CENTERS FOR MEDICARE AND MEDICAID SERVICES

MEDICARE EVIDENCE DEVELOPMENT AND COVERAGE
ADVISORY COMMITTEE

- - - - - - - - - - - - - - - - - - - -x
:
IN RE: GENETIC (GENOMIC) TESTING :
:
- - - - - - - - - - - - - - - - - - - x

Woodlawn, Maryland

Wednesday, February 25, 2009

The above entitled matter came to be heard before the Medicare Evidence Development and Coverage Advisory Committee in the CMS Auditorium, 7500 Security Boulevard, Woodlawn, Maryland.

COMPOFELICE REPORTING SERVICES, INC.
6671 Farbell Row
Columbia, Maryland 21045
(301) 596-2019 (410) 381-2755
(800) 464-3019

FAX (410) 290-7249

PANELISTS:

CHAIR:
BARBARA MC NEIL, MD, Ph.D.

PANEL MEMBERS:

STEVEN PEARSON, MD, MSC
MINA CHUNG, MD
MARION DANIS, MD
CATHERINE ENG, MD, FACP
MARK D. GRANT, MD, MPH
CLIFFORD GOODMAN, PhD
JAMES PUKLIN, MD
MAREN T. SCHEUNER, MD, MPH
TERESA M. SCHROEDER, BS, MBA
DEBORAH SHATIN, PhD

CONSUMER REPRESENTATIVE:
LINDA A. BERGTHOLD, PhD

INDUSTRY REPRESENTATIVE:
ELEANOR M. PERFETTO, PhD, MS

GUEST SPEAKER:
RALPH COATES, PhD

GUEST PANELISTS:
STEVE GUTMAN, MD
NEIL HOLTZMAN, MD

CMW LIAISON:
STEVE E. PHURROUGH, MD, MPA
LOUIS JACQUES MD

EXECUTIVE SECRETARY:
MARIA A. ELLIS

C O N T E N T S

P R O C E E D I N G S

MS. ELLIS: Good morning. We're going to get started.

Good morning and welcome committee chairperson, vice chairperson, members and guests. I am Maria Ellis, the executive secretary for the Medicare Evidence Development and Coverage Advisory Committee MEDCAC.

The committee is here today to discuss the evidence, hear presentations, and public comment, and make recommendations concerning the requirements for evidence to determine if diagnostic use of genomic testing and beneficiaries, with signs or symptoms of disease, improves health outcomes in Medicare beneficiaries.

The meeting will discuss the various kinds of evidence that are useful to support requests for Medicare coverage in this field.

The following announcement addresses conflict-of-interest issues associated with this meeting and is made part of the record.

The conflict-of-interest statutes prohibit special government employees from participating in matters that could affect their or their employer's financial interest. Each member will be asked to disclose any financial conflicts of interest during their introduction.

We ask, in the interest of fairness, that all persons making statements or presentations also disclose any current or previous financial involvement in a company that manufactures or provides devices or other tools for the research of genomic testing.

This includes direct financial investments, consulting fees, and significant institutional support. If you haven't already received a disclosure statement, they are available on the table outside of this room.

We ask that all presenters please adhere to their time limits. We have numerous presenters to hear from today in a very tight agenda, and therefore cannot allow extra time. There is a timer at the podium that you should follow.

The light will begin flashing when there are two minutes remaining, and then turn red when your time is up. Please note that there is a chair for the next speaker, and please proceed to that chair when it is your turn.

We ask that all speakers addressing the panel please speak directly into the mic and state your name.

For the record, voting members present for today's meeting are Steve Pearson, Mina Chung, Marion Danis, Catherine Eng, Mark Grant, Clifford Goodman, James Puklin, Maren Scheuner, Teresa Schroeder, Deborah Shatin.

A quorum is present and no one has been recused because of conflict of interest. The entire panel, including nonvoting members, will participate in the voting. The voting scores will be available on our website following the meeting. Two averages will be calculated, one for voting members and one for the entire panel.

I ask that all panel members please speak

directly into the mics, and you may have to move the mic since we have to share. If you require a taxicab, there is a sign-up sheet at the desk outside of the auditorium. Please submit your request during the lunch break.

And lastly, please remember to discard your trash in the trash cans located outside of this room.

