we suddenly do become interested. We become interested enough that we require a special vehicle called an IDE, or investigational device exemption.

The IDE does not necessarily have to come to the FDA. If in fact you can make the argument that it is not a significant risk to the patient, the IDE can be managed through the institutional review board informed consent. If it is really scary and has powerful information and it does look like it has significant risk, then the agency would want to review it. In the area of in vitro diagnostics for tests in general, there are not very many IDEs. We may get from 3 to 6 in a heavy year or we may get none in a year. So that is the deal for us.

DR. COLLINS: Could I follow up on that? So would it be reasonable – I think that was very helpful to sort of put us in the context – would it be reasonable then – so what we are really talking about as part of deciding that a test deserves high scrutiny, one aspect of that would be whether you do want to insert an IDE-like step between A and B1.

DR. GUTMAN: Yes.

DR.COLLINS: I think what Bob says is, in the majority of circumstances to do that could be quite choking to the enterprise. But in the places where you are most concerned about things going awry, maybe that ought to be one of the things on our list of considerations for looking at those high scrutiny tests.

DR. BURKE: We actually have Reed, Elliott and Joann.

DR. TUCKSON: Just a process check. This is a very useful

place to focus our discussion. Should we be trying to think now of other characteristics to array and see if they fit in there? For example, counseling – is counseling necessary? I think counseling may be an important variable. Should we try to squeeze counseling into these holes here or is that what we are trying to do now, start to develop that list and see if they fit?

DR. BURKE: I think that is a very useful comment and I would tend to say yes. I would also want to insert some sort of – I mean, we have said an IDE type of approval. And also, the issue of data collection, whether it is going to be in some form mandated might be a question, but I think what we are finding is that this table is something we can build some of our discussion around.

DR. COLLINS: Would you have pre- and post-testing information?

DR. BURKE: It could be because that could just be a label – information available.

DR. COLLINS: Where would you put counseling in this category?

DR. LEWIS: Could I just make a comment? It seems to me – I guess it depends on counseling with a big "C" or counseling with a little "c", because it seems to me that every time a patient goes to a health care provide and gets back the results of a PAP test of anything else, you are getting some information. That kind of information, depending on the patient and depending on the provider, that interchange is very unique and very special. And some patients have high levels of understanding and some patients have lower levels of understanding and some people need more explanation and some people need less, and some things are more complicated than others. So I

think that when we start to talk about counseling and to say there are some test that should be done with no counseling – I mean, people get results back and if they don't know what they mean it has been pretty useless. But it seems to me that is a part of all encounters and sometimes it just requires – the level of preparation of the counselor may be what we want to be looking at, not whether counseling takes place.

DR. COLLINS: Aren't we already on the record on this topic? That counseling should accompany genetic testing?

DR. BURKE: Yes, and it may also be that this falls under the issue of how much it is reasonable to expect that we could regulate. I guess the question that really is before us is whether there would be some tests or some testing circumstances where part of the regulation, part of the oversight of the test used, would be that counseling above and beyond what we would consider a normal part of clinical care, some unique kind of counseling, would be required. That may well fall into the category of things that people have already suggested are beyond this regulatory pathway. But it seems to me that is the question that is at least raised.

MR. HILLBACK: I guess I still have a problem with trying to draw lines across a continuum. I don't know what simple definition could be written that would define the transition from B1 to B2. By definition in B2, we are saying we really don't know everything that we will ever know. We don't know very much even of what we will eventually know about a test and I don't know how different that is from B1.

DR. BURKE: Actually, Elliott, I think the table does suggest a very simple definition and that is everything in B1 is under

IRB review.

MR. HILLBACK: By Francis' table a lot of things in B2 are still under IRB.

DR. BURKE: No, I don't think so.

DR. COLLINS: If you are doing follow-up data collection specifically as part of research, but that is not an integral part that is required in B2.

DR. BURKE: I think we need to get back to the data collection. I would offer a simple definition, going from B1 to B2 is that you know want to offer the test not under IRB review.

MR. WATSON: And I think you can adjust the continuum by - is it a diagnostic test in which you know 10 percent of the mutations or is it a predictive test? If you lump all genetic testing together, you are stuck with an unmanageable continuum that I don't think you can ever work your way through. But if you separate out diagnostic applications, where we know this is the disease-causing gene but we only have 5 percent of the mutations, it is great for that 5 percent of people. It is awful for the others. It is not useful for the other 95 percent of people. Nevertheless it is a useful test for some and that is different from predictive where what you don't know is very different in that setting versus the diagnostic setting.

MR. HILLBACK: Okay, but I am still not comfortable.

DR. CHARACHE: I see the definition separation between B1 and B2 as very nicely expressed if we say that B1 requires IRB approval and B2 does not require it. I think that also helps avoid the situation where somebody remains at the B1 level simply to avoid oversight.

When we have said, however, that we need pre-market review

before patient care testing for all genetic tests, I think we will have to be careful in terms of how we define pre-market review. If it is not a full FDA type of review, whether it is very stringent or non-stringent, then perhaps the review would just be at the investigational device level type of review prior to the B1 implementation. So I think that as we flesh these out we will come up with the definitions of where the FDA gets involved and how stringent it should be.

DR. NOLL: When we speak about a test here, I think people have been talking about it as if it was a test for a new disease or something like that. What happens if someone decides to do the same thing and just uses a different restriction enzyme?

DR. CHARACHE: If the predictive value of doing it, the sensitivity and specificity are modified by the use of a different master mix or a different enzyme or a different series of primers, it is a new test. Then you have to show the clinical validity of your new test and it may just be that you say there is no difference between the first one and the second one and you have shown you have clinical validity.

DR. NOLL: I don’t want to be facetious, but I can change the magnesium concentration.

DR. CHARACHE: That is correct. That is one of the challenges of saying the FDA is going to review all of these when they are home brews. That is why they have stayed out of the home brews.

DR. NOLL: It is hard for me to follow the discussion and sort it out in my mind without being able to visualize what is going to be happening at the FDA. What information is going to be asked for and what people will do and so on and so forth. If we get into individual laboratories and recognizing that there are hundreds, if not thousands,

of these laboratories across the country who modify, who change the magnesium concentration or any other thing, are these all new tests?

DR. BURKE: Can we let Joann make her comment and then Steve may have a comment.

DR. BOUGHMAN: I think I can lead into the response there by the FDA. In fact, when Francis first put up the table and he inserted te arrow for FDA review, he simply put something on the far left of the table that was already on the right-hand side of the table. It was when the checkmark started showing up in the FDA column. We could put something out on the left-hand side saying CLIA, between the first two steps and whether that chokes individuals or not, it is in fact currently a requirement that if a laboratory is giving results back to a patient, they have to be a CLIA-approved laboratory. Now there are some investigators out there who are not happy with that.

But one of the other comments I would make before Steve talks about some of this – we are not really talking about inventing new processes. We are talking about adapting processes and whether it is CLIA or IRB or a federally-funded and uniform study in multiple sites that has undergone a different kind of oversight and review process, it really is now filling in the gaps. I want to remind my colleagues around the table that the reason that we have been called here as people whose opinions are valued is that we have a great deal of experience and our perception of what is good and right may be very different from the small laboratory who could purchase a box that has the right agents in it and go merrily along their way not knowing how little they know in the process of giving results back to their patients. From a regulatory mind set, remember it is that end of the spectrum that we are trying to protect our patients from, not the

highly experienced laboratories and researchers and clinicians who have been debating this in depth for years. Very different ends of the spectrum.

DR. GUTMAN: I would just like to put on the table our perspective of what we have to offer. You know the FDA, at least the group that I represent reviews data, looks at thresholds, and then tries to establish labeling and we are quite used to the tension between that and labeling and trying to make sure that products meet standards and are correctly labeled. We normally, for all products, look at analytical validity. It may not be necessary in this product line, as Francis suggested. That may be overkill. We would never approve a clear product based only on analytical validity. We always are looking for some measure of clinical validity, and it may vary in its robustness. We are always looking for some ability for a physician to actually use analytical data. Having analytical data would not, under the current models subject to change, allow a product to be cleared or approved. So we need literature or data to explain how that analytical data would be used and we would want some measure of clinical validity as well, at least a surrogate endpoint or some understanding of how the information would be applied.

What we have never done or almost never done – we don't actually look at outcome. So we never do an assessment of utility. An example I have given for the SACGT before I came is PSA tests, where we approved it for screening with in fact no certainty that was actually going to produce a long-term medical benefit.

We have over the last couple of years have played around with a variety of new review tools and we are better positioned to address whatever the charge might be from this subgroup or from the

group in general from the Department (of HHS). We have a variety of very well-established and fairly labor and data intense tools of PMA and the 510(k) are our usual review processes, but we have on the table a wide and wild variety of abbreviated form of those tools which allow us to do all kinds of things, including conform to standards or approach almost mechanisms of self-certification and to use smaller data sets and different data sets.

So we are in a position, frankly, to create tools that might fill the gap here that we probably would not have brought to the table four or five years ago. And we also have different processes. We have a mechanism that I am not incredibly enthusiastic about, but it is certainly on the table, a process to delegate out authority, to contract third parties, to provide the same kind of deemed status that CLIA does. So there are both, in terms of the packages we use to apply those tools and even the processes we apply to these packages. There are really wide ranging options we can bring to the table and the scenarios you create will clearly create a response from us.

We have done a lot of brainstorming and have models that range from the most conservative to the most avant garde on the table.

To address specifically the question you raise about modifications, we have a lot of familiarity doing this. It may not be that we got it right. If you talk to industry they will tell you we have the wrong thresholds and we are not least burdensome enough, but for modifications there is a very clear path. The path is if you change a buffer or solute or you make a minor change and you don’t change performance – then we don’t want to see it. We want the entity to document that they have made that change and to keep that on file so if we ever inspect they can prove that it did not create havoc. For

God's sake, manufacturers would be shocked. They change weekly or monthly either to improve or cut costs or do something else.

When there is a significant change, we have a new package which allows manufacturers to provide abbreviated 510Ks which essentially say we have a design control quality system in place, we have conformed to that system and we have in fact had significant changes but it has not affected our output in a clinically meaningful way. So we get submissions that are processed in 30 days.

Then of course when you make really big changes, if you make changes in intended use or changes in technology, we do ask for some submissions. Again that also is subject to changes in more conservative or more liberal directions, but we do have a sense of trying to prioritize them knowing that all modifications are not created equal.

DR. KOENIG: I have a smaller point that goes back to the original table that gets at the A1, B1 or A1, A2, whatever we are calling the distinction and pertains to the issue of results being given and the results not being given back. It has always been a concern of mine because I think it is very crucial. I just want to throw out that one of the things we may need to do is to make sure that as we come up with some kind of scheme that we do not forget to include the IRB wing of that, because I am very concerned about different regulatory groups having conflicting kinds of ideas. I have been in many situations. It is a particular issue because so many IRBs don't really understand genetics research very well, and sometimes they are requiring results to be given back in situations where it is illegal. I just want to make sure that people are not in that situation and we may need to take responsibility as the group to try to make that

seamless somehow, so that everyone really understands what everyone's responsibility is. It is not always that clear.

DR. LANIER: I would like to shift the discussion just a little bit. In thinking about the B2 group of tests, I have been trying to think about what levels of scrutiny we might have. I think that is the total purpose of our meeting today. It seems to me that we really need to focus more on the risk-benefit ratio and I think the discussion up to this point has focused more on the benefit side, in terms of what is going to be useful, what is going to be less useful. I think if we really want to get levels of scrutiny we really need to focus more on the risk side of that, and particularly in terms of potential harms that could result from an individual test being done.

So I would like to see us shift in that direction, thinking both in terms of what the emotional risk would be in terms of anxiety, depression to the really more physical harms of mastectomy, oophorectomy, of potential abortions. That gets into a more ethical, value-laden discussion, but I think we need to start thinking about that.

DR. BURKE: I think that is a great segueway because what if we look at this scheme? It seems to me what we are really saying is that B2 is the point where we are going to decide whether something is high scrutiny or not. We have now some sense of what that might mean. I think if I can capture some of the discussion that I sensed there was consensus about, we are also saying that when you go from B1 to B2, and you are now no longer under the IRB detailed informed consent procedures realm of research, where anybody getting the test is also a research subject, we would like to see all tests under a standardized format, provide appropriate pre-test and post-test information. That

is, that the availability of the test comes with some information about its appropriate indication and the test result comes with some information about what that test result might mean.

So the question really is, in test classification, it seems to me, is when does a test require more than that? When does a test trigger concerns greater than that? Your point is that harms potentially deriving from the test might be an important parameter.

DR. LEWIS: I think the other thing is, is it the test or is it the use of the test. Because you get tests and then you start – a test can be well established for one thing and then an off-label use may put the test back up in there in an A or a B1. So I think we need to look not only at the test, but at the use of the test so we are sure that – some tests may be in two or three categories at the same time depending on what they are being used for, as we start to develop new information.

DR. CHARACHE: I would like to second what Judy just said. I think that is very important, but I would like to suggest that the issue of scrutiny really has to begin when the results are given back to patients, in terms of deciding whether this is a very dangerous test or not. So perhaps one could establish categories that would determine when B1 needs more attention as opposed to waiting until you reach B2, because there is an awful lot of harm that is done at the B1 level at the present time, by laboratories that do not mean to do harm, but who don’t understand what is required to give information back.

DR. BURKE: Yes, Bob, why don’t you comment and then we will go into a little more detail.

DR. NUSSBAUM: I was going to say that the issue that Pat brought up is an important one. There are now steps being undertaken

in gene therapy trials, for example, that will probably be spread out over all human experimentation, of having data monitoring boards required for all human experimentation. So I am trying hard not to have a situation where we have multiple people overseeing the same issues. We should try to keep things clean and separate so that we don't get overlapping and perhaps contradictory kinds of things going on. I would hate to se the FDA involved at the B1 level when there are data and safety boards in place and IRB in place overseeing patient care research at that point.

DR. CHARACHE: I am concerned from my own experience with the knowledge that the IRBs often lack the information base that is required to assess the clinical validity before information is given back for patient care, or the analytical validity of the laboratory that is doing the task. We certainly had that experience. It has only been in the last three years that all of the IRB work that is now done in my institution, if it is done in a non-CLIA approved laboratory, actually I review it. And we have been able to assist a lot of our investigators in doing higher quality work and it is now part of the checklist. They do require review in our own institution. But this is not standard. Most IRBs don't have the information base required.

Now I don't think that means that it requires FDA oversight. I think you can set up a system for review that is appropriate and is not burdensome, but does protect patients before information is given back.

DR. BURKE: It does see to me that it is very crucial to test classification to decide this point about the extent to which B1 is adequately oversighted, if that is a word, by IRB. So Wendy is going to comment and then I would like to ask other involved in

clinical research, Richard, Paul and John particularly, to comment on this, too. That is, are we satisfied with IRB oversight at B1? Is it a matter of improving problems or errors in IRB or do we need other oversight there?

MS. UHLMANN: I just wanted to make the point that having coordinated testing for patients for years, that for rare disorders, tests for a lot of our patients don't make it past B1 and that B1 is actually their only way to get anything of clinical use. It is never going to become a clinical test because the condition is so rare and there is no lab that is going to make efforts to have that done commercially.

MR. HILLBACK: I guess reserving my right to still find it difficult to draw lines across a continuum, as I thought about this, to think about these different levels of A, as we have it now, being managed by the IRBs, whether they can do that or not is a crucial question. But in a conceptual scheme, let's accept for a minute that we can get them to a point where they can. Then to accept, back to Pat's point, that once you go to giving results back, CLIA must be involved and, for the moment, let's accept that IRBs are still the other gatekeeper there, along with CLIA.

There is some elegance in this sort of increasing involvement of different groups. Then when you go, at some point, which I think is very hard to define, but maybe we can find a way eventually to get there, to a broader distribution that you what SACGT has requested, which is for FDA to get involved. So I think there is a certain elegance to this, as I sat back and thought about it. I still think it is hard to define when those steps happen and smarter minds than mine can probably figure it out. But I would hate to say let's

move FDA into the first step, because the IRBs don't do a good job. I think that goes back to the point someone made about are we looking at what is or what we want to have happen. I do think Bob is right, that IRBs, because of what has happened in gene therapy and other places, but that one happens to be the hot button, are going to get a whole new set of pressures on them to change and maybe we are going to become part of that pressure.

So I do think there is some logic here that does make sense. Defining what is where I still think is going to be a mess, but I reserve that for a later moment.

DR. FERNHOFF: I think the IRBs, in my own experience, are improving, but many still obviously have a long way to go. I would agree that the area for improvement is at the level of the IRB.

DR. BURKE: In other words, that IRB oversight is appropriate as long as IRBs do what they should be doing.

DR. KING: I can only address my experience with the IRB, but the IRB has been on a rapid learning curve in the last few years in terms of dealing with genetic testing and any aspect of genetics. In fact what Bob suggested, requiring oversight committees, is in fact sort of the critical part of this. That is also true of the research clinical centers that are often the place where the testing is applied and the clinical research centers, the general clinical research center oversight committees are now requiring more IRB and oversight committee reports or status for the testing.

In terms of the B1 and the transition between B1 and B2, in my own experience there are situations where I have information that is interesting to families. It may be relevant to some aspect of their care, but probably not. It is more generally interesting and much more

particularly interesting to me or to people who are doing research with me, and that has to do with certain types of genes we are analyzing. There is other information where we are doing research that is directly relevant to the ultimate care and family counseling of the family. So it is going to be very hard to transition between B1 and B2. It is going to require some definition of the other aspect of the disease or the condition we are testing for and we are going to have to factor in other aspects of the condition, to make that separation.

DR. BURKE: But acknowledging those ambiguities in the element of the continuum, do you think it is workable to use the IRB or oversight definition in this sense, which is to say it usually is crystal clear whether a patient is a subject under an IRB or not?

DR. KING: I think so. From my standpoint this table is actually rather crystal clear to me in terms of that aspect. I think there could be some haziness in one test versus another, but I think the mechanisms to go through the things we discussed earlier, and I really have not participated very much because I have been on a steep learning curve myself, but in terms of labeling and information and counseling and those things, are rather pretty straightforward too many of these conditions. I think the point of where to assign the oversight and to use the consensus of this group to assign the oversight, like the IRB, is a very valid point and a very important point. From my experience, IRBs are responding to that kind of information.

DR. BURKE: And it sounds like – I think part of where I hear this discussion going is that once we are clear on where IRB oversight is a critical piece and where it is not, that part of what we determine about test classification and parameters of concern ought to

be fed back to IRBs, that we ought to work on some mechanisms to develop guidance for them.

DR. WILLIAMS: My only concern is that the level of oversight by IRBs is going to be variable across the board from institution to institution. It is really going to depend on how knowledgeable they are in terms of genetics.

DR. KOENIG: I think this is a very productive and useful discussion and I think we could make a real contribution with some recommendations about the IRB issue. It seems to me that two models have been on the table. One is the idea of a mandatory data safety monitoring kind of thing attached to the IRB. The second would be more the FDA model of the IDE. I guess those are two different models we could think about and we need to think about them as two different things that would help at that step.

One thing, I was just called by somebody in my institution to serve on a data safety monitoring board for a complicated piece of research. It pointed out one problem to me is that, at the moment, if we do decide to make that sort of thing mandatory, we also need to make a recommendation about funding for those kinds of things because at the moment, for example, in NIH-funded research, these are being required but there is no mechanism for funding these, which I think is a problem.

At any rate, just to point that out as an issue, if we do take this on and we do recommend it, we have to think about how to actually operationalize it.

DR. CHARACHE: I think the point that the IRBs currently, in many institutions, lack the ability to provide the oversight that we want just emphasizes the importance of this group to define clearly

what we expect the IRB to do, what we expect CLIA to be monitoring and what we expect the FDA oversight to accomplish. I think all three are going to be important, particularly as the FDA is trying to develop avant garde systems for review because of the tremendous volumes that they are going to have to address. We have to be sure we don't end up with lower oversight of genetic tests than we do for glucose and urea.

DR. MANN: This is along the line of what Pat just said and what Bob Nussbaum previously said about level oversights and not overlapping. I just want to bring up that particularly in gene therapy, elevation of the previous OPRR in the Department, and that office I think is like OHRP – the new name – is that they are looking at this issue. As the Department we are actually working in drafting a response to the ethics advisory commission report that we probably should work in tandem with what other activities within the Department, and particularly that office should probably be consulted into what activities they have regarding IRB.

DR. TUCKSON: Also along the lines of Pat's analysis, I think that what I am hearing and what resonates at this meeting is separating out the IRB in two different ways. First it is a determinant for our classifications because there is a role, a function of what the actual IRB does. It is a place when something becomes – it is research. It is that moment when it starts to make that transition line. It has a purview over a thing. Regardless of what you call it, there is a function that is a significant difference.

Secondly, there is a mechanism, a process called IRB. What I am fearful of is, given the problems and the concerns across the board in this country for whether IRBs can do what they are doing now, given the concern for all the worries that people are looking at how we

are going to shore up IRBs, that if we put all of our eggs in a basket and say, well, the IRB will solve it, what you are going to have is a public concern for this is being adequate today and will make the IRBs worse if you put more pressure on them.

So I think first, intellectually we want to segregate the intellectual function as a cut line. But then secondly we will then have to grapple with the meaning of how do you help IRBs to accomplish this, which is a second order equation. But it is one that I think has to be attended to, maybe not in this forum, but ultimately has to be attended to.

DR. BURKE: I just want to interject a comment that it does seem to me, particularly the comments that Pat and Reed have made and also others, that it is very likely that as we do a job that we are satisfied with in terms of classifying tests and particularly identifying those parameters of tests that are of concern, that generate concern, and outline what kind of data format or educational format and what kind of information ought to be provided before a test and after a test, what kind of information will be required as a result of an FDA process, that that is an appropriate sort of package of information to feed back to IRBs and say to an IRB, when you are considering the possibility that genetic information will be provided to a patient because the research is in the B1 category, bear in mind that once that test becomes commercially available, these are the hoops that it will have to jump through. That may be kind of guidance that really helps IRBs to do a better job.

DR. NOLL: I have difficulty not following the discussion, but in making this a framework for myself, to absorb what is going on. That is just one observation.

The second one is that there are a number of things here that really concern me. I am bringing them up because I think they concern the people I represent as well. So I am just going to say a few things.

First of all, the notion of requiring FDA approval on these home brew tests is disturbing to me. The laboratory that this is going to affect most are the academic laboratories. It is not going to affect commercial laboratories nearly so strongly. Academic laboratories, like my own, and other across the country, what we thrive on, what keeps us going – we are there because we serve a greater purpose in the institution and that is innovation, primarily. Innovation and teaching.

So if we are not constantly innovating and publishing, then we are dead. Requiring that innovation then to be approved before it can be put to use, I see that is potentially a real downer for us. As I said, this will affect academic labs across the country. It will not as strongly affect large commercial labs.

It will squelch innovation. Why would you bother to go ahead and improve performance by X percent when you are a commercial laboratory and you already have the volume and you are doing it, when in fact you have to then go through another approval process. I don't have the details of that, but I can see this is not going to encourage you to make changes that are going to help.

I am concerned, too, about – getting back to this distinction between acquired and heritable genetic disorders – if we are including on the chart here acquired genetic variation, which might be applied to analysis of tumors, for example, there is absolutely no intellectual reason to stop them at nucleic acid changes, but then

there is a whole spectrum of changes that we would not think about as calling genetic right now. There are the chemical studies, immunological studies – all these other kinds of diagnostic studies. If you start treading in there, then you are going to capture the attention of a lot of other people who are going to say, look, do we really want the FDA getting into this kind of thing?

There is no barrier, there is no reason not to extend that oversight to every other kind of home brew assay that is taking place in a biochemical lab or a histological lab, and there are tons of these kinds of things.

I think we are probably most concerned about tests which eventually will acquire a certain volume and be applied to populations. I will make a prediction here. I rather think that most of these, if not all of these tests, are soon going to find their ways into becoming kits. They are not going to be done by my lab with my own gels and things like that or anyone else's. They will be automated tests on platforms that will be run in a clinical chemistry laboratory probably. If that is the case then they already will be going through the process, and that is going to happen in a short period of time. There are established FDA processes for dealing with kits and all that.

I think the important thing here is for all of those things that we do there, is that if we are going to supply information for patient care purposes, that we supply that information with full disclosure of everything that we know about it, everything we don't know about it and we have to take responsibility for it. And that responsibility I think can be assumed in some cases by IRBs and in some cases it can be assumed by whatever.

I am wondering, and I have not thought this through at all,

but those kinds of statements in just acknowledgment that this test has not gone through an FDA approval process has certain meanings to it, and attaching that to a test might be sufficient.

DR. BURKE: I just want to comment that some of the issues that you have raised I think will come up again and will be dealt with under the orphan drug definition discussion that we planned for this afternoon. I think they are very important points that have certainly come up in Committee discussion before. Once we have clarified what we consider, quote, routine versus high scrutiny kind of definition, do we then need to talk about orphan tests and what the implications of a definition for orphan tests would be. So we will get back to that later.

I am going to ask for comment from Emily and then before we break for lunch, I want to move from that comment on to asking Michael, Wendy and Bob Nussbaum to just talk briefly about the potential classification schemes that they have been working on.

DR. WINN-DEEN: I had wanted a point of clarification that I would like to ask this group for, because it is not really clear to me what triggers the transition from B1 to B2. Is the trigger that we now have multiple institutions who are all operating under independent IRBs and at some point, when there is more than some magic number n institutions doing this test that it should now come to FDA review? Is it a scientific trigger based on a certain amount of data that has been collected and, if so, who is going to decide what the trigger is? I think this is a nice schematic, but it needs a little bit more thought as to how you actually make a decision to move.

I mean, A to B1 is a really clear trigger. The trigger is that you are going to report patient results. But I don’t understand

what the trigger is between B1 and B2 to initiate FDA review, nor do I understand if the FDA review is of a test, saying that we now believe that factor V leiden is a legitimate test for multiple labs to be offering, or if each lab independently has to go solicit FDA and say, I would now like to run factor V leiden. So is there a list that FDA is going to compile that says these tests have passed this first level of scrutiny and we now believe that they should be regulated under CLIA instead of under IRBs.

I am just really throwing some questions out because I don't understand exactly how we think that transition is going to be made.

DR. BURKE: I will just make a comment and then let Kathy and then we will move on. From a clinician's perspective, the transition from B1 to B2 is as clear as it could possibly be. It is absolutely clear from a clinician's perspective and that is, you have to have IRB review or you don't. That is a very clear distinction in clinical medicine.

DR. WINN-DEEN: Sure, but who decides that? How do you know? So you are now the tenth institution that wants to institute factor V leiden and you go to your IRB and they so, okay, we think this is a legitimate thing to test. How many institutions do we allow to just run everything under local control before we say this is a test that should now be bumped up to the next category? Within one institution I don't think it is an issue, but it is the idea of multiple institutions and the dissemination of a technology beyond just the initial research study into more common use.

DR. BURKE: Again, I think it is pretty straightforward in the clinical area. I would really like other clinicians to comment on

that. At a certain point the lab needs to charge a fee for it. It is very simple.

DR. WINN-DEEN: The trigger is that you want to charge for it?

DR. BURKE: Yes.

DR. WATSON: In New York it is lab by lab – three peer-reviewed articles in the literature, though I don't think it gets specifically to how those peer-reviewed articles develop performance characteristics. It is the fact that they exist and say there is validity in doing this that it moves to being chargeable.

DR. BURKE: One comment from Kathy and then we really need, before we end this session, to have some discussion on classification schemes. Then we will pick up after lunch.

MS. HUDSON: My only comment was that I thought that one key part of the transition was charging for the test and that you would not market a test under B1, you would not advertise.

DR. WILLIAMS: But again, at what point do you come to that conclusion?

DR. BURKE: Let's tag this as a crucial issue and pick up after we have talked a little bit about test classification schemes, because this is obviously a very important point of attention. So let me just ask the three of you to talk briefly about test classification.

DR. NUSSBAUM: Let me just say briefly up front that this was a draft that I put together quite a while ago. It was translated from Macintosh to PC and back to Macintosh and with that happening there are some errors in it. I will show you what they are so you can get it right on the forms that you have.

It was motivated by a couple of ideas. One was to try to

keep the stringent rigorous review to those tests that really need them, which means the ones that are widely applied and have potential harm that could produce an inappropriate alarm or false reassurance. I decided to establish one that was three-tiered because I was a little afraid that a simple two-tiered with a one point difference between the two-tiers might look a little arbitrary, be able to shift the test just by one point and it would take it from rigorous down to not as rigorous.

I also thought it might be good to have a weighted system so that some factors are more important than others in making a decision about whether it needs a rigorous review. Also the context in which a test is being done – tests don't exist in isolation. Using the test in one setting is quite different than using another, so parenthetically the issue that was brought up – and I know we are not going to talk about it now but which needs to be thought about – the off-label business, does bother me because I don't think you can separate the stringency with which a test has to be reviewed from the context of the application through which it is being put. Which means that if it is reviewed at a low stringency appropriately in one context, and applied in another context where it really should be a high stringency, I am not sure how to deal with that. So I am just raising that as an issue.

Finally, as I said, this is a draft and there is actually no ego involved in this. First of all, please make the following corrections. Under 3 and 4, carrier testing and testing asymptomatic, at-risk individuals, it should say see below, not zero and zero. The see below refers to little numbers 1, 2 and 3. So the idea of the scheme, as you can see here, is that a diagnosis of a symptomatic

individual, a prognostic assessment of a symptomatic individual, I assign a score of zero. I should mention here that the stringency score goes as follows. Level one is the lowest level of review and level three being the highest, and this being some sort of intermediate, and I didn't even presume to figure out what these might look like at the FDA level.

The important factors to take in mind were first of all, how widely used is it going to be? To a certain extent this is designed to help us with the issue of orphan tests. The second is, which I think is really one of the most critical issue, is the potential medical benefit from the testing is. Third, really a critical issue, is the penetrance and/or the predicted value, and it differs. I am sure Muin can talk about that. But taking some of this into account is critical.

Then I did separate out from these four the issue of fetal testing because of the different between a nutrient test of fetuses versus one that is targeted pregnancies that are known to be at increased risk in the general population.

So this is the scheme and these were what I thought the critical factors were. I tried it out on a few of the tests that were in the book and I think it pretty much went the way I would like to have seen it break.

DR. BURKE: Implicit in this would be a decision. If we liked this kind of scheme and agreed with the idea that three zones is useful to capture some gray zone issues, I think there would still be the question of is it only the ones with the highest scores that kick into a sort of high scrutiny FDA kind of thing.

DR. KHOURY: Bob, which way did it cross? Can you tell us

because I tried to do it and I think I liked the results.

DR. NUSSBAUM: You liked the results? For example, APC I-130 13-7K testing – I had prevalence got two points, interventions available got three points, detection rate got four points, so the total score was nine. Medium chain dehydrogenase deficiency – prevalence got two points, interventions three points, penetrance was one for six – that is up there. BRCA 1 and 2 – three points for interventions, two penetrance, so that was five. That is above four, that's in the rigorous area. Same goes for microsoundings of tumor tissue when it is the context of looking for germ line. I did not say anything here about non-germ line mutation because, to be absolutely frank about it, I didn't even think about somatic mutations requiring rigorous review. That is my own personal opinion. I am very happy to be told that is actually wrong. So somatic non-hereditary – to me that was really – part of my bias with this is not to demonize genetic testing. So if you are testing for somatic change, it has none of the characteristics of heritability and implications for other people in the family and the implications for offspring, I saw as something being less impelling.

APO E-4 testing – let's see.

DR. KHOURY: They all fell into the high category.

DR. NUSSBAUM: Pretty much they all did.

DR. KHOURY: Can I make one comment here? This is tremendous effort here and I would say the same about Wendy's a bit later on. But it struck me that I think we have all the right elements there. I live in a world of numbers, fortunately or unfortunately, but in this case I think the numbers might play tricks on us somehow. It seemed to me, following the argument that it looks to me like one or

more of these factors will kick it into high gear? So by your design you give five points when there is no treatment and four points when the penetrance is less than fifty percent, and when you look at all of these you see there is a lot of judgment value. When is it helpful, not helpful.

So anyway, my own take on all of this, although a numbering or weighting system would be extremely useful if it can be managed, I have this idea that I would like to change throughout the day that hopefully somebody else will straighten my mind, that it is probably not workable that much. Maybe a qualitative approach, whereby you put a certain segment or fraction of all the tests in this high scrutiny box – the one I am thinking of right now is something very simple. The ones that are predictive, unknown penetrance without treatment – this is clearly, everyone agrees on those. I am wondering what to do with the rest. This is where I am right now, but maybe we can have some more discussion.

DR. NUSSBAUM: Actually, Muin is right. There are certain things that are weighted that would kick it into high scrutiny. Those are not the ones I am worried about. It is the tests that just on every ground just don’t quite make it, but the points will add up and that will kick them into the high scrutiny.

DR. KHOURY: But when you add them up it look like all genetic tests will end up there and then the system of classification may not work because all of them will end up there.

DR. NUSSBAUM: But they won’t. There are lots of clear examples. For example, any test for a symptomatic individual with a genetic test and there are other examples where you will only get one point.

DR. BURKE: I was going to say that I think once you are above 90 percent, it doesn't matter too much whether you are 93 or 99. I think the bigger distinction comes in the range between 50 and 70 percent and 70 to 90 percent.

DR. WATSON: But the problem is they are not mutually exclusive. If you are diagnosing a symptomatic individual with a unique phenotype, with a test that has 20 percent penetrance, then you have a person who is going to get this high scrutiny despite the fact that they have a phenotype for which this gene is clearly associated. Because penetrance is purely a population issue. Then we get a score of zero because it automatically trumps everything else on the board.

DR. NUSSBAUM: It is only 3, 4 and 5 that is clearly below. 1 and 2 are just zero.

DR. BURKE: I think it would be useful to look at the other schemes and be able to talk about them in the context of each other. Wendy, do you want to comment?

MS. UHLMANN: When I set out to develop an algorithm for oversight of testing, I realized that everybody was trying to duplicate what it is that we do in the clinical setting. What are the variables that we discuss with patients. So in terms of the initial model that I put together, it really has both a combination of analytic variables as well as clinical utility variables. Obviously a ten-variable system is too complex to work. In the ensuing weeks, since I presented this at the June meeting, I've been trying to look at it to see if there was any way to tone it down to just a couple of variables. I know a lot of these can be collapsed, especially the ones that are in the bottom row. What I still keep coming back to is the type of test that really matters and the detection rate. In other words, if you are going to

undertake the test, what are your chances of actually finding something? And then the penetrance. If you find something, what is the chance it is actually going to be meaningful or have clinical implications for that patient?

I think where prevalence still factors in is whether you are going to be doing this on individuals that are high risk, in other words, people who do have a family history or an ethnic background that is going to increase their risk. And in terms of the issue of surveillance and treatment, some of that is actually really back to clinical utilities. How useful the test is going to be for that particular individual. I think as we well know, with many predictive tests, there is no clinical benefit to doing them, but there is a lot of benefit to the patient in terms of using that knowledge to make informed decisions about their life. In other words, it is something that is not being used medically, but it is information that is being used by patients to make decisions that are just as meaningful for them.

I cannot say I have come up with the answers. What I did try to do on the airplane was try it out again, and I think what you have in my handout, how I did it initially with some conditions– where Huntington's disease being done diagnostically in a symptomatic individual, I think we can all agree, it would be a low oversight test. Whereas if you are doing it predictively, it would become high.

The question there, the next one, the familial adenomatous polyposis, if it is being done diagnostically I have it here as high, but you could actually debate that. That is what Dr. Burke and I were trying to talk about earlier today, this whole issue that in some ways it is a very straightforward test you are doing on someone, but where

the implications come is, is that you then want a doctor to say, okay, there is no mutation and therefore you don't need any follow-up, you don't need to be regularly screened. It is a whole issue that it is a test that only picks up 80 percent of the gene changes, which means that 20 percent of the time that person could have FAP and you are not going to be able to detect the gene change with today's technology.

I was able to get some into the low oversight category. I think that is important because not all tests can require high scrutiny. I think there are ones that we can agree on are low oversight, or ones that, for instance, something like sickle cell anemia may start out as high oversight as you are first initiating it, but then can be quickly moved down to low oversight once you have figured out the caveats and how it should be overseen in that population.

As a genetic counselor, I certainly would not want to be counseling everyone that needs to find out whether they are a carrier of sickle cell trait. I think likewise we can say the same for PKU newborn screening, that you don't need to be counseling everyone. What you need is to see the people who have positive results.

I know that this one is going to be a little hard to read, so I apologize in advance. It is my scribbling on the plane, but what I was trying to do is see if I could somehow collapse the different categories down and get it down to just a couple of categories. You see up here at the top I was dealing with type of test, prevalence, surveillance treatment, locus heterogeneity, detection rate and penetrance. Down at the bottom I did take out locus, heterogeneity because I realize that in some ways you can pick that up with detection rate. For example, a condition like HNPCC, where there are greater

than four known genes, that would give you a higher score on the locus heterogeneity, but it is also captured in the detection rate because there are more than four genes, the detection rate right now is only about 60 percent. So you actually do pick that up in that category.

So down there at the bottom, what you see is I was trying to put it through I13O7K and with APO-E4, predictive, et cetera, CF carrier status and I guess I would argue that I think CF diagnostically can be done as a low complexity test. I didn't come up with a cutoff here, but kind of eyeballing this I would say that maybe less than 10 could be considered low complexity and somewhere between 11 and 14 might be medium, and greater than 14 would be high. Those are the discussions I need to have with people here in the room, to sit down and kind of go through and plug in different examples. So this is kind of my first pass at trying to look at all of this and make some sense of it.

DR. BURKE: Great, thanks. I want Mike to comment. What is interesting as we look at these in the context of Muin's comment about qualitative factors is that there is already a fair amount of agreement I think in these schemes about what the key characteristics are. Mike, do you want to comment?

DR. WATSON: I think what I have to say is really not the CDC lab consortium because of communications colliding heavily over the last four weeks. We are nowhere near anywhere on reaching a consensus. But we certainly thought more from the perspective of Muin – simplicity and more qualitative factors. If I had to make a slide, it might have said predictive tests over that way. In fact it is a little bit more than that. I think the two places where we saw the highest need for scrutiny I think in general discussions is that predictive testing area

and in the area of the translation of new tests, where mechanisms to insure both the maximum collection of information and the maximum dissemination of that information so there is some uniformity in the way people say what we know and what we don't know about the test – that would be the two critical areas, I think, where we would see things beyond the current system. I think most of the things -- I can see where these algorithms might help us define where we move to increased levels of oversight within our inspection programs or increased levels of oversight in different regulatory bodies. But that for developing things that don't exist now, it was really in the translational steps and in predictive tests where there is no incentive to any one individual or institution to be able to do what it is going to take for the long term to collect that.

DR. BURKE: It seems to me there may be a convergence on some of the things we have been talking about. That is, that these classifications schemes may really help to direct the kind of data formats or information formats we would like to see the blanks filled in on for pre-test and test result information. It maybe that as we go forward in this discussion, how much can be accomplished in that way will be a very important issue.

We are actually a little overtime for lunch. We are planning a working lunch. What I would like to propose is that people take 5 to 10 minutes to go out and get their lunch boxes, move around, stretch and then come back here and eat lunch and we will talk over lunch, continue this very discussion.

(Recess for lunch)

A F T E R N O O N S E S S I O N [1:05 p.m.]

Social Issues

DR. BURKE: I think we left our conversation having heard some things about classification schemes and also I think having a couple of concepts laid out. One was that whatever classification scheme we come up with will need to be fairly simple to be implementable.

The other is that it might well be that one could take efforts that we have already seen and look at the qualitative points they are making and use those to get to a classification scheme.

Now, Steve is going to talk to us about some ideas that he has developed.

DR. GUTMAN: This is just one more model to throw on the table. This is completely unvetted. It hasn't even been vetted by the

work group at FDA that does this.

We have had probably more heat than light. We have a working group in this division and have gone through classification. The classifications actually failed - I thought they would be relatively easy - they failed because from our perspective they were creating too many tests in the high scrutiny category. We are actually very sensitive to resources. We would frankly like to have few tests in the high scrutiny category. This is totally not well grounded, but I am going to throw it at you anyway. It seems to me like it might be more akin to what Mike might have cooked up had he had more time.

Its starting point is based on whether there are more than 4,000 tests per year because if there are less than 4,000 tests per year right now, the regulatory option that would go to a manufacturer would be a humanitarian device exemption, which the only requirement is to show the safety not the effectiveness of the product.

So, if we were just intuitively looking at tests that would deserve low scrutiny, it, frankly, would be not a great place to put resources. It would be tests that have lower than 4,000. It is not patients per year or disease states per year. It is tests per year. So, you could have a rare test that you were doing a lot of testing on. That wouldn't necessarily count.

So, if it is was now, we would all be horrified or delighted, Huntington's Disease, less than 4,000 tests per year. So, it would go in the low scrutiny category. The next thing is the test predictive. If it is not predictive, it is a diagnostic test and, again, we would put that in some kind of, I am not sure what kind of category - low, intermediate, moderately low, slightly low category - but we would push them off a little bit to the side.

If you stop and just use predictive, of course, virtually all diagnostic tests when we went through them sort of had a predictive -- or lots of diagnostic tests had a predictive flavor. So, we tried to make a second cut. One, when they are predictive, when they are being used to predict in a sort of broad population - Ashkenazi Jews, African Americans, Italians. Are they being used in a general population or are they being used in a select population of people who had identified cases in their family or symptoms. I don't know if this is a real definition or not, but if we thought there was some selection to the predictive use of the test, we said it would go into a lower scrutiny.

If we thought that it was just being used in everybody of a certain category, then it went into a higher scrutiny. I don't know if that is reasonable or not. Then the last was the most nebulous. You will see that I still am very hung up with the difference between having effective therapy or not effective therapy. But we put together two very, very common and difficult FDA terms we have used in risk assessment. Whether a test is used as a stand alone or adjunctive test and whether the impact is significant or not, I merge them and I make the either brilliant or horrendous mistake of saying, for example, hemochromatosis is an interesting disease, but there are other ways of identifying it besides a gene. You probably would be doing CBCs and making decision.

I don't know. That may be too simplistic. I said the same thing about APO-E, I said it doesn't matter whether it is cardiovascular disease or Alzheimer's disease. You are not going to be able to do much except perhaps worry more.

With cardiovascular disease, maybe you would do more

cholesterols. I hope you wouldn't just treat with absolutely no other information. So, I would relegate them to a lower class. Karyotyping isn't population-based screening, but a very odd constellation of tests. I forced BRCA-1 into that category because I wanted to make Francis happy. The other ones that struck me as - this derives from the FDA work. I did this this morning because I got in an hour early.

Canavan's disease and the medium chain fatty acid disease were the ones that I put in high scrutiny from this group and I wasn't quite sure what to do with the others. I wasn't quite sure whether everything else was low scrutiny or whether you made that high scrutiny or whether that was moderate scrutiny. I didn't try to refine it, but I was trying to define in as narrow a sense the highest level of tests that I could.

Then because it seemed that numbers were so beguiling to this group, I took the same scheme and I just forced the numbers to try and agree with the flow charts. We prefer flow charts at the FDA. So, I said that if you were a rare disease, you got zero points and if you were a common disease, you got 30 points because I won't have to make the point that was a really important distinction.

What you could do is change all these numbers, change 30 to 4 and each of the 5's could be changed to 1. So, you would have a simple numerical system. But the deal here was that I wanted to make sure that having all of these combined, if it was a rare disease, would still not allow it to be kicked into a high scrutiny. That may be absolutely awful, but that is what I tried to do.

The bottom line is I came up here with a number. If it is less than 30, it is low scrutiny. If it is more than 45, clearly we wanted that to be a high scrutiny and what I wasn't quite sure is what

to do with the stuff between 30 and 45, where to make the cutoff and whether they have a medium scrutiny or not.

I don't have any allegiance to this. I put it out as an alternative. I thought it was simpler. Maybe it is not that simple. It begs the issue of what to do if you have a disease where you don't know the penetrance. You don't know the locus heterogeneity. You might not know the population distribution exactly because, again, at the FDA, we are very hung up with intended use and we presume that whether you really knew a lot, a little or something in between about a disease, you would have an intended use and presumably indications for use, which are the population which you are about to apply.

DR. BURKE: I would actually like to make a couple of comments on this because I think it is very helpful. First of all, you have shown us a way to look more qualitatively at things that are also looked at in a second way. It may well be that this flow chart might help figure out what is past the threshold may be the more functional approach.

In terms of the last issue, the sort of penetrance issue that you say your scheme doesn't identify, I think including that in the definition of a significant risk test might be a way -- in other words, I think you have created that category as a way to capture some of those issues.

I am sure others will have comments about where you put different examples, but just for starters, I think cystic fibrosis is an interesting example because as far as carrier testing is concerned, it would go all the way; whereas, perhaps some diagnostic testing within a family with no mutations would fall in the earlier range, but identifying it this way, one might be able to figure out where

different uses of the tests fall.

So, I find this kind of attractive and I think it would be interesting to look at the other characteristics that were incorporated in the other schemes and figure out where they fit in your scheme.

Other comments?

DR. WATSON: I am trying to look at it from sort of a worst case scenario perspective, but a disease like perhaps Canavan's or even CF. I am trying to see where you spin out the intended uses so they are clearly dealt with differently because when you define it by disease with multiple intended uses, it is less clear and you don't necessarily want a Canavan's family-based carrier to be constrained all the way out here by a population-based screening study.

DR. BURKE: Then, isn't the solution just what you said, which is that it is the definition of test use that has to be incorporated. What that triggers immediately, though, is the very interesting issue of off-label use; that is to say, to what extent do the harms potentially associated with a test need to include the harms that could generate from off-label use. Because if you okay a test by some streamlined low scrutiny procedure and then it turns out that that same test could then be used under circumstances that we would consider creating a significant risk for patient harm. I don't know if that needs to be figured in the oversight.

MR. HILLBACK: I guess what I am asking is if you put in the lab director or someone at the lab in the position of saying, well, you are doing this to analyze one particular family for Canavan's versus someone who doesn't have a family history necessarily and, therefore, we won't do it for you because we don't have FDA approval to do it where there is no family history but we will do it where there is

a family history.

You are starting to create the lab having this incredible complex web.

DR. BURKE: I think there are two pieces and maybe the APO-E4 will offer us a good example to discuss this. It seems to me that we first have to get back again to that theme of to what extent should this oversight mechanism regulate practice. If we say, for example, that a given test has a given indication - the test has gone through review - and we are clear on what that indication is and we are clear that that is the only indication or that list of indications is the only indication, then the question before us, I think, is whether very standardized, very complete ways of making that clear to ordering physicians are sufficient versus that there needs to be something in the oversight mechanism to catch people and then to get into what the lab responsibilities are.

I must say, I tend to favor a method that clarifies, as well as possible, and in a very standardized way what is the appropriate use of the test.

DR. KING: Wylie, can you, for those of us, at least for me, that hasn't been involved in this, what, in terms of oversight, what would you be considering in terms of oversight as a test once we set some guidelines and directions and information on tests?

DR. BURKE: I think that is actually a central, and not fully resolved, issue. I mean, I think you can say at some point where we are at this point, where the committee is at this point. I will try and say -- and, please, other committee members correct me -- that we have said all genetic tests should have pre-market review, that that pre-market review should be under the aegis of FDA, but potentially

involve the collaboration of other bodies, that that pre-market review should be very streamlined and very simple for tests that meet the low scrutiny category and that in thinking about that streamlined thought that the issue of orphan tests needs to be taken into account, which, obviously, Steve's less than 4,000 a year does.

So, where we are in the discussion today is about what kind of test classification, other than orphan tests or something like that, triggers lower or higher scrutiny. What has come up in discussion with the committee before that isn't part of our discussion today is that there may well be other oversight mechanisms that are important or other activities, perhaps, that are important for use as genetic tests, things like provider education, et cetera, but the test classification scheme at this point is really just to figure out what is high scrutiny.

Is that helpful?

MR. HILLBACK: I think some of the things that we struggled with and I think we continue to do is with a test, in Genzyme’s case, like CF, where we have gone from four mutations to 72 or 80 something I think we are up to now, and we have changed that on a regular basis. Well, is the process we are putting in such that would require going back to FDA each time we added a new mutation or to take it even more simply, each time there is a new paper published, that draws better correlations between the mutations we already test for and the disease, are we, in fact, changing the test because we incorporate all the information by definition, at least some people's definition, as part of the test?