And now I would like to turn the meeting over to Dr. Steve Phurrough.

DR. PHURROUGH: Good morning panel. Thank you very much for agreeing to be part of this. This will be an interesting discussion today. And for those who are out in the audience, we thank you for your interest also.

This is the first of two sessions we are having on genetic testing. There will be another meeting in May that will discuss genetic testing used for screening purposes.

I want to spend just a minute discussing the difference in diagnostic and screening tests as Medicare sees them.

We will hold very rigidly to Medicare's definitions, which may in fact have nothing to do with the reality. They are in fact what Medicare -- what the laws and statutes outline them to be.

A screening test is not a diagnostic test. Even though you may use it to make diagnoses, a screening test is a test that's used in patients who have no signs or symptoms of illness or disease. A diagnostic test is a test used on patients who do have signs and symptoms of illness or disease, somewhat artificial separation, but a separation that's within statute and law, and we will hold to that, and today we're discussing diagnostic tests. And so we are looking at genetic testing used in patients who in fact have some sign or symptom of disease.

And so if discussions are to move into another realm, then Barbara will, as she always does so well, be able to move it back into this very clearly defined focus here.

We look forward to a very vigorous discussion. These meetings typically are vigorous

discussions.

They are time-limited, so Barbara will be very diligent about holding people to time. Recognize that, when you have six minutes to make your presentation, that your sentences may be cut off in midsentence. And so be cognizant of the amount of time that you have and --

DR. MC NEIL: You're finished.

DR. PHURROUGH: -- and move on. See what I tell you. You're good.

Also, I unfortunately need to step out in midmorning. So Dr. Louis Jacques will be the government representative for the rest of the morning.

Barbara?.

DR. MC NEIL: Well, actually, I think Steve said it all. I really don't have too much more to say. As a matter of fact, I don't have anything to say, other than I think it will be an exciting session. I think we are very much looking forward to the original -- the two presentations that you see in the agenda this morning from CMS and from CDC, as well

as from several of the presenters.

I'm hoping that we can have a very rich discussion among the panelists and the various speakers so that we get as many issues out on the table as possible before we start the panel discussion ourselves, which is scheduled to start after lunch, but I'm thinking it would be better if we actually started a little bit before lunch so that we have as much time as possible to think through these issues.

So with that, I would like to have the panelists go around and just introduce themselves with about one sentence, so that we're all clear on who is who. So we'll start.

DR. HOLTZMAN: I'm Neil Holtzman. I'm a guest panelist. I did share the first Department of Energy, National Institutes of Health task force on genetic testing in the 1990s.

DR. GUTMAN: I'm Steve Gutman. I'm also a guest panelist. I'm a professor of pathology at a startup medical school in Florida, and formerly an employee of the FDA.

DR. PERFETTO: I'm Eleanor Perfetto. I'm with Pfizer, and I'm the industry representative to the panel.

DR. BERGTHOLD: And I'm Linda Bergthold, and I am the consumer representative to the panel.

DR. SHATIN: Deborah Shatin of Shatin Associates, background in pharmacoepidemiology and health services research.

MS. SCHROEDER: Teresa Schroeder, Director of Clinical Affairs, Musculoskeletal Clinical Regulatory Advisers, and I'm here on behalf of patient representative.

DR. SCHEUNER: Maren Scheuner. I'm a medical geneticist of health services and policy researcher at the RAND Corporation and also at the VA in Greater Los Angeles.

DR. PUKLIN: I'm Jim Puklin. I'm a Professor of Ophthalmology at Wayne State University's Kresge Eye Institute and also chairman of the human investigation committee at Wayne State University.

DR. GRANT: I'm Mark Grant. I'm an

associate director at the Technology Evaluation Center, Blue Cross Blue Shield Association. My background is geriatrician and epidemiologist.

DR. GOODMAN: Cliff Goodman, Senior Vice President of the Lewin Group, a healthcare policy consulting firm based in Falls Church, Virginia.

DR. ENG: Catherine Eng, internist and geriatrician, Medical Director of On Lok Senior Health Services and also clinical professor at UCSF.

DR. DANIS: Marion Danis, Chief of the Ethics Consultation Service at the Clinical Center of the National Institutes of Health and running that section on ethics and health policy in the Department of Bioethics at the National Institutes of Health.