I think it is that iterative problem that we have had all along. We had it in the Task Force on Genetic Testing. We have had it

here in this group versus the sort of a first approval to get into market.

But all these tests are going to keep changing and that is where I don't think we have found a simple way to deal with that. Some of the proposals have been that you use some idea of a template to require the laboratory to at all times be able to say this is what we know and what we don't know and that in the normal procedures under review by CLIA they are going to be inspected for that.

If they can't do that, they are in big trouble. If that is sufficient, fine. If it is not there are people on both sides of that coin.

But I think those are some of the issues that we talked about. I don't think the first hurdle to get over to get into the market the first time is actually the most complex one. I think it is how do you manage the ongoing iterative process.

DR. BURKE: I think we have Francis, Pat, Muin and Judy.

DR. COLLINS: I want to go back to the point you made, Wylie, about off-label use and to what extent does this represent a potential harm that has to be factored into the analysis of whether or not a test is safe to administer. It might be safe for the group that you intend it for, but then what potential harms are going to fall upon those that you didn't intend it for if it is extensively used in an off-label way.

It would really be nice to have some data here to try to anticipate how big a problem this is going to be. We can think of a couple of circumstances that while they have not been FDA reviewed, have certainly been subjected to a consensus kind of recommendation. So APOE4 comes to mind, for instance, where there are consensus

recommendations saying this ought not to be used for prediction of risk of Alzheimer's in an asymptomatic person, but there is a legitimate case to be used when you are evaluating somebody with dementia and trying to further improve your differential diagnosis.

Is there data in that kind of circumstance to find out whether those recommendations are being largely ignored or whether there is evidence here that that kind of recommendation, once agreed to and propagated, influences practice to a sufficient degree that we might be a little reassured that, depending on labeling, it is going to be a way to go. Because if it is not a way to go, we are in deep trouble. Right?

If we insist that the test not be available outside of research protocols unless every possible use you could think of for that has met our standards of clinical validity and utility, then there is an awful lot of tests where they may have a very appropriate category of use, will not have all possible appropriate categories of use and will get held up.

I don't see much of a solution to that, except for the labeling issue. It would be nice to be able to quantify your question about how much risk are we taking by going down that road.

DR. BURKE: It may also be that the best protection - I think this is another element of data for off-label use - is very energetic efforts on the part not just of regulatory authorities but appropriate professional organizations to explain why certain uses are riskier.

DR. CHARACHE: I am just wondering about a clarification. I like the idea of putting maximum resources into tests that are going to hit larger populations, as opposed to making them all equal, but as

an example, since a lot of these are now home brew, your more than 4,000 tests per year, would that be by diagnosis or would that be by a given laboratory that does more than 4,000? Because if you require the APO-Es all to be done in the same laboratory, you are really down to a very small number that will be under review.

DR. GUTMAN: The way the HDE is written, it is talking about 4,000 total tests per year in the entire country. So, it would be 4,000 in the entire country. There is nothing sacrosanct about this. If you want to add predictive value and take out anything. It was just a way of looking at it.

DR. BURKE: But you made a point before that I think is very important and may in actual fact be the best definition we can get at, that simple definition for the orphan test, which is if the test isn't used more than x number of times a year, it falls outside or underneath or whatever the regulatory system.

DR. KHOURY: As I was looking over Steve's diagram, I am beginning to like it more and more. Actually, I would like to even simplify the process a little bit more. I mean, it is already simple.

First, a couple of comments on -- if we take out the comment at the beginning of this and say we are not going to deal with orphan diseases, sort of put it out of the loop completely, we have these three boxes that essentially characterize the three or four assays that SACGT have already identified, meaning predictive, diagnostic, that third box is stand alone, significant risk - that includes two of the concepts - treatment, no treatment, low penetrance versus high penetrance.

I think, Steve, you put them together from the same box.

That third box is really important -- the population base versus not, which tests the public health, newborn screening or screening all Ashkenazi Jews. So, I like that concept of these three boxes.

When you think about them and turn test by test, you will see that many of those you put down as "no," eventually will become "yes," like hemochromatosis, APO-E, APC, and cystic fibrosis are really yeses rather than noes because APC if it is given to all Ashkenazi Jews, whether they have a family history of colon cancer or not, that is population-based testing and newborn screening for CF or carrier testing. Hemochromatosis, same way and the APO-E in the same context.

So, what I would like to propose is a fairly simplistic method for the high scrutiny. Basically for the non-orphan disease -- I think we have to deal with them in a separate category. Any tests for intended use used for predictive purposes and had this box of whatever Steve calls stand-alone significant risk, which will include the social component. We will come back to it and it affects significant population-based whatever -- we have to define that. If we have yeses on all of these, then this goes to high scrutiny. No numbering, no waiting, nothing. You have to "yes" on all of these.

I bet a significant fraction of genetic tests might still fall in that category, but if you don't start there and if you take away all the orphan diseases and treat them in a separate group, I think we will be flooding the system with too much stuff.

So, these are the important boxes, prediction, significant risk, and a large segment of the population being tested for and you have to have "yes" to all of these to make it into that high scrutiny. I would use this as my starting point for the discussion.

DR. BURKE: Can you clarify why we have to say "yes" to

all? It seems to me that is a crucial point. I think we are simplifying the theme and I think the first box is -- I am just going to put x number of tests a year. So, x number of tests per year just to get at a working definition for orphan, so, if the answer is "no," we have got an orphan test of some kind. If the answer is "yes," your next question is -- I think these are the four points that you just captured, right? But then what we are going to do is say "yes" and get to high scrutiny.

I think it becomes crucial whether it is "yes" to all of them or whether it is "yes" to any of them. I think that is crucial.

DR. KHOURY: The reason it is "yes" to all of them is because first they tend to be highly correlated with one another and I would use this as a starting point for initial classification to see what fraction of tests fall there. I have a feeling that if you answer "yes" to all of them, you would still have a significant fraction. I could be wrong.

I mean, you can make that match the "yeses" and "noes." But I think, Steve, you had that same principle in mind when you flowed from yes to yes, you had in mind that what you really want to see at the FDA are all the yeses and try to get rid of the noes.

DR. GUTMAN: You have to realize that this was deliberately colored by my effort, correctly or incorrectly, to maintain a relatively small number that came out. At the end you just firebomb that as neither here nor there.

DR. KHOURY: If you say yes to all of them, you can be fairly developed to what you called potential significant risk because that factors especially on the low penetrance situation. There is a lot of psychosocial, potentially unmeasured outcomes that you can

immediately say "yes" to. Predictive is one of the significant factors. I think these things visually tend to go together, but I can't predict what numbers right now.

DR. WATSON: If it is stand alone, then it is probably 95 percent analytical concerns that you have. If there are no errors in testing, it is a stand alone test.

DR. COLLINS: Can you define "stand alone"?

DR. KHOURY: Well, Steve, you started this ball. So, tell us.

[Laughter.]

DR. GUTMAN: Well, a stand alone test is a test, which by itself creates an action. So, that would be stand alone. When I was first thinking about this, actually it was a positive test would generate further workup. Actually, stand alone, if it was a negative test, it would cause you to miss a diagnosis, that would actually be a stand alone. What I would call something adjunctive -- I was thinking of APO-E in the context of cardiovascular disease since you don't pick that as the single parameter to predict risk of cardiovascular disease. You look at family history, you look at lipid levels, maybe do a lipoprotein. It is a more complex diagnostic process.

Maybe those aren't good examples. It is easier to put them in the box than to actually take the products and you can see from the clumsy way that I distributed products, I can't even do my own boxes.

DR. KHOURY: Stands alone and significant risk should be kept together in the same box. So, if you answer "yes" to either one of them -- I mean, I like the way you did it, Steve, because you don't have to answer "yes" to all four to make it.

DR. BURKE: I actually want to capture some comments. I

have got Judy, Paul, Joann and Richard and anybody else want to -- okay, Francis.

So, Judy.

DR. LEWIS: I actually like the idea of a flow sheet a whole lot better than I do the idea of a checklist, especially when you look at how the numbers fall out. It seems to me that part of what is missing from the discussion from my perspective is the people that are affected by these. We are spending a lot of time talking about the laboratory issues and talking about the bureaucratic issues and the human issues and the people to whom this is affecting, I think, is something that we need to pay some attention to.

It seems to me -- I like the flow sheet idea, but I am not sure that I wouldn't want to see a test have to have all four of those categories before it became high risk because to me it would be the meaning of things to the individual who is the recipient of the information that we are creating and whether or not anyone alone would put into a high scrutiny and I appreciate your concern for the workload.

But, to me, the issue of what we are supposed to be doing is looking at what makes sense for the people and then we need to figure out how to get the resources to do the job, but that what we are supposed to be doing is getting a sense of what is it that is important for the population, the people that are out there.

So, I just want to pay attention to that and it seems to me that if something is predictive and there is not a whole lot of treatment that can be done, that those are the kind of tests that I want to see especially high scrutiny for.

DR. BURKE: Just to be clear, you could imagine high

scrutiny for something that was predictive, stand alone, with significant risk, but not necessarily being used for population testing. I think that is the description you just used.

DR. FERNHOFF: I like this scheme also in that it seems to at this first step, anything diagnostic sort of falls out immediately. And also, the point just made about for the individual, like BRCA testing, it is not being done on population.

Just back to one other point here about off-label use. I may be wrong on this. I thought it used to be with maternal serum AFP – and I don't know whether that is still has ever gotten approved for -- you are saying a "yes" and a "no"? For neural tube, but not for Down's syndrome, probably one of the most widely used genetic tests in the country and it has been off labeled for 15 years now, whatever it is, for maternal serum for AFP, part of the Down's package.

DR. BOUGHMAN: I am trying to envision what goes into this process and it is not a test. It is a packet of information that goes into the process. In fact, it seems to me that if we go back to our overall goal here, which is to protect the patients from harm, whatever that harm might be; yet, balance that with the issue of access and everybody that people doing the reviewing, the people having their packets reviewed.

Everybody wants a streamlined process, but if we go back to the idea of the packet going in, it seems to me that we can address these issues not whether at each one of these should it be high scrutiny or not, but, in fact, think of this as the packet of information is putting the burden of proof, whatever that level of proof, that they have met the criteria with the information that they have provided.

When you say if it is a predictive test, then it needs high scrutiny. That doesn't tell me what is going to happen at that high scrutiny level. If it is a predictive test, whatever review process that is, needs to make sure that the use and the way that the test is used and the information coming out is adequate to explain that it is a predictive test, then the issue of is it high scrutiny or not is not really our issue.

I mean, some of these things could be high scrutiny, yes, but if, in fact, a perfectly put together package, whether it is a rapid process by internal members of FDA staff trained to do so or an external committee this size that discusses it for two days. It is not going to change the data. It is simply going to meet the criteria that are set at the beginning of the process anyway.

DR. KING: I just want to make the comment that I understand Muin's interest in trying to simplify the scheme, but from a logical, thoughtful, procedural process, I like the process that Steve put out, the multiple boxes that are followed sequentially. There may be some tweaking of them, but I think to me it makes more sense than to put all those into one box, in the middle box, and then say yes to all of them, to go on and be a high scrutiny test. I think that there has to be a reevaluation and we have talked about the fact that the tests will be reevaluated or rethought at various times, but something that will be a low scrutiny at one point, will be a high scrutiny if it is used in a different way and that would come out by following and reevaluating it at the appropriate time.

It just is easier for me to understand in the sequential boxes rather than one big single box.

DR. KHOURY: There is no real difference between what Steve

put forth and what I really put forth. I think what is really throwing people off is the word "routine" that comes out at the end of the “no” to the end of any one of them. And coming back to what Joann was speaking about earlier, what we are talking about here is the process of initial scrutiny that would lead to some perhaps labeling requirement before release of tests, counseling requirements and data collection requirements that basically fall into the higher tier of those that required an initial scrutiny could be released for use with the provisions that this is what we know and what we don't know and continue with that.

So, the fact that the word "routine" is mentioned there is probably a bit of a misnomer because data will have to be kept collected because you need data anyway, but the question is the initial regulatory roadblock or oversight that need to be imposed on the system and we can debate how many of "yeses" do we need or not, but I think we have the elements of the picture here, that an initial review process that is as simple as possible, as transparent as possible, will put a category of tests on the market with the provisions that it would be adequately labeled. There will be information flowing in a standardized way to the consumers and data will be collected in a standardized way in the future.

I am not wedded to the idea of you have to have "yes" to all of them versus "yes" to three or four of them, but I think we have the basic idea in mind and these are primarily the predictive tests that are going to be used for situations where there is no other tests that can either confirm or deny the diagnosis for which treatments may or may not be available or still are being debated, for which penetrance function is not known or is still being debated-- we are

almost there. If we can make the final push this afternoon, I think we will get there by the end.

DR. COLLINS: I think I agree with Richard that it is helpful to have maybe this diagram laid out a little more explicitly in terms of what the decision points are, although it is essentially pretty much equivalent to what is up there, but it is useful to consider each of those decisions separately.

I wanted to raise one other issue and that relates to population-based as a particularly important determinant of how much scrutiny one might expect to see. For an existing population-based test, like newborn screening for PKU, this diagram seems like it gets to where you want to be, but I would have a hard time imaging a brand new test that is being proposed for population-based screening that you would not want to look at pretty carefully.

In that regard, this diagram doesn't quite get you there so that if you had a population-based test, which seemed to pass your stand alone and significant risk category, then would go into the relatively lower scrutiny. I would be a bit uneasy about that particular harm.

DR. GUTMAN: You mean, if it weren't stand alone significant risk. The way it is written is that if it is population-based and it is either stand alone or significant, it does go into the highest category.

DR. COLLINS: But suppose it passes your stand alone significant risk test as being okay?

DR. GUTMAN: Not being significant. But you have another potential, which is you could say part of it isn't how you define stand alone significant risk or you could say that those are being cast off

of that last box, in fact, deserve high scrutiny or moderate scrutiny.

Or you can change this. I mean, you can get rid of the last box if you want. I don't care. I just put it up.

DR. TUCKSON: I am a little confused. Maybe I am not listening carefully. I thought Muin, when he celebrated the box because any one element in the box led you to ties. All, not just one.

DR. BURKE: But I think that is a very helpful comment, Reed, because it seems what I am hearing is we are really warm now and it is just a matter of figuring out if it is all of them or if it is -- in other words, I think it is figuring out what is happening in that middle box arena.

MS. UHLMANN: In terms of high scrutiny when we were at the beginning of the afternoon, where everything was kind of falling into high scrutiny, I think a lot of it is because genetic testing is a moving target. I mean, just look at what happened to cystic fibrosis when the gene was discovered in 1989. You have got a mutation, 75 percent. You thought okay. It will just be a couple of months. We will find a second one we'll be there.

Where are we ten years later? We have what, over 800 mutations that cause cystic fibrosis. So, I think within the midst of all this there has got to be a way to look at tests and realize that they are going to switch categories. That something that may start out as a low complexity test may turn out to be high scrutiny and vice versa. Francis also made the point that I want to make, that I think in general with population screening that when you introduce a test into a population, I think initially that is going to be a high scrutiny test, but I think it can also depending on the type of test can move quickly into a low scrutiny or low oversight, examples being

Tay-Sach's disease and sickle cell anemia, et cetera, which are done in routine doctors' offices. Although those are population-based tests, those are not people that we need to see for high complexity counseling.

DR. BURKE: But for the purpose of regulation is what happens when a test is first made available that is crucial, isn't it, at this point? I mean, that is our concern at this point.

DR. VOELKERDING: A couple of questions and one is - I guess I would ask Steve - we have been talking about the FDA having oversight of all new genetic tests. One question is when he says what he would really like to focus on are the ones that give most concern. So, is it feasible to think of things like pharmacogenetics and new mutations that are valuable for diagnostic purposes that many home brew laboratories would be wanting to employ? Would those even really need to come forward to the FDA? That is question 1, because they would potentially fall into the term "routine," however you want to use that term.

So, that is my first question that I would like to get some clarification on.

DR. GUTMAN: My guess is that if they are home brew tests, they wouldn't come into the FDA. We actually anticipate seeing some DNA chips at some point that might do some of these tests and if they are sold as commercial systems, we are frankly hoping they do come into FDA in terms of marketing chips.

DR. BURKE: Just for clarification, I think we have already determined that all genetic tests will come through some form of FDA review. Now, in that determination, I think that is already a discussion that has been had by the SACGT, but in that discussion, we

allowed for two possibilities that I think address the issues you are raising.

One of them is the so-called orphan test. Steve has just given us a way to approach an operational definition of that, which clearly would cover, I think, a lot of early home brewed tests, that if they would be used less than x number of times a year, they wouldn't yet be ready for FDA review and the other is the idea that many tests, presumably a majority of tests, although they would come for pre-market review, would fall into the routine category.

That is, obviously, the classification issue we are dealing with now. And that routine category, although it would be an FDA review, would be a streamlined process, including standards set by professional organizations who participate with the FDA in that process, including the collaborative process. I just want to say that is where our conversation has taken us to this point.

DR. GUTMAN: That is a reasonable correction. I am making the assumption that you have all bought into my argument that diagnostic tests would be at the low end of review. That is my bias coming out of it.

DR. BURKE: And certainly there seems to be an evolving consensus on that point.

DR. COLLINS: Just to quickly point out, though, that some pharmacogenetic tests will both be useful for predicting drug responsiveness and will also have a predictive property about natural history of disease and risk of future illness.

You can't always cleanly separate these two out. They are often sampling a pathway that has consequences for both.

DR. CHARACHE: I am on the same trail, but I would like to

suggest a change in that word from "routine" to "standard scrutiny," as opposed to "high scrutiny. They are all getting scrutiny. Then we have to define what standard scrutiny will encompass, such as Can you get all of your information from a simple kindred or how many kindreds do you need for a low prevalence test versus a high prevalence test.

MS. BOLDT: I really like the idea of this box and what I see is wanting to break out a little bit of a significant risk and actually having treatment versus not treatment. I do like the idea of having a scoring system and as simple as a "yes," equals one and a "no" equals zero, depending on what it is.

The reason I see that as very helpful is that it is very objective then because some of this, as Judy was saying, is so subjective if someone thinks it is helpful or not helpful. So, I think that makes it clear whenever someone is trying to classify where the test fits in. I think we should not have to have them all be "yes." I think that you have a combination of them that would be a yes to lead to a high scrutiny.

DR. LEWIS: I think one of the other issues I want us to focus on and maybe this will come out later when we need to talk about some of the social issues is that tests that are more specific for certain subgroups of the population than others and that would have the potential of dealing with a particular segment of the population -- I mean, when I heard people talk about Tay Sachs as being straight forward, when you start to look at the number of genetic conditions that are being tested for the Ashkenazi Jewish population, that raises some significance. In sickle cells, when we start talking about the African American population, I think that we have to be sensitive to the fact that looking at prevalence in subpopulations, as opposed to

population-based testing is another thing I want to make sure that we put in the higher scrutiny category.

DR. KHOURY: I want to emphasize the point about population testing, what Francis said earlier. I think it is an important box that by itself needs higher scrutiny, regardless of what else is going on. I mean, if you are going to give out the test to all newborns in this country, somebody has to look at this.

Now, if you remember, what we are trying to do here, we are trying to look at the initial FDA or oversight of pre-market approval. There will be all kinds of opportunities for discussions by different groups, consensus conferences, public health agencies, like CDC and others and NIH to evaluate probably in the post-market testing whether something is useful for population testing, even in general like newborn screening or subsets of the population, like Ashkenazi Jews, who live in a certain geographic area.

So, I would even put that in a separate box by itself and I think that the three axes we are dealing with for initial tests, - the predictive versus diagnostic, the treatment/no treatment - this idea of risks and penetrance- I forgot what it was. These together and then the population is kind of a separate box that would require a scrutiny different from the regulatory scrutiny, I would think, because these things happen because of consensus development processes, rather than an initial FDA process.

MR. HILLBACK: I guess, going to Judy's point, I think later on we have time scheduled to talk about global social issues because I don't think this framework lends itself to trying to overlay social issues. I think you get into real slippery slope and get a lot of confusion. I would like to keep this to what we have considered

science-based, the pre-market approval concept being science-based rather than to try to incorporate all the other things.

I think we really get in trouble when we try to do both at once.

DR. BURKE: Thanks. I actually thought Judy's was a nice segueway into where we need to go next. I am going to ask Barbara Koenig to make some comments in a couple of minutes.

Let me just say, what I am hearing and I think what we will try and do a little bit further on in the afternoon is redraw the boxes and see if we are getting closer to what we want.

But I am hearing a possibility that what we might be talking about is still sticking with the first question being x number of tests per year, with the question perhaps back to FDA about how you find that out, how you determine that, but some measure to determine whether tests are being used very infrequently and should fall under an orphan test category.

Then once you are above that, you go to a central area where a number of questions are asked. One of them it seems to be a cluster that includes whether the test is predictive, whether there is a treatment, whether there are significant risks associated with it, another separate box is whether or not the test is planned for use as a population screening device and maybe there is another box there following how our discussion comes in the next hour for certain kinds of social risks.

In other words, maybe significant risks need to be broken out in terms of potential exposure to medical interventions and also the potential for social risks. We may be able to figure out whether that box should be there after the discussion that we are about to

have.

But it seems to me the other question that we have -- we have those middle boxes that are helping us to determine whether the test is of concern. We then, I think, have to consider whether there is an intermediate step before you go to high scrutiny or maybe a breaking down of what high scrutiny is with Step 1 being disclosure.

That is, to what extent do the concerns that are triggered by that middle set of boxes get solved or get addressed adequately by the right kind of pre-test and post-test information and, once you have seen those sort of disclosures, to what extent do you have any additional concerns.

I just want to throw that out and what I would like to do now is move to what is a social issue.

DR. TUCKSON: Also, please, in your summary, let's also not fail to attend to the language. This word "high" versus "low," "routine versus non," I mean, I think those are very important words to the individual person, who happens to be in a so-called low area test. My God, would this mean that I have reason to be concerned if people aren't paying attention.

So, even routine is important and we have got to find the language. I don't think we need to wordsmith it now, but given our charge of the SACGT, we don't want to undermine people's confidence because of the words we choose.

DR. BURKE: So, we have already had "standards" be proposed in place of "routine." We captured that. What you are proposing is high scrutiny maybe should have a different name.

DR. TUCKSON: Yes.

DR. BURKE: We can think about that. Barbara, do you want

to start us on social issues?

DR. KOENIG: I will do my best, although I am going to start with a warning that I have not solved this. Wylie and I went back and forth about whether I would do a formal presentation and I am so glad Joann has taken my handwritten notes and typed them on my laptop. So, I will try and talk through that.

To give just two seconds of background to the non-SACGT members, we have spent a lot of time talking about social and ethical issues and I am a social scientist. So, I think that is why I got assigned this. And we came up with a sort of foundational premise, which is even tests that meet all the kinds of criteria we have been talking about for the rest of the day, meaning that they have analytic validity, clinical validity and possibly clinical utility, there may be situations where there are social consequences of the test, which would cause them to bump up into a high level of scrutiny where we might want to think about them before they are widely available on the market.

Why did we come to that conclusion? Well, because the one way in which genetic tests are different than others is these historical issues that in the past the main social consequence of genetic tests, for example, eugenics or things like situations where people were tested for sickle cell trait that was used as a way of denying opportunity, those kinds of things.

Some of the main consequences of genetic testing in the past were in the social arena, as opposed to strictly in the biomedical arena. But that doesn't make this any easier, unfortunately.

Also, there were many public comments on this. There is a lot of public concern about this. So, that is one of the reasons that genetics is somewhat different. But then in terms of thinking about

how to put social and ethical criteria into a categorization scheme like this, I just was talking to Kathy before the meeting and other people -- there are very few precedents in government oversight for trying to do this.

So, we are trying to do a very, very difficult thing. It is possible, OTA, for example, when they used to evaluate new technologies used to, in their evaluation, include social and ethical impact, but there was never any attempt to sort of have that be part of oversight, in terms of whether to allow something to move forward.

So, we are really doing something that I hadn't realized quite how hard it was going to be when we started talking about this. So, I think that there is the danger -- well, environmental impact reports, as Kathy pointed out, may be one area where there is a parallel in another arm of the government that we could think about.

But I think we really need to worry about the fact that there is also a danger of doing harm whenever you add something like this, as well as in addressing some of these harms. So, in my own thinking about this, in terms of what the Committee has done, we have spent a lot of time worrying about things like making a distinction between non-medical and medical genetic tests and whether that is an important thing for us to talk about.

We are going to probably talk about that more tomorrow. I think it is important to keep in mind that that is a very hard issue to -- it is not a clear distinction. It never will be a clear distinction. The tests are going to move back and forth between being considered medical, being considered non-medical. It is not a clear distinction.

Then we also need to think about things like what about the

idea of having your full sequence on a chip with all of your SNP variability that you literally carry around with you. I am trying to think forward to the future to whether that is a technical possibility or even a great many points of variability that you would carry around. Is that then considered a medical test? Is that a non-medical test? Is it just sort of a reflection of your own genetics?

The other point that we have to keep in mind is the issue of when we are starting to think about social impact or social consequences of genetic tests is to really -- what is social? Are social impacts just the sum of all the consequences that happen to the individuals in a society?

For example, you just worry about this in terms of the potential discrimination against individuals and you add that up? Or are there other kinds of consequences for society as a whole that we want to be thinking about? For example, if we allow tests of -- and one of the ones that keeps coming up -- ancestry, if we allow marketing the tests for particular kinds of ancestry, might that have the consequence of increasing racism, those kinds of things.

Can we have that on the table or do we need to stick with a more individual summing up of, like this person was discriminated against in Iowa. Those are the kinds of social consequences that we are thinking about.

The other thing we really need to consider is if we are going to implement social consequences in our scheme, who should do this? How should they do it? At what point in the review should it be done? Should it be done before these other things? Should it be done in tandem with these other kinds of things? Should it be an iterative process? That is going to be a very, very complicated thing.

Another thing to think about in terms of categorization, are there certain categories of kinds of tests that we would want to put on the table as automatically putting something into a high social scrutiny. The ones that we have talked about and that people mention are things like diseases that predict mental illness or, excuse me, tests that might predict mental illness, predict other forms of behavior and, in particular, the new tests that might predict addiction or the possibility of addiction to particular kinds of substances.

Then I just went through a whole bunch of other things. What about tests to predict perfect pitch or sexual orientation, all these. There are all kinds of tests like that that are extremely complicated that some people might argue have health consequences, some people wouldn't.

Then the other category that we might want to think of is tests that target particular populations. We actually wrote that into our final document. Whenever you are going to target a particular identifiable population, given the history of genetics, that we need to be careful about that.

Then, I think, finally, so that we can have time for discussion, the other thing that I think we might want to consider as we think about data collection as well, we haven't really raised this, but we could try and put into our oversight scheme, the need to collect data also on social consequences. That is not there. We hadn't really moved to that level, but that is another possibility that we could add that as we are thinking about sort of the category that before would have ended up being B2. We might want a lesson on market, but then really look at the consequences to some extent.

So, I hope that was helpful.

DR. BURKE: Comments. Open for discussion.

DR. BOUGHMAN: We may be able to gain a different kind of insight than at least I certainly have thought about up to this point.

From Francis's table this morning, we are talking about the turning point where, or at least one of the critical differences here is IRB involved, IRB not involved, at least the processes that we are talking about now. We may, in fact, want to ask ourselves what is it about the IRB review and the informed consent process associated with that, that lets us feel comfortable or at least in the year 2000 that somehow the patients or the individuals are being protected in the process. What is it we are letting go of that we need to recapture in these genetic tests that we still have this feeling?

The other dimension that I think we need to just remain cognizant of is that, in fact, at least parts of society, some states and so on are making progress globally on addressing some of these issues. We have gotten the executive order. We do not have a national law regarding the discrimination, but we do have states that are moving in this direction.

Do we feel differently sitting around this table in August of the year 2000 than we did when we had that discussion in June of 1999? Are we moving in the right way to, in fact, be addressing some of these things at the very global level or how would what we are proposing change if a national law with regard to genetic discrimination were passed, would that change our process dramatically?

DR. VOELKERDING: I just wanted to relate something that I found personally very vexing and troublesome and that addresses the issue of social aspects of genetic testing. Probably many of you are aware over the last few years, there has been a lot of new genetic

information on the genetics of hearing and there is a specific mutation gene called the connexin 26 gene associated with non-syndromic congenital hearing loss.

I heard an anecdotal story, which really struck me very hard, which was there was a family that had a child with congenital hearing loss and the family and child were tested and the child was found to be homozygous for the major mutation in the connexin 26 gene. The parents became pregnant for another pregnancy and requested connexin 26 gene testing on the fetus.

It was positive and they elected to abort the child. So, I think that that was a very interesting, challenging issue for me as I struggled trying to ask the question when you ask issues of social testing or social issues -- they all seem related issues -- where individuals are starting to -- in this sense, it just seemed a very interesting and complexing and challenging question.

So, I think it gets at issues of intended use of specific genetic tests and I am not certain that we have really -- it is hard for me to take something like that, having recently just had our second child, and ask myself what decision process would I have gone through as a potential parent and how that could be incorporated into some of the other aspects of the discussion we have had today about intended uses of tests and oversight.

So, I throw that out as I think, a very complex and challenging question as we define more and more potential risks and subtler phenotypes, if you will, for which we have potentially medical interventions and medical therapies available.

MR. HILLBACK: I guess when I look at this and I think we have paid a lot of attention to a lot of these issues, but it seems to

me the fundamental thing we are trying to get at here today is how do we decide when a test should be utilized? I think that is a separate issue from how we utilize it, from the cautions that we include, from the training to the average physician out there that has to use this, from how do we incorporate these social issues into how the genetics community operates and operates in the context of a larger medical community.

I think those are very important issues, but I don't think they fit with an approval process of the test, whether that is by FDA or whether that is CLIA or whether that is home brew or anything else. I would love to hear -- I am sure there are comments on the other side, but I don't want to downplay the issues because they are very important issues, but I don't quite understand how they fit into an approval process.

DR. BURKE: I have Richard, Judy, Reed. Anybody else?

DR. KING: I will disagree with Elliott on that. I think there are very complex issues and how you view them and how other families view them and how other individuals view them and other groups view them can be very different. I don't see that that is what we are trying to address in this particular committee today other than say that somewhere in the information that there are various views about that and the appropriate genetic counseling or whatever is indicated.

DR. BURKE: In fact, I have a question to you on your comment. I wonder if those kinds of social issues up the ante for the kind of information that needs to be provided with the test.

DR. KING: Absolutely.

DR. LEWIS: That was going to be my point was that I think we need to pay specific attention to tests that we tend to stigmatize

particular populations and that that becomes the social issue. At the level of the individual, I think what we are doing is giving individuals information and then whether we agree with the choice they make with the information or not is not necessarily the relevant piece. We are giving people information. We want that information to be valid, accurate and not stigmatizing, but what people choose to do at the level of the individual, whether or not you would have chosen to continue that particular pregnancy or my neighbor across the street, who just had a hydrocephalic spina bifida kid and had prenatal diagnostic testing and chose to continue the pregnancy. I mean, that was their decision. It may or may not be one that any particular person agrees with, but the question becomes what people do with information is an individual decision. I think what we are wanting to do -- at least my sense is what we are wanting to do is pay attention to the broader social issues so that what we are not doing is something that would cause harm to large groups of people that was unintended consequences of the genetic information.

DR. TUCKSON: I think I was encouraged by the large comment that Richard made and, Wylie, your highlighting that. I think that the place in which at least if they do come together is at the point of the requirements and the character of information that is provided and the notion of the requirement for counseling.

So, the information in the package, whatever that level of instruction or guidance is and then, of course that is the place they come together. I would appreciate the general theme, Richard and Elliott, that this is probably not the table to hash out a whole lot more. The only other consideration that I have, though, is that those of us who have to meet tomorrow -- I hope it doesn't apply, but I just

don't want to get caught in the position of the folks who sat around scientific tables like this around the nuclear bomb.

At some point you say, well, let's just deal with the chemistry of it, folks. I think the kind of information we are talking about because of Francis and his crew having, is going to be so powerful about whether or not the person will continue to be or allowed to be, whether the race will be allowed to be or not to be categories of rates and so forth.

I mean, that is pretty important stuff and you can't stay moot on that. However, I would have prescribed that for this task today, perhaps the only linkage is around the notion of information counseling and the quality and character of that information.

DR. WATSON: That was essentially what I would have said and I think she glossed over pretty quickly in the concept of certainly an environmental impact statement. I think there are social risks and although Francis has done a lot to get us where we are, unfortunately, our government and our Congress has done very little to get us where we are, as far as dealing with those sorts of issues. So, I think they are important to be raised because they do inform patients about where we have gaps in our system that can impact them and it will drive the appropriate people to build the changes that are needed.

DR. KOENIG: I just want to interject two things really quickly. One is that just saying that one can simply turn these things over with individual choices is in and of itself a particular political strategy.

I just want to make sure that that idea is on the table. That is basically what has happened in genetics. That is what happened as the genetics profession tried to transform itself from a group,

which had been part of the eugenics movement up until the fifties, and then transformed itself into their modern medical genetics. I mean, that was done by moving these issues from the realm of social choices into individual choices and to basically, I think, engaging in a fiction that these can be purely individual choices, which is not to say that I am opposed to individual choice because I think these things do vary enormously, especially in the reproductive area.

But we are now not just talking about reproductive issues. So, I think we don't want to let our fear of the reproductive domain affect the way we discuss these genetics in other areas. In terms of a response to Richard and maybe Elliott, I think the one exception to why this is important in oversight up front might be the issue of targeted tests to particular populations and that is going to be an area where there are going to be things written in whatever is the equivalent of the package insert, what kind of language is used, who is targeted, what that means, those kinds of things perhaps are things that could be considered up front and might be--.

That is not a matter of individual choice because once it is out there and it is marketed to a particular group, you can't control that as an individual period, bottom line.

DR. CHARACHE: I am going to continue on along that line of thinking, but I would like to emphasize that we cannot overemphasize the level of ignorance of the medical community in terms of how to use a given genetic test. One of the challenges with the FDA procedure is that the information of how the test should be used and interpreted is not in the hands of the clinician who uses the test. It is in the hands of the laboratorian.

This is why I have emphasized the role of the laboratorian

in communicating this information and ensuring that it gets across. But I think in order to be sure that that information is disseminated to the community that is ordering the test or the public that is demanding the test, we bring ourselves to the issue of informed consent of some type of interaction, which ensures that that information has been in the hands of the person who has ordered the test and the person who has having the test performed on them.

I think that we have to make sure that when we emphasize what information has to be obtained and available, that it is, in fact, communicated to those who need it in a way that can be certain that it gets there.

DR. BURKE: Wendy. We have got a long list of folks and I have got everybody's name down, I think.

MS. UHLMANN: I just wanted to point out that I think what Barbara was raising in terms of social issues is actually a subcategory of predictive tests, which are going to be tests that are going to predict different traits, be it obesity, be it homosexuality, be it perfect pitch, et cetera. I think that that is maybe a categorization to think of it as being a special category under predictive tests, which will deserve high scrutiny.

DR. CAGGANA: That is exactly what I was going to say. I mean, I see this as going towards this multifactorial test, which are sprinkled with all sorts of social issues. In the context of the boxes up there, assuming that there is counseling and assuming that we do the right thing, when you start talking about a series of genes that are going to give us the result, that will give us a prediction of whether or not that individual will have any one of the multifactorial traits, do we put that through as a panel or do we -- I mean, we already sort

of see it with thrombotic risk factors. People order a panel of three tests. People order groups of tests to rule out. So, how does that fit in that scheme?

DR. FERNHOFF: It seems that regardless of where a test winds up, either in standard or whatever we are going to call it, is there or has the group considered mechanisms for reporting adverse effects, such as in the drug adverse effect, adverse social effect of a test. Some of these may be real, some may not be real, but some mechanism where either individual or groups this has been applied wrongly or misapplied, that whatever system eventually comes down to some way of easily reporting that and can be reviewed by an oversight group and how do we correct this in the next round of labeling.

DR. BURKE: And that certainly speaks to the issue of how do we define harms that can derive from tests.

DR. VOELKERDING: It was my hope that by raising this issue of connexin that it would help sort of get some feedback and clarification, but it still seems to me that it does drive at the issue of what is the intended use of a specific test. I was wondering in one setting, as we said earlier -- and the connexin testing could be used diagnostically, in that case it would have been potentially standard review; whereas in this case it was used for a very different purpose, which I think the consensus of the room would have probably pushed it into a very high level of oversight because it was being used in the setting of prenatal diagnosis.

So, I guess I am understanding that in part individuals that were submitting tests for oversight or for approval would have to specifically list their intended uses and that they would get approval for specific intended uses of the test and that if they are getting

pressured in their home institution to use it for other purposes, it would have to be rethought. What would that be? And they would have to come back into essentially the oversight mechanism, as I understand it.

The issue of reporting adverse effects, just as an analogy, as someone who spent ten years doing transfusion medicine, the adverse effects of blood products, for example, it is a voluntary mechanism. When there is an adverse event of a transfusion, for example, it is a voluntary mechanism of reporting to the FDA. Essentially I can initial a report within the first 72 hours.

Of course, we are talking perhaps about a much longer time line, but that responsibility largely resides in the laboratory in the setting of the blood banks. So, there is kind of an interesting analogy in that regard.

MS. DAVIDSON: I wanted to speak -- actually, this goes way back -- I am sorry to be out of sequence here -- to Joann's remarks about genetic determination because I just wanted to say, from my perspective, while there has been movement over this past year, we are really just a lot of times circling around and beating on the chest in terms of what is happening in Congress now.

With respect to what is happening in the states, I think that that eventually will contribute to some federal action. But, again, I don't think it is anything that really takes away from how we should look at genetic discrimination on the risk of discrimination in terms of oversight issues.

But that being said, it is better this year than it was last year.

The other thing is that I just want to join several people

who struggle to look at this issue and the quality testing issue because I can say that, particularly from where I sit in dealing with the public and dealing with people that it is almost literally, viscerally impossible to break this down into a quality testing lab issue versus kind of a quality experience or non-quality experience for the patient and family.

While I know that our remedy right now is to look and to make these decisions about categorization of tests, and how we apply ways of applying scrutiny and oversight. It is critical to always be thinking about informed consent, counseling, and information. Karl, I was particularly thinking about the particular situation that you talked about. For me, it is literally impossible to think about all the delivery of the service or of the tests without taking it all the way to someone sitting either in my office or at the end of our help line and whether they really have had quality service, not only with respect to tests and the adequacy of tests, but all the pieces of information and counseling and informed consent that go with it.

DR. BURKE: But it is larger context we are working on.

DR. KHOURY: I wanted to come back to a discussion about how to use the social issues in the initial classification with all the points that have been raised and I agree with all the discussion here and it seems to me that we need to find the initial hook under the significant risk discussion that would include both medical and non-medical and using some criteria that would make the shift go this way versus standard or routine or whatever type of scrutiny. For that, it has to be made operational to the point that is this testing for behavioral genetic trait or something very complete because all genetic information has potential social impact, be it SNPs or otherwise.

It helps classify people and populations, et cetera. For that to be useful and pragmatic, it has to be cut down to the level of a few items that could be used along the spectrum on -- when we are asking initial questions -- is this a predictive test, is this a stand-alone test where there is no treatment or is the treatment bad enough, are there significant social sequelae, and is this going to be used for population testing? Yes or no?

Upon further reflection, I think, the way I fall under this, if the answer is "yes" to any one of these questions, rather than all of them as I said earlier, it goes automatically to the high box or the high scrutiny box because I think many of them are correlated and I don't see that would create more work for you, but I think it is very important that we do it and we do it correctly. I think each one of these factors alone might deserve pushing it into the highest group box. But in order for this to be useful, it has to be systematically thought through with some specific definitions on what we mean by predictive, what we mean by stand alone, what we mean by social risks and I don't know if we have time to do it today, but the basic principles can be laid down.

MR. HILLBACK: I was going mostly where I thought Muin was going to with his last comments about pushing everything to high, but I find it very, very difficult to figure out how you define, in a practical, operational way, social concern.

I mean, if you say, well, every time there is a bad outcome, an individual is going to feel bad, then everything falls in that category. There is a social risk with every test. I don't know what you would do. I don't know how you would define it.

So, I guess my question back to anyone was how do you make

that a real criteria, other than it defines someone who is different from someone else, since by definition, we have all agreed we all different from everyone else, at least in one form or another of our genome.

I just can't find a practical way of doing it. I think the other point, back to having this follow-up on social issues, the labs can't even follow-up on the clinical outcomes because we don't interface with the patients. To think about trying to follow-up on social outcomes, I think it is, again, very difficult. I think it goes back to the rest of the medical population, not the laboratory population.

DR. BURKE: I would like to make a comment on how you define a social issue. I agree it is hard. I think it is hard to define medical risk sometimes, too. Yet, at the same time, I don't think anybody has any difficulty seeing that a predictive test for mental illness is different than a predictive test for cardiovascular disease.

Now, it might be that you get to a point and a threshold where it is very hard to define. But I actually think that some of the examples Barbara gave us, it would be easy at least to pick out those that generate the highest concern, which is probably what we are the most concerned about.

I would like to say we have about five minutes until our break. I have been trying as this discussion went forward to try and figure out how that diagram gets changed. So, when we come back from our break, I will have another transparency to show you.

Anybody can show theirs, too. So, we have time now for about two or three comments before we break.

DR. NUSSBAUM: This is pretty crude, but one approach in

thinking about what could have a social implication would be whether the result of this testing would be to have an impact on individuals outside the family of the person being tested, so it would have implications for people outside of the family. You might make the argument maybe beyond the individual being tested, but to me that is a definition of a social impact.

If, as a result of this testing, it has implications for people who didn't consent to have the testing, are not even related by family to the person having testing and yet will still be impacted by the results of testing.

DR. BURKE: Persons outside of the biologic family. Okay.

DR. KHOURY: The other issue about orphan tests versus multiplex will be a point we discuss later?

DR. BURKE: Maybe we should talk a moment about orphan test and ask whether the operational definition that Steve gave us works; that is, it is a matter of figuring out how many tests a year it is. But is that a reasonable way to approach the orphan test definition?

DR. BOUGHMAN: My suggestion is that, in fact, because this is a work in progress, we do accept that on two bases. First of all, it relates from the frequency issues that we address regularly in genetics and also because it is one of the mechanisms or one of the points in current mechanisms that we can come back and modify later, but I think we have other agenda items that we should focus on.

DR. BURKE: So, that is a reasonable working definition. I would just say from the lab and professional organizations' point of view, if we are going to take that approach, that may be a place where the organization weigh in about what number of tests per year it should be and how that gets defined.

DR. KOENIG: My only concern about doing it with a number is that part of the issue is the potential market, as well. So even if there is something that is quite rare, you could imagine a situation where the people were very much afraid of it and often it isn't rare, but something like that, you might imagine that even though the legitimate use might be less than 4,000, it could be why don't you get this because everyone should get it.

DR. BURKE: Wouldn't that then trigger –?

DR. KOENIG: It becomes a situation of one set number gets the prize?

DR. BURKE: Steve, if you had a test that was below 4,000 and then it went above 4,000, wouldn't that trigger?

DR. GUTMAN: That would trigger.

DR. BURKE: So that triggers the review then?

DR. KOENIG: But it may trigger it too late because it wouldn’t have had any initial review. Is this a hypothetical notion of we imagine this will have a use of 4,000, once it is on the market?

DR. BURKE: It makes sense to me that you might have an early stage where tests aren't being used very often and there is accumulating information about them. It sounds like we may have to talk about this more after the break.

More comment on that in the next couple minutes?

DR. NOLL: I just want to ask another question, if I may, before we break.

We talked about information beyond the family, but I haven't heard any discussion, and you have probably done this in another, what about the information that does pertain to the family? Who owns that information? And what are the obligations of probably

not the laboratory but the physician who gets the information to do with that information with respect to the other members of the family?

Have you had those discussions? I think it is an absolutely critical thing about genetic tests and I never hear it discussed. People always assume that that information belongs only to that individual in the United States at least.

DR. BURKE: We actually have discussed in the past that tests that generate additional information relevant to family members are part of what makes a test potentially deserving of higher scrutiny. I don't think we have had discussions about the issue of duty to disclose, which is clearly a very live issue in genetics.

DR. COLLINS: I think we may have touched on it in the past, but it did not seem like it was quite directly in our mandate. There are other groups who have looked at this very explicitly and have published statements on the topic and there is case law out there in terms of whether physicians are responsible. That doesn't completely agree with itself, but it is out there.

So, I think we had sort of concluded that this was maybe a little far away from our central mission.

DR. BURKE: It does, although I think your point raises the question whether the potential of the test to produce implications for other family members is one of the things that goes into one of those boxes somewhere. We haven't come back to that issue in this discussion.

DR. KING: It also goes on their information sheet.

DR. BURKE: Yes. It definitely needs to be part of the information sheet.

MS. BOLDT: Another issue we can take up further is what is

the duty to recontact. As we get new information, what is the responsibility to contact the families again?

DR. BURKE: I think we are ready for a break, 15 minutes, back at 2:45.

[Brief recess.]

DR. BURKE: Okay. Are we ready to get started again?

I have made a couple of transparencies that I think reflect where the discussion has been going and points to some of the questions that we need to resolve. But I wanted to start with -- I am going to show a schema, with help from Francis, that I think clarifies our discussion. Before I do that, I want to talk about two levels of scrutiny using the neutral term that Pat proposed. So, we have got a Scrutiny Level 1 and a Scrutiny Level 2. Scrutiny Level 1 was what we called routine before and Scrutiny Level 2 was what we called high scrutiny before.

We can talk about which terminology is better, but just to be very concrete about what I think we are saying. My effort here is just to capture what I think the discussion has been -- that Scrutiny Level 1 is, first of all, a streamlined process. It is a process designed to be relatively quick and relatively easy for a test offerer to go through.

It is a process that is based on preset standards. Basically a test offerer knows ahead of time that there are certain kinds of standards that the FDA expects a test to meet in an entry market approval. So, those are all clearly laid out and those standards are, in fact, set in consultation with professional organizations and potentially other organizations.

That is a matter for further discussion to decide who is

involved, but the point I would really make here is that we are talking about a circumstance where professional organizations very concerned about lab proficiency and lab testing would have a place at the table, helping to determine what standards should be part of that approval.

Again, from our discussion today, this process would include an assurance that there is pre-test and post-test information, i.e., information about the indications for testing and ways of reporting test results that meet a standardized template. That is what, when we say "routine" or "Level 1," that is what we are talking about.

When we talk about Scrutiny Level 2, I think there is more discussion to be had about what constitutes the scrutiny that would occur. So, I am just in some senses earmarking this as something yet to be discussed, but as I hear it, the elements include, first of all, that there is a detailed review of pre-test/post-test information. In other words, the pre-test/post-test, test indications, test result reporting have to meet a template, just as they do for the standard review, but there is particular concern when something has been triggered into this higher level of scrutiny about the accuracy and completeness of the information and in particular, that all the caveats are there about testing. For example, test results would be reported in such a way that non-informative results are very easily determined, et cetera.

But then after that, there are questions about what this scrutiny might constitute. Certainly what outcomes a scrutiny might be because one of them might be that data collection is mandated or encouraged in some form. There might be others and actually I haven't put on, but I will put on the transparency reporting at least in some

voluntary form of adverse consequences of testing and as we begin to think about the outcomes of this scrutiny, we might also, I think, address a question that has come up a few times, which is how do we ensure informed consent.

Is there, perhaps, something additional that comes out when a text like this has been scrutinized to assure that informed consent might occur? I am just laying those questions out. Let me show the scheme that I think we have been talking about that puts a test into one or the other of the scrutiny levels and then we can come back to these questions.

Okay. Here, I think, is what the scheme looks like. So, on one side you have got Scrutiny Level 1. On the other side, we have got Scrutiny Level 2. We start with test volume. We put in the 4,000 tests per year as the threshold just at the marker, clearly, what the threshold should be and how it gets determined; that is, whether it is a matter of the offerers reporting it or some sort of surveillance mechanism, I think it is going to be an important discussion because this, in essence, is our definition of an orphan test.

So, there is really going to have to be some attention put upon it. But I think what we are saying here is this would be a very easily implementable way of setting a line between tests that bypass this because of the orphan test category versus tests that don't.