DR. CHUNG: Mina Chung. I'm a cardiologist, cardiac electrophysiologist at the Cleveland Clinic.

DR. PEARSON: And I'm Steve Pearson. I'm an internist and President of the Institute for Clinical and Economic Review at Massachusetts General Hospital in Boston.

DR. MC NEIL: All right. I'm Barbara McNeil

and I'm head of the Department of Health Policy at Harvard Medical School and a radiologist at the Brigham and Women's Hospital in Boston.

So let us start. We're going to first start off with hearing about the CMS voting questions from Maria Ciccanti -- Ciccanti rather, and Jeffrey Roche. There they are. Yeah.

And I assume all of you have these either in your packets or from the table outside.

MS. CICCANTI: Hi there and welcome. I'm simply going to read the questions that will also appear up here. Okay.

So this is -- the whole thing is about getting your discussions, your opinions about the desirable characteristics of evidence, or again, as Steve mentioned, diagnostic genetics, including genomic tests.

Our team, which includes Steve, as you met this morning. Also Dr. Louis Jacque, who is sitting in the audience, is division director for drugs, items and devices. Jeffrey Roche, which is right over here,

and myself, Maria C. when Maria E. and I are in the same room together.

We wish to obtain the MEDCAC's recommendations regarding the desirable characteristics of evidence that could be used by the Medicare program to determine whether genetic testing as a laboratory diagnostic service improves health outcomes. The questions should be addressed in the context of diagnostic testing, as Steve mentioned.

So question number one. Are the desirable characteristics of evidence for diagnostic genetic testing different than the desirable characteristics of diagnostic testing in general?

Question number two. What are the desirable characteristics of evidence for determining the analytical validity of genetic diagnostic tests?

Question number three. Beyond aspects of analytical validity considered in question two, are there meaningful differences in the desirable and/or necessary characteristics of evidence about the effect of genetic testing on outcomes for three testing

paradigms listed in the following slide? If yes, please consider question four separately for each paradigm. If not, please consider question four to apply equally to all three, diagnostic assessment, prognostic assessment, pharmacogenomic assessment.

Question number four. For each type of outcome on the following slide, how confident are you that methodologically rigorous evidence on the outcome is sufficient to infer whether or not diagnostic genetic testing improves patient-centered health outcomes? Note, for each letter outcome type, assign a number from one to five to indicate your vote. A lower number indicates lower confidence; a higher number indicates higher confidence.

4-A, changes in physician directed patient management. 4-B, indirect or intermediate healthcare outcomes; for example, changes in laboratory test results such as hemoglobin or time to achieve a target value. 4-C, direct patient centered healthcare outcomes; for example, mortality, functional status, adverse events.

Question five, are there ethical issues particular to genetic testing that may alter the methodological rigor -- I can say this word any other day of the week -- of studies of genetic testing. Please discuss the existence, relevance, and impact of such issues.

Question six, does the age of the Medicare beneficiary population present particular challenges that may compromise the generation and/or interpretation of evidence regarding genetic testing. Please discuss the existence, relevance, and impact of such issues.

Now I would just like to bring Jeff Roche up. He's a Board certified pathologist. He's also worked in public health, epidemiology, and he joins CMS as a medical officer here in 2007.

MR. ROCHE: Good morning everyone. I hope to provide very briefly some context for the questions that you've been asked to review today. Many of us are aware, because we've used them for many, many years, of what diagnostic tests do. They provide

information, and they help physicians make decisions about patient care.

Now, there are different aspects of various types of diagnostic tests. Some of those tests identify the illness itself. Some of them give us an idea of the extent or burden or likely future course of an illness, and still others help us to assess the best therapy for a patient.

There are many tests available for diagnostic use. This slide just lists a few of them, and each of them answers a question, a question that helps a physician decide on patient care. But the way that physicians decide what the meaning of a lab test is and in many ways the value of a laboratory test in a system of coverage is its effect on being able to provide good information, and information quality is assessed by evidence.

This is an example of a clinical study which has looked at the quality of evidence about, in this case, proteinuria as a potential prognostic indicator for patients with type 2 diabetes melitis.

As you can see from this slide, and I hope this is visible in the audience, the group with the highest level of proteinuria, the group with the curve toward the bottom of the slide, that's the group with the highest mortality.