Once the test has met the target volume, there are two questions that get asked immediately. Question No. 1 is this test going to be used for population-based testing programs? There doesn't have to be universal screening, but it does mean screening of a population as opposed to using a test for clinical criteria. So, it would certainly mean screening all people of Ashkenazi Jewish descent,

for example, screening newborns for that matter.

The newborn example is worth commenting upon because a good newborn test will at a certain point become standardized and not something that generates a lot of concern. But when it is first proposed for use, what I think we are saying is it should receive careful scrutiny.

So, Question No. 1 is is population-based use proposed? If so, automatically you get into Scrutiny Level 2. If you pass that hurdle and the test is not being proposed for population-based use, then the next question is is it predictive versus diagnostic? So, this is, again, trying to capture what we have said.

If a test is used solely for diagnostic purposes, then it is triggered automatically into Scrutiny Level 1. Here is now where we have to have our discussion because we now have two options. What we could do is stop here. We could say if it is predictive -- you ask the question predictive. If it is not predictive, it is Scrutiny Level 1 and if it is predictive, you could bump it automatically on that basis into Scrutiny Level 2.

That would be one approach. The alternative approach would either to be to consider some or all of these characteristics. In other words, predictive, yes, and then determine whether one or more of these need to be present. I haven't drawn the lines and I haven't put "ands" or "ors" because that is what we have to have a discussion about. But if we are going to say, no, we are not going to automatically trigger Scrutiny Level 2 for all predictive tests, then we have to determine what characteristics of a predictive test would trigger that higher level scrutiny.

The kinds of characteristics that we have been talking

about I think have been captured here. One is treatment is unproven or non-existent. Another is that there is low positive predictive value. Now, the implication here is that this may address the stand alone issue. That is, what we are really concerned about when we go down a genetic testing pathway is how certain we are to get to an answer at the end of it.

If you have a genetic test that in and of itself is low penetrance, but is followed by a confirmatory test, then the testing procedure itself has high predictive value, even though the initial genetic test might not have in and of itself high predictive value. It is another way of saying why we care about whether something is stand alone or not.

Whereas, if at the end of the testing pathway you have got limited predictive value, that would be a consideration. Then, finally, is there a significant potential for harm either in a medical sense, meaning that the interventions that would be applied to test positive people are risky or in a social sense, meaning that the implementation of this test might generate enhanced opportunities for stigmatization of a group or of individuals, just recognizing that not all of those elements are fully defined.

But I think these were the categories of concerns that we raised that might trigger something. So, one possibility is to say if predictive, then consider these characteristics and if any of the three is present it triggers, or if some combination of the three is present. So, I think that is what we need to discuss now.

DR. COLLINS: I think I would advocate that if any of those three is present, that would be a reasonable argument to go onto Scrutiny Level 2. But I wouldn't say that you should just stop at

predictive and say that automatically carries you to Scrutiny Level 2. It is probably helpful to think of an example. So, for instance, let's talk about familial polyposis predictive testing, where there is an intervention, where the positive predictive value is pretty high if you had a mutation in APC and where the intervention may, in fact, be fairly drastic because it's a colectomy, but I think you would still say on your last category there, significant potential for harm compared to the benefit. It is probably not going to trigger that either. So, that would be an argument that that is the kind of test, which could make it through the whole scheme and still get bumped to the left here, to Scrutiny Level 1, whereas, if you didn't have those opportunities to go through those three check offs, everything would end up over in Scrutiny Level 2.

I would say, again, being concerned, we don't want to choke the FDA system with tests that maybe we are more comfortable with because of the criteria just mentioned, and have them, instead, focus on the things we are not so comfortable with. It would be good to have that option of having such a test end up on Scrutiny Level 1 side. Others may advocate that we should carry that argument even further and have an "and" instead of "or" in those last three. But I think "or" is the right answer.

DR. BURKE: Okay.

DR. KHOURY: Francis, the answer to the familial polyposis, it goes to "no" because its treatment is proven. It has a high PPV and there is no significant potential harm.

DR. COLLINS: Assuming that you would do maybe more than 4,000, for the sake of argument here, I'm not sure that is right.

DR. KHOURY: If we just say predictive for the sake of the

argument and don't use any of these other criteria because these other criteria, especially when we get down to significant potential harm, medical and social, will require a lot of discussion.

Let's say for the sake of the argument, we just use the simple scheme of volume, population based yes/no, and predictive yes/no. Is there any evidence or can people think about whether that very simply classification will overwhelm the FDA system. I can't think so, but maybe people will disagree with me, that most predictive tests today have either a low predictive value or treatments that are still in the sort of proven/unproven or have some potential harm.

So, I think the answer to one of these three questions is going to be yes, most likely. So it will automatically bump it into high.

DR. COLLINS: If I can just respond, I think that is probably true. You could also say another example would be hemochromatosis testing in a family where the disease has already occurred. Not population screening, but where you are looking at a sibling to see whether, in fact, they are also affected. I don't know what the volume is for that, but I suspect it is getting up there, 4,000 by now.

So, that would be also one that I think most of us would feel pretty comfortable not going to Level 2 scrutiny yet if you proposed that everything is predictive and it ends up in that category, you force that. I guess I like to be forward looking here and imagine that we should not set up a system that is predicated on current ignorance, but is actually prepared for future elimination. One hopes that will come.

DR. BURKE: Actually, let me just say that I think there

actually will be some discussion about what is predictive or diagnostic because in those examples -- in the FAP example, for sure, I think there are some people that would say the genetic test is diagnostic in a family with a known mutation.

So that is also another out for making sure that we don't overwhelm the system.

DR. WINN-DEEN: My example was the three predictive thrombolytic mutations - Factor V leiden, the prothrombin mutation. If you took any one or the combination of those three predispositions of possible thrombolytic events, we would say those were probably predictive tests and, yet, I think probably even in sum you would say that all of them could be Scrutiny Level 1. So, I think there will be examples where you wouldn't want to just throw all predictive tests over into Scrutiny Level 2.

DR. KHOURY: But don't they have low predictive value even in combination? I'm not sure what eh intervention would be -- should you give anticoagulation or ask women to stop oral contraceptive? Even with that example, you will answer "yes" to at least one of these three questions.

DR. BURKE: And I think there might be an important distinction between using a thrombolytic panel in someone who has had a thromboembolism at a young age versus someone, say a healthy young woman who wants to know before she --

DR. WINN-DEEN: I was thinking of the 20-year-old woman debating oral contraceptives, should every woman before they go on oral contraceptives have this screening, which would definitely put it in a high test volume kind of mode.

DR. BURKE: Right, but I think you could argue that this

might be a useful example to discuss because putting aside what would overwhelm the system, it seems to me you might well argue that should be higher scrutiny at least at the level of ensuring that the pre-test information that is provided has all the messiness of the information in it, you know those people who understand the limitations of the test.

DR. WATSON: I think the likelihood is that at the time when they predict a test is likely to come for review, but -- imagine if Factor V leiden testing had come for review at the time we all began testing for it, I don't think the issues you mentioned would really come into play. What you probably want is the things that triggered review for Scrutiny Level 2, pretty early in that developmental pathway that Francis described, but then in Scrutiny Level 2, what knowledge about that test that takes you back to Scrutiny Level 1 so that you don't leave it stuck there forever.

DR. BURKE: And what do you do at Scrutiny Level 2? Isn't that also implicit?

DR. WATSON: Oh, yes, because I think there is a lot of information collection to facilitate the transition back to 1.

MS. UHLMANN: I like this model because I think it is getting more simplified and that is what we need given the task at hand.

I think that the issue that I am struggling with a bit is the whole issue of predictive testing and I agree, we don't want to make every predictive test a Scrutiny Level 2.

So, I was wondering if the division, though, becomes that Scrutiny Level 1 is the predictive test that do have known treatment, which is consistent with the FAP example. But then the question

becomes what about something with HNPCC, where that is just not as good a test.

I think some of that comes in maybe in the pre-test and the post-test counseling because with their HNPCC, I would say that you have the treatment available and that would push that into a Scrutiny Level 1 category, but because of the fact that you are missing 40 percent of the mutations in known families, does that make it Scrutiny Level 2?

I mean, I throw that out for consideration.

DR. BURKE: Well, it certainly limits the positive predictive value. So, you capture it by our criterion.

MS. BOLDT: I just want to throw out consideration for early childhood testing, too, and where we should consider that with the schematic?

DR. BURKE: So, actually, that raises also, I think, the question of prenatal. In other words, we said population testing automatically triggers high scrutiny. Should proposed childhood testing and/or prenatal testing do the same?

DR. COLLINS: Does childhood testing fall into your category of significant medical and/or social risk and so is it covered already? I am not sure where to fit it.

DR. BURKE: I think significant risk is yet to be fully defined. So, we could define it as testing occurring before 18 and we could do the same with prenatal as well. My sense of the issue with prenatal is that there is a different threshold for predictive value in prenatal. When you have a test that you are thinking of using with prenatal, there really is a standard of predictive value that is virtually higher than almost any other testing situation and you want

to be sure that that is there. Now, I think that is so ingrained in how we do business, so to speak, that generally you don't see tests offered for prenatal diagnosis unless they are meeting that standard. I am not sure that means it shouldn't have the scrutiny to be sure that it meets that standard.

DR. BOUGHMAN: I want to make sure that we first of all aren't mixing apples and oranges here and that we understand it and remind ourselves that scrutiny level and amount of information are two different issues. Scrutiny Level 1 in a streamlined process can still demand a certain amount of information, disclaimer, disclosure, certain kinds of caveat as can Scrutiny Level 2. By saying something needs to be at a higher scrutiny level, what we are saying is that there are issues that need to be more fully and broadly discussed in the context in which this information is being presented for approval.

We have to be careful -- I am not sure, Mike, if something that went through and got approved on Scrutiny Level 2 ever slides back to Scrutiny Level 1. Scrutiny Level 2 can still end up with an approval, given those caveats. It may or may not every have to come back through certainly for the same intended use, unless there are major changes or --

DR. WATSON: But it depends on how you -- I mean, if you are reviewing this on a lab by lab basis, it may change as knowledge accrues.

DR. BOUGHMAN: That is right and the last point I wanted to make was I hope we don't get so bogged down in this group with some of the what ifs and details of this that we don't recognize the amazing progress has been made here, what was a real mixture of major issues. In that same context, given the comments that you made when in my mind,

one of the things that we have done is reminded everybody that population-based genetic testing has societal implications that are not available in the others. So, at first thought, that is one of the major reasons that that one gets kicked out. So, it is in there at the top and it is back in there at the bottom for others.

DR. BURKE: I just want to follow-up on one thing you said, Joann, and that is we want to be very clear that we expect there to be a template that is met for providing information prior to the use of testing guiding its appropriate use and for reporting test results in a way that makes them very understandable to the average clinician. I would think that would be the kind of standards we are after in either level scrutiny.

It seems to me that by defining that, I sense that we have reached a comfort level, that that has solved a lot of concerns that people have about genetic testing, so that when we get to Scrutiny Level 2, that is not where that kicks in. That kicked in from the beginning, but in Scrutiny Level 2, it seems to me there are two things that are happening. I recognize we haven't fully talked this through, but one thing that is happening is that there are enough caveats about the test that we are very concerned about how that information is presented.

But the other thing is that at least there is a consideration that something additional to that should occur. It may be that certain things will go to Scrutiny Level 2 and all that needs to happen is that the information needs to be appropriate. But I think we have said all along that there are going to be some tests, maybe a small number, where there is a societal concern that data collection go forward in some form or another.

As we mentioned the adverse consequences of testing might mean we might want special procedures for informed consent. Maybe there are other things. So, it seems to me the implications of Scrutiny Level 2 are twofold. One is sort of enough concern about the information about the tests that there is an extra level of review about what is said about it, what information comes with it and that it is at least a consideration of additional circumstances.

DR. COLLINS: There is one element of Scrutiny Level 2 that I didn't hear you say and that is the possibility that the review will conclude that this test is not appropriate for marketing. Right? So, it is not just a review of the information to decide whether they had it right. You could imagine situations where the information was right; namely, this test has no clinical validity or utility and you would decide this is not one for marketing, even though the information was accurately described. So, I assume that is intrinsic to what you are describing as far as Level 2, but there is a yes/no quality there.

DR. BURKE: Yes, I think that is true of both levels, isn't it? Scrutiny Level 2 says either not released or released with the following. Scrutiny Level 1 is not released but to Scrutiny Level 2 or released/not released or, bumped to Scrutiny Level 2.

DR. KHOURY: What we have here is amazing. Others might look at t as regression, but I think this is progress. Scrutiny Levels 1 and 2 share a lot in common, more than we think because if this template of presenting the data comes in and there is an initial process that puts it very quickly in one or the other boxes, but that, to me, is a major deal. If we can get that truth in marketing using some pre-approved criteria, professional organizations and industry and consumers and everybody working in government works on together, it

seems to me the major difference for the higher test scrutiny level from a public health perspective is that the tradeoff of allowing certain tests to be used, knowing that they are still far from being imperfect, that will allow that additional data collection to be used for future and prospective settings that will be required -- I don't know how it is going to be used -- that will allow the ultimate validation in the long run.

So, I think that data collection piece is the central component that should be there. But having said that, I don't think we need to throw out the content of data collection from Scrutiny Level 1 and because data will continue to be collected by CLIA as a genetic testing process. So, data is here in all of this, but in terms of priorities for spending either private or public health dollars will definitely go into Scrutiny Level 2, especially for big population-based testing which you had on the previous slide.

DR. BURKE: So, the kind of data collection that is maybe likely to come in Scrutiny Level 1 is more routine surveillance stuff; whereas, Level 2 might be targeted projects.

DR. KHOURY: NIH-type projects, public/private partnership.

MR. HILLBACK: I am sorry I missed a few minutes of this. This looks pretty interesting to me. I guess what my reaction is that in both cases standards are going to be set for what data needs to be looked at and the real difference is who looks at that data and the fact that there may be somewhat more data involved in Level 2 than Level 1, but I would hope that the consortium or whatever would help FDA and industry and academia and everyone else would also help with the definition of what ought to be looked at in Level 2.

But it seems to me it is a similar process. It is only a

question of who does the heavy lifting in the process in terms of the review and Level 2 kicks it up to a more formal -- I hate to use that word, but a more formal review at FDA or a less formal review.

DR. BURKE: I do want to sort of mention that we have sort of agreed the actual content of the review process wasn't part of our discussion today just because of all the other things we have to sort of figure out, but I think your point is well taken. In other words, that represents an unresolved issue that will be resolved at the end of our day that we need to reapply.

MR. HILLBACK: But I don't see that as a problem. I think what we have said all along is that if we are really primarily focusing on making sure that there is the ability to reproduce a test, which everybody assumes is back to our analytical entity, but it is still a crucial piece and that we have enough data to show there is some utility and enough utility to make it useful to the patient or physician and that we are presenting a transparent picture of the pluses and minuses of how this test might be used, we are basically saying the requirements for the two categories may be similar just more data and a review by a more formal body in Level 2.

To me, that has quite a bit of logic to it. So I think that is a pretty appropriate process.

DR. COLLINS: Can you put your flow sheet back up since it is --

DR. BURKE: Everybody has got a copy of it, I think.

DR. COLLINS: I just wanted to raise the issue of carrier testing and where it falls out here. Presumably if you are talking about carrier testing in the context of a family history, not population-based, and if it were a relatively common disorder, you

could imagine it wouldn't be an orphan test. So, it would come down from the top there and end up in your box that is marked predictive because certainly carrier testing is not diagnostic of an illness. It is predictive of the risk, not to that individual but to their offspring.

Here, again, I think there is a good reason why not to automatically then bump that to Scrutiny Level 2 because I think there will be plenty of carrier tests for recessive disorders that are highly penetrant for which therefore positive predictive value is high. Maybe our words ought to be chosen a little more carefully here in terms of treatment, unproven or nonexistent, I think here is an intervention that would be potentially of interest to some parents. Maybe that would be the substitution there for treatment.

So, getting at the significant potential for harm, non-medical or social, well, again, you have to decide how that applied in a particular circumstance. Again, if you are talking about the whole category of carrier testing for recessive disorders, I think I am going to argue once again that those last three categories are good to consider individually once you get down to predictive and the answer is "yes," then you ought to look at those three. If any one of them are "yes, they go to the right. If none of them are "yes," you go to the left.

DR. BURKE: It seems to me that, assuming that there is agreement about that around the table, that might be a kind of example that we would want to include to illustrate what we are talking about. So, for example, a carrier test for an autosomal recessive disease that is highly penetrant, that is not being proposed for population screening, would be an example of a test that we would expect in this

paradigm would go to Scrutiny Level 1.

DR. CHARACHE: This is something I just think we have to be sure is being remembered, not necessarily addressed right now, but with that whole diagram, we are speaking of the subject as though we are talking about tests, but we are really not. We are talking about diseases or genetic modifications.

What I mean by that is that if you talk about a given, whether it is say familial adenomatous polyposis, that is not a single test. The test that is being done at Hopkins is fairly different from the test that is being done in Lab Corps, totally different. So that in fact we have got a lot of tests here.

So, we are going to have to think, as we think how to implement this, that each one of these entities may represent about a thousand tests if it is done in a thousand labs and that has to be built in to how we construct the oversight.

DR. VOELKERDING: I just wanted to return to your Scrutiny Level 1 and 2 and Elliott touched on this point. I would assume that in Level 1, you would put standards set in consultation for professional organizations. I would perhaps assume that that would also be applicable to Level 2? Just for the record of documentation.

DR. BURKE: Actually, I think that is a point that is very compatible with what was just said, that we assume that these are very comparable processes, but I think that is a good point. Let's be sure to put that in as well.

DR. VOELKERDING: Then my second issue, which gets back to an issue that was raised in the morning and I don't know if this is the correct forum for it, but we did try to drive at this question of -- we have some categorizations, some processes, but we still haven't hit

that issue of what Elliott referred to as the point of inflection; that is, when would a test essentially be ready to come forward for either Level 1 or Level 2 scrutiny and that may not be what we are here to address today. But I just at least wanted to raise that again as an issue, which, again might be something when a new genetic discovery is found out.

The question is when will people start bringing things forward and getting some feedback, perhaps even from the FDA and other consultative groups about, well, okay, it is sort of time that it is okay to bring this stuff forward for review.

DR. BURKE: Let's make sure to note that. I have got three people to make comments, but let's be sure to come back to that because I think that is an issue that has come up.

DR. KOENIG: I have two comments and they are not related and the first in a way, to follow up again on this of what the 4,000 per year actually means and how that is determined or ascertained. Steve, can you say something about how FDA does that, because I am uncomfortable having it be our very first threshold if I have no idea how it really works.

If that is literally at the point that 4,000 tests are done and then, of course, how you know and keep track of those kinds of things.

DR. GUTMAN: Yes. We have very little experience. We have only had a single HDE and the case was one where the product manufacturer was in the room and was involved in the review of it, some clotting tests and it was clear it was being used for compassionate use and they were at most a couple of hundred cases a year. It was a no-brainer.

I presume, although I haven't thought this out at all, that the mechanism for understanding volumes would be that it is part of whatever program you put in place, you probably would need to have registration and listing of these tests and you would track volume through registration and listing. You would be able to track over time if a particular test became hot and went from 300 to 25,000.

I don't know of another mechanism. I guess you could also survey the literature, survey the Internet, talk to experts in the field, but without registration and listing, I actually don't know how you guys should know which ones are 3,800 versus 5,000.

DR. BURKE: I would be really interested in hearing from the professional organizations on this. I mean, we agree four thousand may or may not be the right number, but would it be viable? Would it be possible, doable to set a number and figure out what is below it and what is above it, roughly speaking?

DR. KOENIG: Can I just add to that question, though? Are we talking about the eventual, potential market overall or are we talking about a threshold of tests already done?

DR. BURKE: It seems to me it has to be the number of tests being done at the time.

DR. GUTMAN: It is volume per year. I mean, again, you can make any recommendation. You can change the number to 3,000, 3 million over two years. You can do whatever you want.

I just pulled this out of the HDE reg. God only knows how they actually created that number.

DR. BURKE: But I think, Barbara, for the concern you are getting at, it might well be that what the initial discussion, if the test is being offered and not coming for review because it is under,

that there may need to be some assurance that it is expected to stay under for x number of years to come.

DR. KOENIG: Right, because otherwise there are two different things. There are tests that start small and then there are tests for orphan diseases. I don't think they are always the same thing. But the volume is going to be small initially for both kinds of tests.

I wanted to also try and put the multiplex step into this.

DR. BURKE: Okay. Actually what I would like to do is hear a little discussion about that and then have Judy and Wendy comment and then come back to multiplex if that is okay.

DR. LEWIS: My comment was also on the same subject and part of what I was trying to sort out in --. I don't know enough about it to be educated but I am a little bit concerned about is there any reason why any of these orphan tests or any of the low volume tests might require higher scrutiny because it would seem to me that tests that have the lowest volume are the ones where there aren't as many labs doing them. There isn't as much interrater reliability, for lack of a better term, and are there any of those tests that might require higher scrutiny because patients really need more protection.

So, I don't know if it is the number alone that is the issue or are there other factors that we need to look at in terms of the orphan diseases, to make sure that the tests continue to exist, but to make sure that there is sufficient protection that people are getting good information about the test.

DR. BURKE: I think we are raising a crucial issue.

DR. BOUGHMAN: It is much, much more than a number. We are actually talking about, if you will, a culture change or mindset change

in the part of the laboratory -- the whole community that is not used to, in fact, going through putting forward this process.

The whole FDA approval process is something that a company seeks when it has a new device. They want it to be FDA approved. They want it to have gone through this process and have the Good Housekeeping Seal of Approval on it.

In fact, that is not the paradigm that we have been operating under in genetic testing. But what we are talking about here is that somebody who would want to put forward their test because they really believe even though I am only doing 1,500 now, that, in fact, this is going to be done everywhere, then, in fact, it is important to get into this process and get that stamp of approval for the test as it is going forward and the details of the process of course are going to have to be worked out in exactly that point.

But remember that back to Francis's big chart, you go from IRB into this category. So, in fact, that is where I would say that the patient protection would be, that if, in fact, it goes get kicked out because it is an orphan or the very low volume, that that, once again, doesn't mean that there would not be a caveat or an assumption associated with that that a small number of people or a certain amount of information in going with it.

DR. LEWIS: So, I would be comfortable knowing that this small test -- the test for the orphan diseases, that there is some kind of oversight, that it is the IRB rather than FDA, that there is oversight somewhere.

DR. WINN-DEEN: My thought on the orphan category is that what we really need to establish for these orphans is some kinds of centers for excellence category or stamp of approval or whatever we

want to call it, where labs that are offering these rare disease testing, that maybe we need a CDC-compiled directory of labs if the existing set of volunteer directories aren't sufficient.

So that people who need that testing would know not only that this lab offers it, but that they have some kind of stamp of approval on their process. And it may be that the only stamp of approval is that you go through. They have an IRB, they have fact sheets, whatever really minimum things that we require, you don't have to go to the FDA for that.

MS. UHLMANN: I just have a couple of points. I was looking at the Scrutiny Level 1 and the Scrutiny Level 2 and I just wanted to say we should take the question mark out of data collection for Scrutiny Level 1 because I think it has been shown over time that tests may initially be low oversight, but actually may become Scrutiny Level 2 over the years. I think we have seen that with a number of tests.

Also, the SACGT has set as one of its overarching principles, the involvement of the consumers. So, I think that that voice needs to be heard also in the standard setting in consultation with professional organizations, that consumers need to be there as well.

I look at carrier testing as also being a subset, in some ways, of the population-based. It is also not just for autosomal recessive conditions, but for X-linked recessive conditions as well. I am not sure where we want to go with the preimplantation and prenatal -- where these boxes go, but that is something that should also be considered in terms of a higher level of scrutiny.

DR. MARTIN: I just wanted to comment on the question about

test volume and the various types of tests being done. There are a couple of ways to do that. Steven mentioned the development of a registry for collecting that information. Also, out of our group in the public health practice program office, we do a survey about once every four years of laboratories, on laboratory testing in general and it is a sampling. So, we have to consider some other method to do this, but there is a possibility of piggybacking it in a sense on that survey to collect that information if the resources were there to do it.