And this association between the test result and its clinical implications is very key to our ability to analyze evidence about diagnostic tests of whatever type.

Now, we draw from a number of other sources as well when we look at a diagnostic test. Some of them are listed here. And although we recognize that genetic testing raises new questions that challenge us to some degree about testing and what it means and its value for the patient, we know that these kinds of tests are here, they're here to stay, and that many laboratories, as shown in this slide from a survey of more than 800 laboratories in Europe, North America and Asia -- this was taken, by the way, in 2003 -- indicate that these tests are widespread. They're being used in clinical medicine, and our challenge is

to look at how they will help in patient care.

We are not alone in this. A number of groups are looking at the evidence for the value of genetic testing. Some of those groups are listed on this slide, and they have chosen to look very carefully at specific uses of genetic testing in particular clinical applications.

Some of the evidence they use, for example, is shown on this graphic. This is a graphic study of patients with lymph node negative estrogen receptor positive breast cancer. And these patients were tested for the expression of certain groups of genes.

For example, at the top of the slide, and it may be in somewhat small print for folks, but it's the proliferation gene group. Notice that the increased expression of the genes in this group is linked to a higher relative risk of mortality. Again, the type of evidence that we would look for to look at the value of genetic testing.

Now, many groups are looking at the value of genetic testing. For example, the EGAPP group from

the CDC has published not only specific evidence reviews about particular uses of genetic testing but also about a more general framework based in part on earlier work which helps us look in a general way at the value of genetic testing for diagnostic use.

Others have helped physicians individually decide if genetic testing might be worthwhile for their individual patients by providing a series of questions similar to those used in evidence-based medicine to help assess the value of the test for that patient.

In addition, we note that professional societies are increasingly looking at evidence also to find out if genetic testing might be valuable for physicians in practice. The American Society of Clinical Oncology, for example, the American Society of Chest Medicine has also been looking at that.

The value of such general approaches is -- although frankly they're not easy to form into fashion, they can lead to a clear and consistent approach to looking at new evidence about genetic

testing. They tend to favor clinical study designs which minimize bias and which focus on improved patient outcomes.

And finally they can be examined, debated, refined, and hopefully over the years become standards for clinical care.

As I say, I would like to wrap up this very brief presentation by thanking many organizations in government who have helped extend and explore the genetic testing world, especially the EGAPP group at CDC, the technology assessment program at the Agency for Healthcare Research and Quality, the online Mendelian Inheritance in Man site sponsored by the National Institutes of Health, as well as sites like gene reviews and genedesk.org, as well as the many activities of the National Institute of Health, which are trying to extend our knowledge of this important subject. Thank you.

DR. MC NEIL: Thank you very much, Dr. Roche. All right. Are there any questions or clarification for Dr. Roche? If not, we will move on

and presume that you'll be around if other questions arise?

Great. Okay, so let's move on to our next presenter, Dr. Thomas Trikalinos from Tufts-New England Medical Center -- there you are -- who is going to give us -- we have his slides as well as some general facts and considerations.

DR. TRIKALINOS: Good morning. My name is Tom Trikalinos, and I'm going to present you a technology assessment that was performed by the Tufts Evidence Based Practice Center. This technology assessment was conducted by a group that has no conflicts of interest and it was funded -- it was commissioned by CMS and funded through the AHRQ program.

The technology assessment is a compilation of four systematic reviews and pharmacogenetic tests for cancer and noncancer conditions. Let me give you a starting point for this technology assessment and how it came about.

Tufts EPC has performed several horizon

scans. Horizon scans are broad views of the literature to identify what status exists out there and what kinds of features are being addressed for specific topics.

A 2006 horizon scan that we did focused on identifying genetic tests on cancer conditions. And we identified in that report 62 of them according to criteria that are specified there.

In 2007 we performed another report that focused on genetic tests, not only pharmacogenetic tests but genetic tests in general, and for noncancer conditions, we identified several hundred, but 91 of them were relevant perhaps the Medicare population.

So this was the starting point, the starting pool of genetic tests from which it was convenient to sample to perform -- to select topics for this review.

The inclusion criteria for tests to use in this case were -- we were trying to find examples. The aim was to find examples of tests that were highlighted of being relevant to the Medicare beneficiary population, and they could also serve as

general examples to stimulate discussions for a framework for the general evaluation of genetic tests.