MR. HILLBACK: I think what we need to remember is that we have two escalating systems, both that went together, I think, pretty well. One is the one that Francis drew that talks about whatever we call those various phases, one phase that is governed by IRB. One that is governed by IRB plus CLIA. Remember that CLIA and home brew today, which, in effect, you are still in in home brew in that level, requires that you know what you are doing the test for and it requires this information.

There is just no outside external formal review of that. What we have added is then another escalation that once you move beyond that phase, whatever we call it, into the next phase, you then have a formal review and we have two levels of formal review of the information and of the utility of the test, Level 1 and Level 2 of this scrutiny.

But up until then, it isn't like you are operating with no control, no overview, no oversight at all or operating in the system that we have today, which a lot of us believe worked pretty darn well anyway and what we are really saying is that some crossover point, we then add a formal, external step, which is what has bothered lots of

people is that it has always just been the labs saying, oh, we are ready to go to the next step, whether that is a university lab or NIH lab or a commercial lab.

Now, we are saying, well, no, someone else has to help you say that and it could be CLIA helping say that or it could be FDA helping say that.

But somebody does. So, I don't think you want to assume that there is nothing and then all of the sudden we get to this point because that isn't accurate. And I think it does escalate sort of appropriately as things move through this continuum of both time and these other measures we used.

DR. CHARACHE: If I have now introduced a test which is low volume and then takes off, does that mean I now have to step back and have my original test reviewed at the Level 2 level or does it just mean that everybody else who wants to introduce the test now that I am over 4,000, has to go through Level 2.

Now, that also raises a very key issue, which is what about the bad tests that are already on the market. If we are going to start thinking about Levels 1 and 2, we really have to think about some of the junk that is out there.

DR. VOELKERDING: One thing that hasn't been discussed today is that -- and I know many of you are aware, is that the CLIAC has a notice of intent for formation of a new genetic specialty under their review and if some of the recommendations of CLIAC include sort of what I would call a beefing up of the requirements for pre- and post-analytical information associated with tests. So, in this interim phase where a test we are referring to is low volume or orphan, given that I assume that to some degree there will be some increased review

by CLIAC on coming, that will add some additional assurances in that realm.

Secondarily with regards to numbers of tests that are performed, as far as I know, we really don't have, other than what Bob has alluded to with the CDC, the professional organizations probably don't have the kind of survey information that you would actually like to have in hand. There is some information via CAP proficiency testing, but real survey data on numbers of tests being done is something we really don't have.

I would make an estimate, just as a barometer to think about, if you look at Factor V Leiden testing, I would probably make an assumption that somewhere between a hundred and a hundred fifty thousand Leiden tests are being performed per year right now in the United States, just to give you a kind of a ball park is what I would assume.

DR. BURKE: I just wanted to capture some things that I think have come forward from the discussion. First of all, on the diagram that you have, you may want to note a correction here or at least something that wasn't clear when I first drew it. That is if the test volume is less than 4,000 a year, it doesn't trigger Scrutiny Level 1. It actually triggers no FDA review.

That is just what we have been talking about here. I think what I have heard the discussion saying is -- so, what we are saying is first you figure test volume at a certain level, and I am not sure we have completed the discussion about whether that is feasible, but we want some measure that is at least comparable to that. If it is below a certain level, there is not going to be an FDA review.

There is, of course, going to be CLIA supervision and in

terms of the comment that was made about centers of excellence, I just want to capture -- maybe there is something else. For example, a test offerer perhaps had to submit paperwork that says this is why we are not going for FDA review and this is what we are doing and it might be that over time we want to develop a mechanism that permits folks offering those kinds of tests, some sort of organizational structure to be part of, that ensures good information, delivery and so on.

The other things that are here, I think, are just things that capture our discussion, which is that we said predictive if "yes," then triggered this kind of evaluation. So, I think these are different.

DR. KHOURY: I guess one of the things we have been recognizing more and more, especially hearing Pat's comments about all of this is that this is -- this schema, while it is workable, is in initial process of going through the bottleneck, if you will, of initial review. What is going to happen is that there will be a lot of post-market kind of information that will come that will make things move across the different boxes, things that are under 4,000 per year might move to another box.

Things not population-based may become population-based, et cetera. What this underscores in my mind is the need for ongoing surveillance or data collection, both phase 1 and phase 2, across the board and this is sort of the kind of efforts that CDC and the states have done in many other settings. I mean, we collect data on infectious diseases. We collect data on all kinds of things to help the policy and the regulation process to protect public health.

The point that Bob Martin raised earlier is that there is an existing system to collect some information on lab tests, although

it is not geared toward genetic tests and it only occurs once every four years. But it is an important system that exists and it can be very useful to be beefed up and evaluate volume and the impact of testing so that things can be classified accordingly. We can work with states. New York is a prime example to make sure that this data can be captured. So, it comes back to my point about the two types of data, the data on the genetic testing process, which captures the volume, captures the pre-analytic and post-analytic phase and then as more data gets captured on the parameters of the test, that would be helpful for consortia through NIH- or CDC-funded research or private money coming together.

To me, this initial phase has to be tied intimately with a process of data collection across the board. Otherwise, we will be stuck with a system that we can't implement, but won't know its impact, especially as time moves on and things will change invariably even within the year for each test or each use of a test.

DR. BURKE: I also want to comment that our goal here is to come up with a test classification scheme. I don't want to shortcut any element of the discussion but I don't think we can work out every detail. So, for example, if we are comfortable saying that there is an initial evaluation that ought to be based on test volume in order to treat tests that are used relatively little in a different way, yet we recognize there may be both technical and other difficulties as to how you define that, I think we may have done our work in that regard in getting to the test classification scheme.

DR. COLLINS: I have two comments but they are connected. One is sort of a semantic one. I guess I am bothered by your issue of no FDA review for the low volume tests for two reasons. One is we are

on record as saying all genetic tests should undergo some level of FDA review and this sort of goes counter to that, but also I think several people have said there might be low volume tests that we would have concerns about and you wouldn't want them exempted in this sort of -- they are completely off the table kind of arrow as you have currently got there.

I think this all fits in with Scrutiny Level 1 being a relatively low level of review. Maybe this is 1A or 1.1, as opposed to 1.5 or something like that. But I think it still for our own internal consistency would involve FDA having a quick look at this. So, that if we take that away, we may not be able to get it back.

DR. BURKE: I know you have got another point to make, but why don't we talk a little bit more about this because this is very important in our global scheme. We probably want to talk then about whether as we are envisioning Scrutiny Level 1, it makes sense to just fairly automatically put anything that is low volume into the Scrutiny Level 1 box, because bear in mind that we have said Scrutiny Level 1 has these characteristics, streamlined, standards, consultation, blah, blah, blah, and can include bumping to Scrutiny Level 2.

So, certainly, Francis's point would be well-taken. If we are comfortable that Scrutiny Level 1 is a pretty easy kind of process for the test offerer to do. I think if we feel we need something even simpler, then we have got to talk about some exemption paperwork or something like that.

DR. COLLINS: Until we see it, we don't know, but I think conceptually you are okay.

DR. BURKE: And conceptually we know we want Scrutiny Level 1 to be simple. Okay. Good. So, that changes that completely. That

is, we were right the first time. Okay.

You have a second point, Francis.

DR. COLLINS: It is connected to that and I think we are spending some time trying to wrestle with it but I don't think we are going to and that really is the issue of what to do about rare diseases and orphan tests and I think we have come away with that by having this arrow. That is a good thing, but there is a whole host of issues here in terms of how to keep laboratories in business that are offering these kinds of tests and are feeling increasingly beleaguered about that process.

I actually think maybe tomorrow when the full committee meets, we might consider taking that on as a new task, a new focus, because we have touched on it at various points and had discussions and we certainly touched on it today. But I don't think we have really fleshed out this very important issue of what to do about rare diseases and tests that are done at low volume, how to make sure that whatever process we put in place doesn't just drive all of those opportunities completely away.

DR. BURKE: And I think that comes very nicely under the heading of access, which we said was not our main focus today, but was a very important issue to target for attention.

DR. VOELKERDING: Actually Francis hit the issue right on the sort of the head of the nail, which was this issue of the consistency of the previous language in FDA review and given what you have sort of revised now, does test volume really make an issue, is test volume even relevant at this point?

If it actually goes for a quick Level 1 review and it is a new genetic test, what you are saying is that it is going to a Level 1

review, regardless, and then there is going to be a triage.

DR. BURKE: If it is low volume.

DR. VOELKERDING: No, if it is any volume. If it is any offering of any new genetic test. I just want to -- I am not arguing one way or the other.

DR. BURKE: I think a new genetic test that comes in at high volume will have this evaluation occur. It depends upon what volume it comes in at, right?

DR. VOELKERDING: I guess I still don't understand how that relates to your statement that all new genetic tests will be reviewed by the FDA.

DR. BURKE: If I am understanding what we are saying here, we are saying all new tests for review and the first branch point decision is a volume decision and by making that, what we are doing is we are agreeing that if it is a low volume test, it is reasonable to start with a Level 1 review.

DR. VOELKERDING: Which is still an FDA review.

DR. BURKE: But it is still an FDA review.

DR. VOELKERDING: I am just trying to --

DR. COLLINS: Level 1 and Level 2 are both FDA. Level 1 is streamlined. Level 2 is not.

DR. BURKE: Does that answer the question?

DR. TUCKSON: If the low volume test has any of those criteria in Level 2, doesn't that kick you down?

DR. BURKE: What we haven't done -- what we have said will be one of the outcomes of the Scrutiny Level 1 might be that the people looking at it say, no, these has got to be Level 2. But we haven't defined what criteria might make them do that. It is not going to be

population -- well, I suppose it could. Yes, we could use exactly the same pathway.

PARTICIPANT: We could, yes.

DR. BURKE: The problem is I think that is going to happen all the time and then everything is Scrutiny Level 2.

DR. COLLINS: Again, I think, when it is rare test, there are all these other consequences that may be particularly important, like driving the test completely off the market by applying too much in the way of oversight on a small laboratory that is not prepared to deal with that. That is why I think this may be more than we wrestle with in the midst of trying to deal with the majority pathways here.

I think if we sort of stop and said, okay, Scrutiny Level 1 does not mean automatically streamlined. We have already said that. It could also mean, well, this particular one deserves a little more attention and when we get to the broader discussion about rare diseases, which I hope the Committee will, we could then try to define a little bit more about what the criteria would be to cause that kind of elevation on the test.

DR. FERNHOFF: There might be some instance where even though it would be low volume, there would be some absurd social testing for some absurd social reason that we would all say needs to go to Level 2.

DR. BURKE: So, it seems to me what I understand this to be it is probably not going to be a capturing of exactly the same stuff that got you to Level 2. But it is going to some direction that has to occur in the -- sorry. Let me stop here and say some tests are going to get Level 1 scrutiny just because they are low volume tests. And others are going to get to Level 1 Scrutiny because you went through

the pathway and they didn't meet any of the criteria for Level 2.

Those are different categories. What I think we are saying is that they got to Level 1 scrutiny because it was low volume, needs a somewhat unique review of -- and it probably needs to include things that will be discussed when we discuss access issues.

So, I think what we are marking here is that if you go to Level 1 scrutiny because you are low volume, there is a piece that needs to be fleshed out, which is how do you look at a low volume test and how you look at a low volume test may include these factors but may also include some critical access issues.

DR. KOENIG: Can I just interject but could we also actually have a category in which at that initial scrutinizing, that we have a category of CLIA only for a certain very low volume orphan diseases.

MR. HILLBACK: I think we ought to see how the regulators put all this together because Level 1 may not be all that different from CLIA anyway. I would rather not mandate to them how they do it.

DR. BURKE: Actually in a previous discussion we have said it might well just be having all your CLIA stuff looked at ahead of time.

DR. CHARACHE: I think one thing I am hearing is concern because the low volume test may not have adequate oversight, but I think that what we are saying overall is that we can't give the same level of oversight to all genetic tests without clogging the system and making it impossible for anything to get through.

So, what we are doing is using the public health approach, which is saying we are going to address first those tests that affect the largest group of individuals, which is not to say that the others

are not important. It is a strategy that we have to adopt to get anything done.

DR. BURKE: I would like to at this point go back to the point that Barbara brought up earlier and that is the multiplex testing and make sure that we spend a few minutes just thinking about that. Do you have any comments to make on that, Barbara?

DR. KOENIG: Well, no, just that one obvious possible strategy is just to consider when you a multiplex test, to take the test and categorize the multiplex test according to the highest level scrutiny piece in the package. That would be one sort of conventional way to do this. But then another way to think about it would be to ask the question but is there something qualitatively different about multiplex testing that means that you would automatically want it to have higher scrutiny.

I think that is maybe what we need to discuss. For example, one of the things that happens if we are concerned about things like consent, it is impossible if you have a test that has 20 things bundled together, there is no way that you can have the same level of information and counseling in a meaningful way with a person being tested or screened than you can if you are doing one test at a time.

So that that then creates a potential for people having a lower level of voluntariness for those tests, which they don't know much or either they get the results that they hadn't necessarily thought about very much or planned for. So, I think those are two ways of thinking about it and it would be interesting to hear if people have other ways of thinking about it.

DR. BURKE: Before we get into this conversation, Steve,

can you is the multiplex or -- this is my naivete about the technology -- would it automatically qualify as a kit?

DR. GUTMAN: I think so, yes.

DR. COLLINS: What if it is being done in house by some lab that is just basically 20 different tests on the same DNA sample.

DR. WATSON: we do that. That is what most multiplex tests are.

DR. COLLINS: We are talking about multi-allelic with multiple loci and multiple phenotypes.

DR. GUTMAN: If it is not on a chip, it wouldn't be.

DR. BURKE: Okay. So, we just to want to respond to some of these issues. Should a multiplex test be evaluated in terms of the highest scrutiny test on it or does it have inherently some other properties?

DR. WINN-DEEN: I think we have to differentiate in multiplex. There are two categories. One is one disease, multiple mutations, which is one category for which we can do all the same kind of counseling but on the lab scale it is a multiplex test.

The other is where you are doing risk panel in a population like the, a risk panel that one might order for an Ashkenazi Jewish population or some other at-risk populations. I think we need to talk about those in two different ways because the single disease, multiple mutations in the lab, that is a lab issue and the other is more, can we possibly counsel on ten different diseases and how do we handle that.

DR. BURKE: So is everybody comfortable with a multiplex test that is really just testing multiple different mutations? Obviously, there are test complexity issues that have to do with analytic validity, but it is not triggering other concerns.

DR. COLLINS: We can expand that to say a test like HNPCC, where you are looking for mutations in three or four different loci, but they are all connected with the same phenotype. That is more familiar. The counseling is sort of direct and straightforward. You are talking about things like positive predictive value, it just happens to be looking at a few more places. I think what you are raising, though, is a situation where it is multiple phenotypes that are being tested for as predictive.

DR. BURKE: So, we are comfortable that a multiplex test isn't the dividing line. A multiplex test that is testing for a single disease entity in one fashion or another doesn't trigger new and different concerns.

So, we are talking about a panel where you might be -- as Francis says, testing for multiple different phenotypes.

MR. HILLBACK: If you have a situation where you have five diseases you are already testing for that are all well-known, Gauche disease and a number of diseases that are common in Ashkenazim, and then you put them altogether in one test that you do together, are you trying to look for a major differentiation there just because they are done at the same time versus done five tests independently?

DR. BURKE: Yes. It is an interesting issue.

MR. HILLBACK: Versus SNP analysis, which is sort of fishing in the ocean and hoping to catch something.

DR. BURKE: For the moment we will be talking about that.

I do think you could distinguish between let's say a carrier panel that is testing for the carrier state for multiple different conditions, all of which are relatively common in the Ashkenazi Jewish population. And that is a test that might have a role

in prenatal counseling for a couple.

But it would seem to me that that test becomes just automatically very different if you add to that panel the APC1307K mutation. I mean, it has become a very different kind of test with really very complex counseling issues.

MR. HILLBACK: But I don't think by being multiplex even when it is multiple diseases, it doesn't automatically become more complex. It is a question of what is on that panel.

DR. BURKE: So, what we are really talking about is that -- I mean, it does seem to me that we are saying multiplex testing for different phenotypes includes, at minimum, some sort of evaluation of what the clinical implications of the different tests are.

MR. HILLBACK: I think it would go to the highest level. In other words, if you have a test as part of the panel by itself, you would consider high scrutiny, it sucks the rest of the panel to high scrutiny. If all the tests that are on that panel would be low scrutiny, then I don't know why the panel is any more than low scrutiny, but I am not hung on this big time. I just think that there is a certain logic I like with that.

DR. BOUGHMAN: I think it may have been Barbara or Wylie that made the comment that if you were screening for multiple diseases, you couldn't actually include all the same information as if you, in fact, were screening

-- I am going to challenge that comment because if I go in and I am screened for five different disorders because of my ethnic background, I think I deserve the pieces of information for all five, and it may be a longer form. It should be. I have been tested for five different disorders and I need to know something about that. I think that is the

balance.

DR. BURKE: I wonder if what we are saying here is, number one, that the scrutiny level is determined by the highest scrutiny test on the panel and, number two, that the expectation in terms of information is that the same level of detail is provided about each phenotype, as would have been provided if the test were done as a single one.

MR. HILLBACK: I think you are right. I think primarily we are doing it for our own convenience, not because we want to do something that wouldn't short circuit the process because we are doing a panel. So, I don't have a problem with it. I think you are right.

DR. LEWIS: And then the other thing is, it seems to me, that while we are doing the consent for that panel, the patient would have a right to say I want to know about four of these five and this one that I prefer not to have the information on.

DR. BURKE: So, that is part of the counseling challenge.

DR. LEWIS: Just because it is easier for us to do all five at once, if for some reason somebody doesn't want one of those tests done, that they have the right to say "no, thank you."

MR. FERNHOFF: Just to give an example from the newborn screening world now that many states are struggling with is the expansion of tandem mass spec, where you can test for 33, I believe, different disorders in states, in addition to the seven or eight that are already being tested for and states trying to inform all parents about all of these 40 or so disorders in about ten minutes on their way out of the hospital is a challenge. But it is happening.

It is the kind of example where you can -- it is one thing telling people about PKU. It is another trying to explain all 40

diseases, some of which we have no idea really what the treatment is and et cetera. But it is happening.

DR. BURKE: And I think that in terms of what we were just saying before, that raises a challenge whether what we have just said, which is that the level of information should be the same for all tests as if each test were given individually, if doable, beyond a certain number of different tests.

DR. BOUGHMAN: In the diagnostic mode, we are not talking about carrier screening, right?

PARTICIPANT: It is diagnostic, but it is population-based.

DR. BOUGHMAN: Right. But, in fact, the pieces of information in trying to figure out what would go on in this process, one would say we are screening for these 33 and should any of these become positive or show up as positive or need further follow-up, then we would go back, whether it is 7 or 33 that you are screening for.

Carrier screening is a different kettle of fish, I believe, because, in fact, it is a different level of predictive information that we are talking about.

DR. KHOURY: One aspect of multiplex testing, which maybe I was asleep when you said this before, is in relation to multiple gene loci that has something to do with the same disease or ability to intervene. If you have, for example, the methylene tetrahydrofolate reductase gene and some others, these genes are involved in the folate pathway and the combination of polymorphisms around these three loci would determine who will need the folic acid pill everyday for the rest of their lives to prevent cardiovascular disease or prevent a neural tube defect pregnancy. That is a multiplex situation similar to the Factor V Leiden in combination with a prothrombin gene and other genes

in relation to other -- a woman should or should not oral contraceptives.

So, this is another multiplex and maybe you discussed this first and you decided this was not useful, but the point from a scrutiny level is that this will add not only analytic complexity but the recalculations of the predictive penetrance. Predictive values will change and will require more data to be collected on the risks of various diseases if you carry Gene A and variant B, C and D at the same time, with or without this potential exposure or intervention.

I would think the combination of testing in contrast to what other people think will improve the predictive value, if any. So basically people who have a prothrombin tetrahydrofolate reductase gene plus this other gene plus this other gene will probably be more predictive as to who will need folic acid, which is different from what Tony Holtzman wrote in the New England Journal of Medicine saying how all genetic tests will have low predictive value. So, this is one way to actually the predictive value, multiplex testing along with or without exposure. I don't know whether this belongs in our discussion, but I am very excited about multiplex testing as a result of that actually.

DR. BURKE: What we have said was that when you are only testing for one phenotype that we have less concern about multiplex testing than we do when you are testing for multiple phenotypes. But you are adding an extra point that, in fact, clearly the test characteristics of a given multiplex test need to be considered as that test. That test needs to be considered in terms of its predictive value and may sometimes be enhanced.

DR. COLLINS: So, I guess I wanted to endorse the wonderful

ideals here of being able to counsel adequately when we have multiplex testing for multiple phenotypes, but to express deep concern about whether those ideals are achievable and maybe to suggest an alternative here that multiplex testing should, in fact, have its level of scrutiny driven by the test, which is in the media the highest level of scrutiny, but also a consideration about whether the multiplexing itself decreases the likelihood of effective information exchange with the patient about what this all means because, let's speak futuristically here, it is certainly possible there will be 30 or 40 predictive tests for a whole host of different phenotypes that will be offered to lots and lots of people of all sorts of different backgrounds and the idea that we could actually adequately counsel each of those people about each of those 30 or 40 possible phenotypes and have them make a truly informed decision about whether they want that test or not.

I hope that is an achievable goal but I think in many instances, we will fall short of that. It does seem that when you get to this kind of panel testing, that has to be a real consideration about how much scrutiny ought to -- in this kind of situation, you might want to do not only data collection on clinical utility and validity but on what was the understanding of this information. That is part of the process. You don't have any clinical utility if you didn't know what the result meant.

DR. BURKE: I was wondering, in fact, particularly where we are in the development of multiplex testing that if it is sort of an early stage where we can see it coming, that it might be useful for our Committee to note that as a particular concern in data collection; that is, that there is a particular need for data collection on this point

because it might ultimately be a very important parameter in determining the need for test scrutiny.

DR. BOUGHMAN: And I would just add that FDA and others are already in discussions, Francis, on these issues and why Steve made the response as quickly as he did because, in fact, when you talk about chip technologies that would come forward, rather than independent or different tests done simply because of the convenience of the sample they came in.

Those that would go through the full-blown review process, the labeling, the implications and so on are going to be kicked up to a very -- the 510(k) pre-market kind of level, which, in fact, does include a much broader series of questions and data availability.

MS. UHLMANN: I was going to make points very similar to Francis. I did want to point out, too, that just in terms of our routine medical care. Just looking at general medical practice, you go to a doctor. You get a blood sample drawn. You get a urine sample drawn. Lots of tests are done on those samples. So, in other words, kind of a similar comparison to multiplex testing being done and they are, obviously, not informed of the different tests that are being done on a blood sample and the different tests that are being done on a urine sample.

But the point is is that I think that multiplex testing is going to become standard of care. You are going to go into a doctor and they are going to run hundreds of genetic tests on one chip. We are not going to be able to counsel it. I mean, I could see our sessions now -- OK, well we are now at condition #52. I think we are certainly going to have to go to other models of getting informed consent, where patients are given brochures that summarize the type of

testing that is going to be done, but the key thing that we are going to want to be able to counsel about are the people that have the positive results. But to able to counsel about all the conditions that you are going to be screened for, many of which are rare and have less of a chance is not going to be practical.

DR. KING: I guess one thing about the term "multiplexing," I think it is a semantic term, but I think there is a difference between tests that are bundled together for convenience in the laboratory, like the Ashkenazi screening of putting four or five tests together versus a true multiplex testing for a phenotype or a group of phenotypes. I think maybe the term "multiplex" should be really reserved for those groups of tests that are being done like for the clotting disorders or for cardiovascular disease and not for just the convenience of running four or five tests together at the same time.

PARTICIPANT: Just running things in parallel versus --

DR. KING: Right. And I think the other aspect is that we, in fact, have to remember that we have been doing genetic testing on people on a routine basis without counseling all the time, when patients come in and they get a hemoglobin and they get a cholesterol and they get a protein, electrophoresis and some other clinical tests. Those are all basically genetic tests or largely genetically driven, much to the same level that we are talking about with some of the fine molecular tests that we are doing.

We haven't dealt with that very well in terms of counseling, but we have in terms of education of the health care professionals that deal with that kind of information.

So, I think that there are models of doing broad genetic testing where we can't counsel for everything, where things are being

tested currently and there are other ways in which the usefulness and the reliability and the appropriate use of that information is being dealt with in medical communities. I think that we have to think about that.

DR. TUCKSON: Wylie, in addition, I should also make the point -- I am glad you made the point -- we need to define this multiplex word very carefully. It is sort of made me laugh as I thought about until the time I was in my residency in the early eighties, we used to order -- you wanted to work up a thyroid disease, you just order a gigantic panel of thyroid tests. Nobody gave a darn about all the flotsam. You picked what you wanted from it and then it was in the early eighties when we started realizing that we needed to order the test that we wanted to have.

So, I think it is a little ridiculous to think that we would go back to an era where you could just order 500 tests and pick the one you want in terms it gives you the answer you are looking for. I don't think that unless Elliott knows something about the economies of his business, that it is so unusual that that would be in any way practical or cost effective.

So, I think that the kind of language that Richard just used, we need to define this term very carefully about what we mean by this.

DR. BURKE: Is it sufficient to call it a multiplex test to test for multiple different mutations, for multiple different phenotypes? I actually need a little help making sure that I define the difference between that and what you say, just bundling.

DR. WINN-DEEN: I think multiplex is one way and panel testing is another.

DR. BURKE: So, panel test is where we have been discussing and where we have our concern. Okay.

So, it seems to me -- yes?

DR. BOUGHMAN: Let me ask a different question for those of us who are trying to imagine the chips that might come forward. Would you look at an adult female chip, an adult male chip in different categories than you would a cancer chip for all the different cancers or a cancer chip that would apply --

[Multiple discussions.]

DR. WATSON: It sounds like the bundling is around the population kind of issue of ethnicity or some other population feature that bundles a bunch of things that are pulled together --

DR. BURKE: I think we are talking about two different kinds of panel tests. One kind of panel test is an ethnic or population-based panel test. That is where you might do Canavan's or Gauche's disease and certain disease mutations.

DR. WATSON: CF mutations.

DR. BURKE: So, that would be one reason for creating a panel and another reason for creating a panel really is the futuristic one. It is the genetic report card, according to The New York Times. Right? I mean, that is the one that is eliciting our greatest concern but we are also calling that a panel test and if I am understanding it, we are suggesting that multiplex should be reserved for phenotype specific tests.

I am not sure how the term is used now.

[Multiple discussions.]

Should we avoid the word multiplex because it has different meanings to different people, talks about different kinds of panels?

MR. HILLBACK: I would define them. I think if you are trying to evaluate a particular problem by testing for lots of genes that all relate to that one problem, that is one sort of thing. If you have another that is oriented towards a specific group of patients that looks at an array, that is a second.

Then I think the third one is the ultimate fishing trip, which is really non-specific and not necessarily huge. It could be ten or it could be ten thousand, but that one is not -- the other two are here.

[Multiple discussions.]

DR. WINN-DEEN: I think the other panel that is likely to come along and maybe it will be the first one that is actually on a chip is your sort of pharmacogenetic, predisposition panel, where you just get your whole SNPs and NAPs and everything else recorded for history so that anytime in your future life that you might need a drug that is handled by one of those enzymes, the physician can choose the right one or the right dose. That is probably a pretty low-risk panel, but it is still a panel of multiple things with multiple intended uses.

DR. BURKE: I just want to be clear that we will define the three different kinds of panels. We will avoid the multiplex. We will talk about the fact that we are less concerned about a panel that tests for multiple alleles, even for multiple genes when one phenotype is being assessed but when multiple phenotypes are being assessed, number one, their level of scrutiny is at minimum defined by the highest scrutiny level test on the panel, but we also have a particular concern about whether appropriate counseling can occur, appropriate informed consent can occur because our expectation would be that the same levels of information would be provided for each test on the panel, as would

be provided individually and that we have concerns about the need for data to determine whether that is feasible.

DR. KING: And the information on this point is going to need to be reiterated in the future as we learn how to do it.

DR. BURKE: As things develop, right, and also addressing Muin's point, that there may actually be some panels that may counseling easier because they enhance predictive value.

MS. UHLMANN: I just wanted to make the point, I think that the first category there, I think really is just referring to a methodology of doing a test. It is not really a panel -- that you are looking at multiple alleles to detect one condition or you are looking at different gene changes that detect one disease but the second category I still think that if we say that it is going to be defined by a test in the panel that has the high scrutiny, then I think by definition that almost all multiplex testing is going to become high scrutiny. So, again, I think we need to relook at that and consider that multiplex testing is going to become standard, it is going to become the most efficient way to test for multiple genetic conditions simultaneously and should all multiplex testing then be high scrutiny because that is what is going to happen if we use this definition.

DR. BURKE: I haven't heard any discussion around the table. You are raising a practical concern, but I haven't heard any discussion around the table that would say we have any rationale for reducing scrutiny. In other words, if we didn't say that the scrutiny was determined by the highest level scrutiny individual test, we would then be saying that the way to avoid high scrutiny is to get your test onto a multiplex. I am not sensing any willingness to do that.

DR. KING: Furthermore, we don't know about how to deal

with these issues. I think they all are high scrutiny right now because we really don't know how to deal with them.

DR. BURKE: Given these concerns about information exchange.

DR. LEWIS: It seems to me the issue is the level of scrutiny for specific tests, not how many tests are done at once. So, if it was a multiplex test and there was one test that was high scrutiny, if that test had passed approval, then is the issue the methodology or is the issue the concerns around the test?

So, I think that that is a piece that needs to get sorted out because if it is the fact that one particular condition, the test for that condition puts it to high scrutiny, then that test is for high scrutiny, it may be that the rest of the panel isn't. It is just that particular test.

DR. CHARACHE: I think this brings us back to that initial diagram that we had this morning. I think the question of the technology has to be thought of separately from the question of the importance of the test. So, we have been dealing this afternoon with the importance of the test. We have to remember the technology issue has to be brought into play.

DR. GUTMAN: Actually, Judy is just right on target, at least in terms of the intraregulatory paradigm that we apply. We see tests bundled all the time and, frankly, are more than happy to differentially treat them, depending on the intended use. Obviously, if they all have the same intended use, they are treated in the same way, but if they have different intended uses off of the same platform, we would be happy to process them as a PMA or as a 510(k) or to exempt them, depending on what the frame was or what the analyte was or, for

the same analyte, you would say, for example, depending on the frame, it might be a Class 2 or a Class 3 product.

DR. BURKE: I guess I am hearing enough discussion that sounds like there may be a reason to have a special consideration panel test for Level 2 review at this point in time unless they deal with a single phenotype. I am talking about panel tests for multiple phenotypes.

MR. HILLBACK: I think the tests that we do today -- I don't even know exactly how many different disease states we are looking at, but we are looking at seven or eight, but I think that they are all tests that would be considered in Level 1 right now. I am not certain of that. I would have to go check.

DR. BURKE: And these are tests for multiple phenotypes.

DR. KING: There are different diseases, but with the tests put together --

MR. HILLBACK: Put together using a technology that allows that a simple way to do that. So, it is not separate tests for each patient run sequentially, but it is all done concurrently.

DR. BURKE: So, we have agreed that we want to put in some language and this probably is a point for discussion or maybe just a point to be earmarked for continuing surveillance, but we have put in a point about our concern about information exchange and counseling.

Having said that, are we comfortable otherwise to say that an appropriate working or operational approach at this point in time is that panel tests should be accorded scrutiny according to the highest scrutiny test in their panel?

DR. KOENIG: I am not sure I am comfortable with that yet just because I do think that the quantity issue creates a qualitative

difference in the experience for the consumer. So, I will just put that --

[Multiple discussions.]

I am not sure, but I think it is an empirical question. My hypothesis would be that in the situation where you are testing for 30 different phenotypes at once on a particular -- the fact that you are making such a huge quantitative volume increase in the amount of information that needs to be talked about, exchanged, whatever, that that makes a qualitative difference in the experience for the patient, regardless of whether each individual test may be lower in scrutiny.

DR. WILLIAMS: As opposed to the actual conducting of the testing.

DR. LEWIS: The other question I would have is what is the reason for putting all those tests on the same panel and is there some kind of social issue, like stigmatization that might come out. For example, if you were doing panel testing for Ashkenazi Jews and they could come out and say, if you were testing for each disease separately, you would end up with frequencies for each disease, but if you were doing the panel testing, you could then say X percent of Ashkenazi Jews have genetic--, it would be real easy to make some statements that might look like they had some eugenic implications. I would want to make really sure that if looking at the reason why we were bundling the tests --

MR. HILLBACK: Who is going to make that statement?

DR. LEWIS: I don't know the answer to that, but I have read lots of interesting research that people come out and say crazy things.

MR. HILLBACK: Well, but that is not --

DR. BURKE: Let me propose an approach to this. First of all, we have got to sort of figure out what to bring to the Committee tomorrow morning and I think this is one of the things that we might have to identify as not fully resolved. I think what we are saying is that a panel test that tests for multiple alleles even in multiple genes for a single phenotype does not trigger additional concerns beyond what the genetic test would trigger to begin with.

But a panel test that tests for multiple different phenotypes is a more complicated entity because we can imagine circumstances in which it would not trigger additional concerns. It might easily go into Level 1 scrutiny and we can imagine circumstances in which it would not by virtue of its panel nature, not by virtue of the individual tests. I think perhaps the proper thing to do right now would be to identify this as an issue that isn't fully worked out, but to identify it as an issue that we think is possible to work out.

In other words, I am guessing that there is something additional about panel tests that needs to be added that helps us to figure out what goes to Scrutiny Level 1 and what goes to Scrutiny Level 2.

MR. HILLBACK: Remember in both scrutiny levels, we have an assumption that the laboratory doing the test is going to present the pre- and post-information and that is consistent through both levels of scrutiny. So, instead of presupposing that by doing a bunch of tests together, they are going to do a bad job, you review the information as presented. If they do a bad job, it is going to get kicked to Level 2 or they are going to be told to go back and fix your package if you want to get it approved.

So, I don't think you should assume that it can't be done

and, therefore, will make it automatically a 2. You make sure that they follow the rules. The rules are you need to counsel people for each of these tests that you are doing. The fact that you do them altogether is no different than doing ten of them independently, not multiplexed, and saying I am going to report on all ten independently. There is no difference.

DR. KING: I guess I would agree with that, but I think I agree with what Barbara is saying and our experience in the Ashkenazi population is going to be different from looking at 30 different genes in just somebody that comes into the clinic. We have a different driven population, a different level of education, different interests, different backgrounds. I think that there is something about multiplexing in that circumstance that makes it qualitatively different than just doing each test individually at the time.

MR. HILLBACK: But there is still the requirement to say this is what we are going to talk about before the test and what we are going to talk about after the test and either it is adequate or it isn't, whether you have multiplexed or not. I think we are getting the multiplexing in the way of the fundamental issue. Are we telling people the right things?

DR. BURKE: Yes, Elliott, I actually don't think there is disagreement and I think if we had another two or three hours, we could nail this one. But the point I want to make is it is interesting that you are saying 10 tests and Barbara is saying 30 and I think that might, in fact, be the dividing line. Once you have got 30 tests, it is too many and -- but I think probably it is an empirical point. So, I think what we are really doing is we are saying --

DR. TUCKSON: I think you are right. We will leave it

unresolved, but I think one of the things that we will need in the next time we look at this, it will be some sense of, again, the economics of this, the reason being is that I just cannot imagine, especially with the access issues we are concerned about now with the escalation of health care costs now, with the escalation of pharmaceutical costs now, I think you are talking about something very different about somebody ordering one test, two tests or three tests and having somebody to pay for that versus getting a panel back where you get 40 answers back. I mean, these are considerably different levels of information and it cannot be, I think, in practical real world terms looked at as, well, do you want us to run one test or 40 tests in a row.

They aren't going to order 40 tests in a row because they don't have that kind of money, but if you get 40 tests back on one dime, all bundled as a cost, that is a considerable different level of information with the potential for interrelationships between that information that changed the nature of the discussion.

I think Barbara is absolutely right. That changed the nature of the debate because you cannot sit at this table and predict the interrelationship between information that is multiplied 40 times versus one or two times. It is just unknowable and that, I think, throws it automatically into a different level of discussion.

DR. FERNHOFF: I think it is important we define panel testing just the concept, but again just to specify panel testing for diagnostic, for the newborn screening panel, testing for carrier screening, which is often on an ethnic basis and then panel testing for predictive testing, the report card. Those three are all panel tests, but they are aimed at different areas. We should just be clear about it.

DR. BURKE: Yes, that is helpful. I think what we will be able to do is sort of lay out the complexity of the issues, why it is not simple and why there clearly are some panel tests that we expect will readily fit into the low scrutiny and I think we have also identified need for empiric research.

DR. WINN-DEEN: I just wanted to make two points. One is that I think we can probably create some kind of a numerical cutoff as you were mentioning, whether it is 10 or 30 or something, some number that genetic counselors feel comfortable that they could have a reasonable interaction with a patient about. At what point is it overwhelming. The other thing to consider is that when you get into these really high density multiplexes, that most of these are going to be devices that are going to go through a different FDA process. They are not going to be home brews from CLIA-certified labs anymore and we may be worrying about something where there is already a mechanism and other people worrying about it.

I know FDA is not at all sure how they are going to array-based testing. But that is not a problem for this Committee as much as it is for how FDA is going to handle things when they get there.

MS. UHLMANN: I guess the point I had is in some ways more of a point of clarification, but just looking at the Ashkenazi Jewish panel, that panel is for tests that are individually already well-established. I guess that is what I looked at in terms of multiplex testing, that it is going to be put together after the test and individually already gone through a review process. But maybe I am mistaken on that. It seems that what you are really then bundling is existing tests that have already been somehow evaluated.

DR. BURKE: It seems to me in laying out the issues, we

have to encompass the possibility that that might not be the case, but clearly that might be one of the reasons why a test would readily go into Level 1.

MS. BOLDT: When I look at these panels though, we are going to bump it into high scrutiny when we start doing this for prenatal testing, though, and that is what I get concerned with the liability issues and the practicality of having someone come in because they are at an increased risk for triple screen or subscreen and then having to offer an Ashkenazi Jewish panel and talk about the ten different things that are available and the liability and standard of care issues.

So, I would think that bumps it up to the highest scrutiny, just the fact that we are doing prenatal, even if carrier testing establishes low scrutiny.

DR. COLLINS: I would be skeptical that we could come up with a threshold where you go above this threshold and then it kicks in a more complicated situation. It is going to depend on what kind of test you are talking about. I can imagine if you are looking at carrier screening, the threshold might be in a very different place than if you are looking at predictive tests for adult onset diseases, with or without a therapy available. I also would attach some skepticism to whether we are going to see kits burst forth anytime soon to do multiplex testing, using genetic variation.

I mean, if it has been as slow as it has for simple genetic tests to find their way out of home brews and into kits, admittedly maybe part of that was to avoid FDA oversight and we are about to take care of that, but from a purely technical perspective, multiplex testing on a chip is not going to be trivial. There are all kinds of

ways to screw that up and I would actually doubt that this will end up being a kit form for at least five or ten years. I think this is going to be primarily an in-house home brew situation for quite awhile.

MR. HILLBACK: I would just like to go back and say first of all I don't know why we are trying to prejudge this. I think each of these things, whatever it is, needs to stand on its own merits. I think we have built the outline of a system today that will do that and we shouldn't try to say, well, if it was me sitting there, here is how I would push this one around that system. I think we each do it a little differently, but I think we ought to rely on the system that we are creating. And there is a lot -- the devil is in the details in any system. What we have done is create a big skeleton and not much more. I think until we see more of the details, which Steve and other people have the excitement to try to put together over the near term, we are going to come back, I believe, and find some things when we see the details we are not going to like and we are going to want to make some other suggestions.

I don't think we can anticipate every plus and minus as we go. I also agree with Francis. I don't think there will be lots of these things turning into devices real soon and it wasn't just because we are trying to avoid FDA. The economics of it weren't there to begin with.

DR. VOELKERDING: Just because I have to leave in a few minutes, I was wondering if you were going to by the end of the day readdress the issue of will it be the FDA's role in the scrutiny process to determine which tests would require written, informed consent. I bring that up because it is also an issue that is in the CLIAC notice of intent.

The second aspect of that is that will this evolve into the issue of how to monitor compliance, both for informed consent and it gets at issues of post-market surveillance.

DR. BURKE: I think those are very helpful questions at this point. What I would like to do now is basically go back over -- we have got about 15 minutes and I actually think we are there. I think we are where we need to be, but what I want to do is just go over where I think we are, make sure that we have consensus and be clear about what we want to report.

What I am anticipating that we want to report is the scheme of different regulatory oversights that Francis showed us in his table. So, we are going to look at that table again. Then our definitions of Scrutiny Levels 1 and 2 and our scheme for the process of determining whether a test goes into Scrutiny Level 1 and 2 and, hopefully, as we do that, we can address the questions that Karl just raised.

Once we have done that, as we go, what I would like to be doing is capturing what we have got a consensus on and what issues we feel need further investigation or further points for discussion. Okay?

So, let's look first at this scheme. What this scheme says is that we can imagine genetic testing occurring in four different settings. Setting No. 1 is research and no data is returned to patients. That is Step A.

The next step is Step B1 --

DR. LEWIS: Can we call them 1, 2, 3, 4?

[Multiple discussions.]

DR. BURKE: We can discuss this as much as people want to.

The next stage is genetic testing is done in a research

setting on patients who are human subjects and test results are given back. The next stage after that, which we are calling B2 for the moment is that a test is now commercially available. Again, from a clinician's point of view, it is a very straightforward distinction because now you can order a test without there being a research protocol in which a patient is a human subject.

So, from a clinician's point of view, that operational definition is very clear. Then finally, there are two stages, but it might be some tests go directly to Stage C. So, I will just incorporate that possibility. So, B2 is where the test is now available, but there are still many questions around the test that amongst other things trigger a concern with what kind of information that doctors and patients are given about the test and also trigger a particular concern with ongoing data collection to resolve uncertainties.

Then there is the stage at which a test is approved. A fair amount is known about that test and, while there is always interest in ongoing data collection, there are sort of major unresolved questions. So, I think it is important to capture that you could imagine tests that go from A to B1 to C and other tests that go A, B1, B2 and stay in B2 for a long time. Sort of both sequences could occur.

DR. BURKE: You could imagine the tests going from A to B2.

DR. MANN: Wylie, I just have a question. I see a question mark in A for the consent. Is that to distinguish the kind of consent you need?

PARTICIPANT: You are not going back to patients, so --

[Multiple discussions.]

DR. BURKE: What sort of consent you need is really an IRB

issue. You are not giving data back to patients. You may be gathering identifiable information on identifiable subjects, in which case you would need consent.

The consent would be determined by the nature of the study. It is not actually intrinsically related to the fact that you are doing genetic testing.

Okay. So, the important thing about this table is that IRB oversight applies to Stage A and Stage B1 and not to subsequent stages. CLIA oversight applies to B1, B2 and C. And FDA oversight applies to B2 and C.

Are we all in agreement about that? I am not going to talk about information and informed consent issues here because I think they actually -- I think when we present this table, the purpose of this table is to show where IRB, CLIA and FDA occur. And actually we will get to the information counseling issues in our oversight scheme.

DR. KOENIG: The category of tests that sort of linger in B1 because they are orphan-like and stay there forever, is that going to be an issue? And do we need to think about that? That was what I was getting at when I said the other option. If the original scrutiny for FDA is that perhaps some will never -- is that part of the low volume?

DR. BURKE: Yes. I think I would like to talk a little bit more about lingering in B2, as opposed to becoming an orphan test. To my mind, lingering in B1 means it stays always within research protocol, where a person is a human subject. I don't think that happens long term, does it? I mean, at a certain point if a test still is being used, it is offered outside of protocol. People aren't human subjects and at that point if it is a low volume test, it has become an orphan test.

But please correct me if that is an incorrect understanding. I mean, research funding runs out. People stop offering it as a research possibility.

DR. COLLINS: I think the plan we put in place here actually was to make it possible for that kind of a test to move from B1 to B2 in this new system and FDA oversight we worried that wouldn't be possible.

DR. BURKE: So, again, we will show this table tomorrow, but what we will limit it to the columns of IRB, CLIA and FDA. I think that is the real purpose of this table. Do people agree?

DR. BOUGHMAN: Francis, in that set of issues, I understand why you have the sub-check mark under IRB in what we call B2 because, in fact, there may be continuing IRB kind of general oversight. But I wonder if we might have that be a caveat and remove the check mark because three months from now I might not remember that part of the conversation.

DR. COLLINS: Yes. Or even put a question mark after it because some of that data collection won't require IRB approval. It will be a matter of collecting information from laboratories and doing these FDA approved tests about what happened.

DR. BURKE: I wonder if we even should footnote it, in other words, take this line out of the major table and do it as a footnote here, as an asterisk, because what I think we have said is that one possible outcome of test review will be to ask for additional data collection, depending upon the nature of the data collection, it would be under IRB or not.

DR. COLLINS: I don't think you ought to take the data needed out of there or it will confuse people about the difference

between B2 and C.

DR. BURKE: Okay.

DR. COLLINS: Just footnote that particular more data needed with a little asterisk. You could do the same with consent. I guess I am a little worried about a question mark up there. It makes it sound as if we are not sure. Maybe there is not a process. Maybe that also just needs a footnote saying that that is under IRB oversight.

DR. BURKE: I am not sure we need the consent column. I am not sure it is informative for what we are using the table for.

DR. CHARACHE: I wonder if rather than put a question mark there or a footnote, whether you can just make that a plus/minus because sometimes you will want it, sometimes not. So, it is not a requirement.

DR. BURKE: I think the issue is is it clearer to keep it in or not keep it in.

MR. HILLBACK: I would rather footnote it personally and say "as needed" or something.

DR. KOENIG: I just want to make sure that when we present this tomorrow you make it clear that, even though most of the discussion is then going to be around the B2 at that juncture, that we are not ignoring the IRB issues and that our overall oversight effort is going to be really address possibly some of the issues of how IRBs should be involved at the A and B1 level.

DR. BURKE: I think you have just identified the first major discussion point or at least point of consideration that we want to bring back to the Committee, that IRBs need to be appropriately educated, offered appropriate guidance and coordinated as part of the

system.

So, that table and then --

DR. COLLINS: Wylie, there is one other thing. We didn't come back to that question mark IDE. I thought we would talk about it as we got into the scrutiny levels and I guess I didn't hear anybody advocating the need for that in a high scrutiny situation but I am not sure we really talked about it.

DR. BURKE: It seems to me that maybe we want to look at our scheme and then ask with that thought in mind do we have to go back to IDE. I think what I heard was that IDE was a possibility for dealing with oversight at Level B1, bringing FDA in at that point. On the other hand, we have said two things. One is that we want IRB, CLIA and FDA to be coordinated. Secondly, that we already have research subject protections through IRB at this level and that we might be creating an unnecessary redundancy if you brought FDA with an IDE at this point.

I think what we have also said that underscores the importance of Barbara's comment, we have said that there is a fair amount of concern around this table about whether IRBs are doing an adequate job of providing appropriate oversight at Level B1. So, I think part of what we are saying is this scheme works in theory, but it works in practice only if IRBs are brought up to speed.

So that one of the things that is an important goal of oversight is to figure out how to bring IRBs up to speed. One way to do it will be to educate IRBs about what we are coming up with in terms of test classification and concerns, but there may be other actions that need to be taken as well.

Is that satisfactory to people?

DR. KOENIG: We might even want to bring up the issue at some point of the idea at some time that data safety monitoring mechanisms are another possibility just to put out there.

DR. BURKE: So that is another issue to bring up under the IRB heading.

Then we are now talking about what happens when you enter into this B2 or C phase and that is you have now moved to a situation where a genetic test is going to be offered outside of human protection that is under IRB review. The decision has been made to offer the test outside of research protocols and once that decision is made, the test has to undergo a pre-market review.

That pre-market review is going to occur at one of two levels. I think I can go over this fairly quickly. I think we had good consensus on this. Scrutiny Level 1, we recognize we are not laying out all the details at this point, but we are laying out expectations in general about how this should work. Scrutiny Level 1 should be streamlined. It should include standards that are set in consultation with professional organizations, and others, including consumers. I have that on a note that came up in discussion.