Because of the nature of this report, tests that were already reviewed or tests that had been -- were addressed at ongoing reviews by AHRQ, CMS, or other teams were not considered. And also it was decided not to tackle tests that had a very, very large evidentiary basis for practicalities.

So after a series of iterations between AHRQ, CMS, and Tufts EPC that performed vetting -- pre-vetting of several topics, it was decided that the report should address four specific pharmacogenetic tests, CYP2C9 and warfarin therapy. CYP2C9 is Cytochrome P450 family 2, subfamily C, polypeptide 9. VKORC1 and warfarin therapy. VKORC1 is short for the Vitamin K epoxide reductase protein subunit one. APOE and statin therapy. APOE is apoliproprotein E gene, NTH a foreign chemotherapy of the folate metabolic pathway.

Common variations in these genes and their role as pharmacogenetic tests in this specific

context. As you can understand, the framing of -- the selection of the topics does not necessarily mean that these tests are used in everyday clinical practice, but that it could be conceived as pharmacogenetic tests.

The report has a compilation of four different topics, and this is an outline -- I did an outline of the key questions that are addressed in the four different topics just to show you how they mapped in parallel. So I'm going to show you the exact questions in the next slides.

But the first addresses associations between the gene types in the population and clinical or biochemical outcomes.

The second key question pertains only to the warfarin example, and it's essentially a breakdown of the first question, but essentially it's again associations with what we call their adverse outcomes.

Question three asks whether there are modifying factors that modify the strength of these associations, and question four goes beyond the

associations and asks whether actually testing affects therapeutic decisions, and question five addresses benefits and harms of testing versus no testing.

So the actual key question one would flow like this among patients who take one of the three drugs -- one of the three treatments, is there any association between common genetic variations, and these happen to be all SNPs, single-nucleotide polymorphisms, for the four genes and clinical variables. So clinical variables where maintenance dose for warfarin, clinical outcomes for the warfarin example coming to the key question two, as I mapped them here, CVD mortality or mortality for statins and chemo respectively, or biochemical variables.

Key question two is essentially as I told you. I mapped it here as a separate question. But essentially it assesses associations for warfarin, for the warfarin topic between common genetic variations in these two genes and adverse outcomes. And here we are talking about clinical outcomes that are of interest to this particular topic, and they have to do

with bleeding and thromboembolism.

Key questions three and four. What are -- three is what are demographic or clinical variables that mediate the association between pharmacogenetic test results and biochemical or clinical outcomes among patient who use the three treatments?

Four, how does a pharmacogenetic test result affect the decision to use treatments? And five, what are the benefits, harms, or adverse effects that are experienced by patients who are on the therapies and who have been managed based on the test result. Let me go back.

So, as you can see, these key questions are not the same key questions that the MEDCAC panel will answer today. And this report -- I'm presenting this report in a way that tries to lift specific insights that we gained during this process that may be of relevance to the choice of a framework and perhaps highlight some issues that we came across.

The EPC did not perform a systematic review on the specific key questions that you are going to

answer today.

So we are talking about diagnostic tests in general, and I'm going to talk about genetic tests in the next slide. But diagnostic tests in general, when we are studying them, we need often a framework to contextualize the evidence and to help us interpret the evidence that exists.

Several frameworks have been proposed for the evaluation of diagnostic tests in general, and one of them that's very well known is shown on the slide.

Recent systematic review of different frameworks for diagnostic tests has found 17 or 18 different frameworks that are specifically tailored to different topics and different aspects.

However, many of them follow similar rationale as displayed on the screen, and if I'm not mistaken, even people who are among the panel have proposed frameworks that look very much like this.

So we usually refer to this as the Fryback and Thornberry framework for the assessment of diagnostic tests, and it's, in reality, very, very

applicable to imaging tests and diagnostic tests that are not genetic tests. However, the general framework is there.

So Fryback and Thornbury discuss that you can map the available evidence into six levels, and you can use these levels to help you interpret your findings.

Level 1 has to do with the technical feasibility and essentially asks the question of whether the test actually measures what it's supposed to measure. Does the test perform reliably and deliver accurate information? Here you have all these things like repeatability, reproducibility, and other terms that mean different things but sound alike.