It includes the assurance of appropriate information about indications given prior to the use of the test and appropriate reporting of test results after the test is done that follow a standard template and are aimed at the consumers, i.e., the non-genetics who might be using this test and includes the possibility of some kind of ongoing data collection. The result of Scrutiny Level 1 might be three different possibilities. A test is not released, a test is released or a test is bumped to Scrutiny Level 2 because of a bunch of questions that came up in this streamlined review.

Scrutiny Level 2, you get into Scrutiny Level 2 because you got to 1 and got bumped up or because you had test characteristics that put you into Scrutiny Level 2 to begin with. It requires also a review of the pre-test and post-test information, but by virtue of being in Scrutiny Level 2, we know that there is greater concern about the accuracy and detail and nature of information provided. We also know that there is a possibility that this may result in data collection and a greater likelihood that there may be particular targeted data collection procedures that are put in place around certain tests because of the concerns about them.

There may be other consequences at Scrutiny Level 2. The only specifics -- maybe we need to talk about this a little bit more. We talked about the possibility that there might be a voluntary reporting mechanism for adverse consequences. We talked about the need to ensure informed consent and I think this is where we get back to Karl's point about written informed consent.

Are we saying that it is that once you get to Scrutiny Level 2 that we could imagine some test having a requirement for written informed consent and only once you are at Scrutiny Level 2?

DR. KHOURY: I don't want to answer that question, but I think you should remove the question mark on the data collection because otherwise we won't be able to establish penetrance or clinical utility. That really is a situation like BRCA1 or even APOE4, if people decide that there is time. So take that question mark --

DR. BURKE: So, Scrutiny Level 2 results in some recommendations for ongoing data collection. Does Scrutiny Level 1 automatically involve data collection as well? Or is a question mark appropriate there?

MS. UHLMANN: I think there should be data collection for Scrutiny Level 1 just because it may over time as more that is learned about it, it may move to Scrutiny Level 2.

DR. BURKE: But we recognize that it is very likely that the data collection engendered by Scrutiny Level 1 review is maybe less and maybe more routine.

DR. KHOURY: I would characterize Level 1 as more passive data collection and Level 2 as more active data collection. You have to define those terms but passive data collection involves the forcing or linking of records at the diagnoses. With active, it is more people coming together around consortia and getting labs and clinicians working together.

DR. CHARACHE: I am going to disagree with that. I also think that informed consent can apply to Category 1 as well. Remember, the first cut point here is whether you have high volume test or low volume test and there are plenty of low volume tests that are going to require informed consent if not more so than the high volume ones because don't know as much about them.

So, I would say that both those items, data collection and informed consent, should be in both locations, both an as needed concept.

DR. BURKE: So, what we are saying is we could envision -- obviously, I am actually remember Judy's comments of earlier today and, of course, informed consent is always part of this process, but there may be some tests that we want to mark as requiring particularly focused efforts on informed consent and one way to make note of that is to require a signed informed consent form.

DR. CHARACHE: There will be a lot of tests in Level 2

assessment that will not require informed consent and there will be some in Level 1 assessment that will.

DR. BURKE: Okay. So, informed consent is an outcome of review. The particular or specified requirements for informed consent can be reviewed at either level.

DR. GUTMAN: Once you have characterized these as devices, actually medical device reporting requirements would go into place for all of them. So, you actually gain little by not suggesting that for Scrutiny Level 1 tests if there is an adverse impact, it makes no sense for me that it wouldn't be reported to the agency.

DR. BURKE: So, adverse consequences should be reported?

DR. GUTMAN: Yes. The other thing is and I think it is quite all right for this group to make a separate recommendation about informed consent, but my understanding is that the notice of intent for the CLIA program, in fact, puts that issue aside because it requires informed consent across the board. Is that not correct?

DR. CHARACHE: No, that is not correct. It says that informed consent will be required according to the categorization and the categorization is not addressed by CLIA.

DR. BURKE: You mean written informed consent, right?

DR. CHARACHE: Any informed consent, whether it is written or verbal. But the fact is that it is not required across the board by the notice of intent. It is according to this type of decision-making.

DR. KOENIG: Let's make sure we are not confusing documentation of informed consent with informed consent, going back to Judy's, because I think what you are talking about is documentation of informed consent.

DR. CHARACHE: Yes. Where informed consent is required,

the notice of intent says that it should be documented.

DR. BURKE: So, documentation of informed consent is what we are talking about.

DR. CHARACHE: Right, the documentation is required if informed consent is required. How it is documented was not specified. There could be a check mark on the requisitions that says the clinician who orders the test had gotten informed consent or it might be that the laboratory had to have a copy. That was not specified.

MR. HILLBACK: Well, I am a huge fan of collection of information of data, both on patients and I guess now on adverse events. I think everybody has to remember what onus we put on the laboratory and what we put on system-wide because the laboratory in most cases is not interfacing with the patient. We don't have access to patient records after we send the test result out.

So, while we can talk about collection of more data, I think we are going to have to really be careful how we define that because whether it is a commercial lab, a university lab, an NIH lab, if you start requiring the lab to spend money to follow patients, no one will do it. It is very expensive. We don't have the automatic access and I don't know what the adverse event reporting is, but if you think about, here is the patient. You have got this diagnosis and a year later something happens, we wouldn't know it. We never would or rarely would. Occasionally we do.

So, I think we just want to be careful where we are going with this.

DR. BURKE: I think it is very important to note that we are capturing the importance of ongoing data collection but we have not in any point in our discussion described how it would occur.

DR. KOENIG: On data collection, since there seems to be agreement that we don't know how to take social issues seriously, but we think in some ways they should be, I make the point that including social consequences in data collection as a category that should be collected and possibly reported might be a way of raising these issues higher in the public conscience.

DR. BURKE: So, in fact, one thing that should occur is the discussion of how we determine the social consequences of testing.

DR. LEWIS: I am not sure that we have decided that it is the laboratory that has to do the data collection. It seems to me that if we want the data to be collected, some of it is going to come from the laboratory. Some of it is going to come from the clinician. Putting the onus on the laboratory, who doesn't follow the patient, it seems to me the issue is what is the result of this process for the human that underwent it and for the clinician and for the course of care. It needs to be some kind of a partnership. So, I just didn't see this as terribly burdensome on the laboratory but maybe I am being over simplistic.

MR. HILLBACK: I am only reacting to some earlier comments on other days and other times.

DR. BURKE: But I think it is also true, Elliott, that we have had those conversations and I think people have acknowledged that that is an issue.

MR. HILLBACK: I just was doing my job.

DR. BURKE: Duly acknowledged. I think we have defined Scrutiny Level 1 and Scrutiny Level 2 and now I want to make sure that the diagram we have got fits our discussion. What we have said is that there is a test volume threshold that needs to be defined and it seems

to me this is another discussion point that needs to come. That is, we need to point out to the committee that it makes a lot of sense to define a line threshold. Yet, there remain important questions about how you measure that volume and what the threshold should be.

Reed, do you want to comment?

DR. TUCKSON: For tomorrow, I would hope and I won't open it up here, we have done it, but I would hope that Barbara, our ethicist, would at least assure us tomorrow that ethically that that is appropriate. That practicality so well expressed by Pat Charache, who made a wonderful sermon on that, don't clog the system, all those other things, but if you suddenly read the newspaper two years from now and 10 percent of the people -- so, 40 people got really screwed over and we said that was okay to put it in this group. Bottom line, from a patient/physician relationship and our responsibility as physicians to be concerned about the individual patient, is it okay to do a herd lumping that says it was 4,000, it was okay, not the number, the concept, ethically, is it okay? I leave it for tomorrow.

DR. KOENIG: Well, I did want to say one thing about it earlier and silenced myself, but it is very quick. I think one of the ways to deal with that is to say the very last thing on your chart, you could say that with the exception unless there is significant potential for harm and you could leave open the fact that even with low volume tests, that if you could imagine that there were some that had really significant potential for harm, that you could override the volume thing. I think that probably would be more reasonable.

DR. BURKE: I think there are two ways to approach that. One way is to put a box above test volume, which says is there significant potential for harm. The other way is to make sure that in

the Scrutiny Level 1, that is applied to low volume tests, that point is considered. Because I feel like that issue, the issue that Reed just mentioned, was really changed in nature when we agreed that every test would go through Scrutiny Level 1.

Actually, we should go back and define this, I guess, in Scrutiny Level 1. There is a Scrutiny Level 1A and 1B. 1A is Scrutiny Level 1, as we just defined it. 1B is Scrutiny Level 1 for low volume tests. I think one of the things we have said, that isn't fully defined, but I think you have just articulated what it was that we were most concerned about, which is when a test goes to Level 1 scrutiny because it is low volume, there may be a particular concern about its potential to harm and if there is potential to harm, it goes to Scrutiny Level 2.

DR. KHOURY: Potentially a second approach, which would be to put the box test volume after you put the other two boxes, the population-based and predictive. That way, you might shunt some of the low volume tests based on other criteria to the Level 2. In other words, don't use it as a screening process.

DR. BOUGHMAN: Let me flip the paradigm for just a second.

As somebody who is putting forth a test that wants FDA approval so I can tell my patients FDA has looked at this, I am going to make the assumption that this is going to be scrutinized at the highest level, Level 2, until I complete the data form and the questions that the FDA asks me to prove that it is lower scrutiny.

If it is, it is not population-based. It is predictive, but it does not meet any of the following criteria or my first one might be this is a very low volume test and it should come out of the informed consent research protocol category for the following reasons,

but the FDA might, in fact, suggest that along with that test, one of the pieces of information that goes with it is that it is clearly approved as a low or under this special category.

So, in fact, that is part of the information that is given, that this is a very special test. Only a few people do it. It is not the same kind of thing as cholesterol testing.

DR. COLLINS: I think that pretty much says what I was going to say and I think we have come back to the same kind of conclusion about what to do about this about an hour ago as well. Well, let me then say what I said an hour ago in case you forgot. That is, I think there are larger issues here about rare tests that we don't probably in the next two minutes have a chance to get into, that are going to warrant some additional attention to this. Not only from the point of making sure that harms are not done to patients because tests are being offered and marketed prematurely, but also that harms are not done because we drive the laboratories out of the market by applying overly intense oversight to laboratories that are already a bit embattled and doing only a few tests for a rare disease.

That is where I think the calculus here of how this Level 1B or 1A, whichever you called it, which is what kicks in when you have a rare test, needs some additional thought that I think probably the SACGT needs to pursue.

DR. BURKE: We have already earmarked that as something that we would bring up as something that deserves further discussion.

So, it seems to me that until that discussion occurs and might lead itself to some other conclusions, it is probably prudent to say there is an early test volume criterion and so, again, I think to realize -- well, we will need to go back and correct this, but test

volume low scrutiny -- let's call that Scrutiny 1A because that is the first thing that happens, right? If it is a low volume test, it goes to Scrutiny 1A. Scrutiny 1A is all the things that are in Scrutiny 1, plus a special menu for low volume tests.

That menu may include careful attention to its ability to do harm, but may also include careful attention to access issues that might actually lower the threshold of test acceptability. Okay?

So, that is point No. 1. Okay? If the volume criterion is met, that is, it is 4,000 or above or whatever volume that turns out to be, you then go to the next question. Is it being proposed for population-based use? Yes. Goes to Scrutiny Level 2.

No, we had a series of tests of questions that get asked. The first question is is it predictive or diagnostic. If the answer is it is not predictive, it is diagnostic, it goes to Scrutiny Level 1. If it is, in fact, planned for predictive, then we have three different criteria and these are stated as "ors," if any one of these results in concern it goes to Level 2. Otherwise it goes to Level 1.

Intervention, unproven or non-existent, low predictive value -- do we want to say low predictive value or low positive predicted value?

[Multiple discussions.]

This is not something we have decided and I don't think we can decide it.

[Multiple discussions.]

It is predictive value or significant potential for harm, either medical; that is, in terms of exposing people to interventions with significant risks or social. And we have talked a little bit about what those social harms might be. I think there are a number of

ways in which we know that there are details that need to be worked out, but what I think this scheme does that is much more concrete than anything we have done before is it tells us exactly what threshold conversations we need to have and it does that in a logical way.

So, I think what we wanted to do was to bring tomorrow a reasonable scheme for classification and I think we have that and we wanted to bring points for discussion. My notes are that a major point for discussion is the IRB issue, all the different issues that we have raised about IRB processes being coordinated, IRBs being brought up to speed, IRBs needing to be able to deal with oversight of genetic tests in that 2A step.

We have brought up the issue of data collection as something that is extremely important but will require attention to resolve a number of issues and those include how you do it. Who is responsible for it? What kind of data get collected? What you decide for different kinds of scrutiny? We are just acknowledging that we haven't worked out all those details.

We had another point for discussion is this whole how do you decide what the test volume threshold should be and then how do you measure it. In other words, we can capture that as a very important concept, but we don't know how to operationalize that at this point.

I think that is it. And then all those points about panel testing. Let me just write that down and then we will go to comments.

DR. CAGGANA: Just to reiterate -- I might have missed it -- for informed consent at Scrutiny Level 1, if we say carrier testing is Scrutiny Level 1, that needs some kind of informed consent, is that going to be operationalized so that there is sort of flow chart to use?

DR. BURKE: That is a good point. It is a point for

discussion. That is, we have noted that at either level, we might want to have a system for reporting adverse consequences and some tests might result in informed consent, but we have not defined how to decide which tests those are.

We haven't also decided yet what at Scrutiny Level 1 what would bump you to Scrutiny Level 2. So, there certainly are issues within our scheme that require some working out.

MS. UHLMANN: I think that was my point of clarification that at the arrow that says "predictive," that if it is not a predictive test, meaning it is a diagnostic test, it is going over to Scrutiny Level 1. There needs to be a mechanism where there is a low detection rate or has a low penetrant allele or a condition with a low penetrance can be bumped up to Scrutiny Level 2.

DR. FERNHOFF: I think that is in there because in Scrutiny Level 1 you had that option up there.

DR. BURKE: I think the other point I would make, coming back to some of these issues that we acknowledge aren't resolved, having to decide which tests need informed consent, for example, what we have said is that review procedures will occur using standard set in consultation with professional organizations, other appropriate agencies and consumers and these are certainly issues that the professional organizations have been dealing with a lot.

I think it is not unreasonable to expect that you could bring people together and get a consensus on those issues. Does that seem reasonable?

Other comments? Ann.

MS. BOLDT: In terms of counseling and education and how that ties in with the final recommendations for high scrutiny. I'm

assuming that that is Scrutiny Level 2. I want to make sure our language is consistent there, but I think will also be times when Scrutiny Level 1 may also require genetic education and counseling.

DR. BURKE: Let's identify that as a point for discussion. We have already sort of earmarked that. That is, we know that in Scrutiny Level 2 there may be a particular concern with informed consent and counseling. I think we should acknowledge that this scheme will probably result in a small number of testing being identified as having very great concerns about adequate counseling and that there needs to be a part of the working out at Scrutiny Level 2, there needs to be some guidance on those issues.

MS. DAVIDSON: Under unresolved issues, I just wanted to add orphan diseases as well. Under "Data Collection," I think we need to put something about privacy concerns, discrimination.

DR. BURKE: So, what we have said is that we haven't resolved a lot of questions about data collection and maybe other things, but I will also say we certainly know that any solutions have to include appropriate protections.

DR. CAGGANA: In New York, once again, there is a confidentiality law that is actually our civil rights law, that every genetic test requires informed consent. That is why I just keep getting hung on this as to where that falls in this scheme.

Also, getting back to the transition from A to B1, while we know that that goes on, by law, it is not supposed to and how does that get rationalized in our system?

DR. BURKE: It is hard to know. I am not aware that there is another state that makes that illegal. So, it is hard to know. To some extent any given locality you may have local laws that add

additional regulation, right? I mean, that is really what is happening.

DR. CAGGANA: Because if you are in the New York system then the rules are more stringent than the current CLIA rules and that is how you get exemptions. So, I don't know that all will fold into this.

DR. BURKE: I am not sure we can deal with it except we do acknowledge that there are some regions, some states, in particular, where there are more stringent requirements than federal regulation requires.

DR. KOENIG: I just want to raise one issue about the sequence of this and David was just pointing some of this out, too, but the original point when Steve put this up and started with test volume -- and correct me if I am wrong, but the point was to try and make this a more streamlined system and to get out some tests that shouldn't get full review.

So, by taking that out, that is the justification for starting with volume and assuming that low volume tests then don't get reviewed, they are sort of out. So, I just wanted to point that out. I don't know what to do about it.

DR. BURKE: We discussed that and we basically said our Level 1 review should be streamlined. Didn't we resolve it that way?

DR. KOENIG: That is how we resolved it, but I am not sure that is logical. I am not sure then if it still makes sense to start with volume unless you really are going to do something very different with very low volume tests, why would you start with volume at your first point on your flow chart?

[Multiple discussions.]

DR. BOUGHMAN: I think we would do that because all of these other issues that we were talking about, especially the data collection kinds of requirements, all of those issues don't follow for these low volume tests in the same way that they would for high volume tests. You would have exceptions for low volume tests in every other category. So, let's go ahead and identify those, identify a couple of issues, in fact, that are unique and most important about that. If it drops through that one, then it goes on into the more routine process.

DR. BURKE: In fact, I think we also did really say we have got a two level 1 scrutinies, one for low volume tests and one for the rest.

[Multiple discussions.]

DR. KHOURY: If you put back the diagram from earlier, instead of 1A and 1B, I was thinking of the predictive situation -- Huntington disease in a predictive sense. To me, that is a fairly high scrutiny situation. It is predictive and I was wondering whether by putting that first box under Boxes 1 and 2, you might shift away some of the rare conditions that could be used for either population-based testing or predictive with some of the caveats into the Type 2, which people now call 1B and keep that exemption for most low volume testing into one. In other words, let's not create 1A and 1B. Some of the low volume tests would be in 1 and some will be in 2 because I don't see a difference between 1B and 2.

DR. BURKE: Let me answer that as per the points that have come up around orphan tests and let others comment on it. Let’s not use Huntington’s because that is already out. I will use another example that is already out, but sort of has been less talked about and that is von Hippel-Lindau. If you have a situation like that, I think

you have precisely the situation that Francis was describing, where you want to be sure about access. That is, you have a situation where you have got a highly penetrant genetic condition. You have counseling situations in which their disease is known to occur in the family, where you can see the testing would have a tremendous utility for individuals within the family.

We probably need to have that discussion about what you need to do to ensure access and prevent labs from going out of business before you take the action you took because I think if you said first – I think von Hippel-Lindau would go to Scrutiny Level 2 if you put it through this rubric. I think there might be a legitimate argument for saying this is an extremely rare test in a high value family and we want to be careful of how we do that.

DR. KHOURY: Diagnostic or predictive?

DR. BURKE: Well, is it diagnostic or predictive when there is an asymptomatic person in the family? That is why the testing is done in families is to find the asymptomatics early in life.

I just want to throw that out as another example and let other people comment on it.

DR. BOUGHMAN: I hope, if not consensus, at least convergence that the Committee has come to by the end of the day is really a, what I think, is a very thoughtful and comprehensive, although not yet complete pointing out or discussion of the issues and what we believe is a good starting point for a flow diagram kind of approach to this and that we don't overproscribe the order or whatever, at least at this point so that what we do is we hogtie the FDA. I think that to give them the broad brush strokes and highlight the issues that need to be covered is where we need to be at the end of

today.

DR. BURKE: Is it also fair to say vis-a-vis your point and Muin's point that we don't know yet until we have had that discussion what kind of special concerns we need to have about low volume tests?

DR. COLLINS: I wanted to underline that point again and also to say it is amazing how far we have come here on this. We have been trying to do this for five years and every time ended up crashed into a stone wall and I think we have finally gotten through the wall, but let's not let the perfect be the enemy of something pretty good by trying to find every possible tweaking of this that might improve it and in the process we might actually have the whole thing collapse under its own weight. I think this structure captures the major points that the group wanted to make and leaves some flexibility for tuning as we begin to consider special things, like rare diseases and balancing the issue of access versus the issue of not doing harm and also gives the FDA some breathing room in terms of how to implement this.

Finally, I would say I think there is another big step here, which is to consider what really are the consequences of this if we were to apply this to a long list of possible tests that are either out there now or soon to be out there and also fold into that what would FDA actually do with regard to Scrutiny Level 1 and 2. What would that translate into in terms of steps and months and all of those specifics that this Committee, I think, is very concerned about.

Then you may very well have to come back and do some tuning, but I think, Wylie, you deserve a lot of credit for having led this group through an amazingly difficult conversation.

[Applause.]

DR. BURKE: I just want to follow up briefly on Francis's

comment and say I think in our comments about low volume/orphan diseases, we need to say, number one, that there is a discussion that has to be had and, number two, that we think it is very likely that that discussion will lead to validating this kind of approach, at least for starters, until we see how it comes out, but that it is possible that that discussion may lead to no, this isn't the right place to start. You should just go right to that.

So, I think we should capture that both possibilities might occur.

MR. HILLBACK: I am going to ruin Francis's day again and agree with him. I really think we have come a long way. I also want to reiterate what I said earlier, which is I think the devil is in the details. The devil is in, as Francis said, taking this a little further, letting the regulatory agencies take it further, trying to model it, trying to play with it, trying to understand what it means in practice. There are lots of little questions we raised today about, well, you have had something out for awhile and now you want to add more mutations. Well, I think the answer is if you want to add more mutations, the answer is variable, depending on what the mutations are and what it does to the test.

So, I think there are a lot of questions like that that we can all speculate, tie this up. We could be here all night. Nobody wants to do that. I think we have made good progress and we ought to turn it over to the folks and say take this into another level of detail, take some of the issues further, and then we will look at it again. I think that is the way this will be successful.

DR. BURKE: Two very important points. You are saying there needs to oversight of the oversight process and the other is I do

think that this whole issue of upgrading tests, that is, what kind of process involved, is something that we will want the regulatory agencies to look at.

MR. HILLBACK: There is a set of laws on ASRs and I don't know where that fits in all of this. One of the things in preparation for this week, I asked what is it taking Genzyme genetics to deal with ASRs, which is registering all our reagents and validating them all. We will add at least two full time people that will do nothing but that for their entire life until they can escape. That is not a non-trivial cost for us and that is just registering the reagents.

So, we are, as we did ask in the full Committee the last time, we are going to have to look at what is the cost impact, get at some of these other access points. But you have got to start somewhere and I would like to put the peg in the ground and I am sure sooner or later some part of this I won't like, but we can all do that today. We can all do that tomorrow. I think we need to get on with it and get moving.

DR. KHOURY: I would like to personally thank you for doing what you did today because you took us in one day more than five years, as Francis said. And as you all remember my optimism this morning. So I am optimistic. To go back a little bit to what Elliott said, the devil is in the details, but we have the beginning of a framework and I think the implementation of this will require the management of the regulatory and non-regulatory agencies within HHS because all that stuff and implementation of data collection, we all have to work together, NIH, CDC, HRSA, HCFA and all of us, along with the private sector and the professional organizations and everyone.

So, while the devil is in the details, but the details will

be worked out and I can assure we will do whatever we can from where I sit to help in that process.

MR. HILLBACK: Can I follow on that? I think it is very important that all of us can find some little thing wrong here and I am sure we will and we will work this out, but I think it is very important that there is lots of professional organizations represented here today, lots of other groups, that what we have to go away from here is to say we have a good start and it is part of a process of working together to get to a result and make sure that we keep the positive attitude rather than end up starting big squabbles based on very limited data. I think we have got to give this process a chance to go another couple of levels of detail and then we will find something to argue over, but that is okay.

DR. BURKE: It is also reasonable to anticipate that we will develop the best program we can, implement it and find it has got problems that need to be fine-tuned.

DR. MARTIN: Given these comments that we have had in the past two minutes about where do we go from here and turning some of the responsibilities over to the federal agencies and professional organizations reminded me that the next meeting of the Laboratory Forum will be coming up in September and that is exactly the incentive of that group is to take some of the considerations that have been dealt with here and to help move those forward.

DR. MCCABE: I just wanted to thank the entire working group. I really appreciate the work that you have put in today. I have been very impressed with the progress that has been made. One of my calls this afternoon was to Dr. Beverly Malone, deputy assistant secretary for Health and Human Services, who will be meeting with us

tomorrow.

I reported to her that we would have a flow diagram to discuss tomorrow because I was so impressed with the progress to that you were making. So, thanks to all of you and especially thank you, Wylie.

[Applause.]

[Whereupon, at 5:35 p.m., the meeting was concluded.]