Level 2 asks information about test performance and test accuracy. Does the test contribute to making an accurate diagnosis? All the usual studies that talk about sensitivity, specificity, post-predictive values and whatnot, would fall into this category.

Afterwards, the next step is how does a test

affect diagnostic thinking -- what is the diagnostic input of the test. So does the test result change subsequent diagnostic workup that the physician wants to do? How much does it inform to change diagnostic thinking?

Next is the therapeutic impact level. How much does the test result influence the selection among alternative therapies.

And Level 5 goes to clinical outcomes and it essentially measures how much impact you have on patients' health by using versus not using a test and managing the patient with the test or without.

Level 6 zooms out and talks about the societal impact, and here is where you usually would find cost-effectiveness analysis and the discussion of ethical issues and the like.

This is a general framework that is particularly suitable to identify -- to study tests like CT scan and other imaging tests, but it's not very, very well suited to study genetic tests.

So there are variations to study genetic

tests that follow more or less the same process but they are specifically tailored for genetic tests. This is the ACCE evaluation process for genetic tests that is adopted by EGAPP -- that has been adopted by EGAPP and essentially is a framework that breaks down evidence into four components, analytic validity, clinical validity, clinical utility, and ethicolegal issues.

This framework is a very detailed framework that has 44 different questions that have to be addressed in order to cover the whole spectrum.

Let me just guide you through it very, very briefly. At the center you have the disorder and a setting that's being studied.

Analytic validity means does the genetic test measure exactly the thing that it is supposed to measure, and here you would have the ability of genotype in technology let's say to identify the same genetic variations as the gold standard, like by the directional sequencing or what have you. And it has analytic sensitivity, specificity. It talks about

quality control and it examines the ACCEs that have been used.

This -- you can see much to level 1 in the Fryback classification. So it's like the technical visibility in the Fryback classification.

The next step is clinical validity, which is the first C. And clinical validity is essentially Level 2 of the Fryback classification, if you want to see it this way, and it talks about whether the test actually -- whether people who have differences in their genotypes or in their expression profile or whatever, have different likelihood for disease or no disease for the target condition.

And here there are six items that are being evaluated, clinical sensitivity, clinical specificity, positive and negative predicted values, penetrance, which is the old genetic term for measures like odds ratio and strength of association and prevalence of the disease.

The next step is clinical utility and have many dimensions, and clinical utility measures

essentially how does the test affect patient outcomes, not only for the patient but also for the patient's relatives in the genetic tests and in the ACCE framework, and they have a very, very comprehensive list of things that have to be assessed.

This we could say correspond to levels 3, 4 and 5 of the Fryback framework, and then you've got the cross issues of ethical and legal and social considerations which correspond roughly to the level 6 of the Fryback classification.

In the specific report, as I told you, we do not address the same key questions as you do, so I'm going to present to you some insights from the four topics. I'm not going to present to you the results in any detail to answer the specific questions.

The last search date of our literature search was September in 2007. So the numbers that you may see in the number of studies may not include -- may not include recent ones.

So CYP2C9 and warfarin, VKORC1 and warfarin, what do these two have to do with warfarin? Common

variations in the CYP2C9 gene affect the metabolism of warfarin. These variations were usually referred to as star 2 and star 3. Now they have different names, but these names persist in the literature and this is why they are listed like this.

Affecting the metabolism of warfarin, they confer sensitivity to warfarin doses. For the VKORC1, the explanation of the postulated mechanism is more complicated and has to do with pharmacokinetics, but an intuitive way to go about it is that there are some rare mutations in VKORC1 that confer susceptibility to a rare condition that's called resistance to warfarin, and this is a Mendelian disorder. This is not a common disease. So it was postulated that common variations in the same gene could affect the sensitivity to warfarin.

Warfarin is a blood thinner, as you know, and it's being used when you want to control the coagulation system of a patient, and warfarin has a narrow therapeutic index.

If you give too much of the blood thinner,

you risk having hemorrhages and hemorrhagic complications. If you give too little, then you don't attain the necessary anticoagulation stages, so you risk thromboembolisms.

Most people -- well, not most people. The average dose -- and an average dose for warfarin would be 5 milligrams per day. And most people -- the average person, let's say, would fall there. However, there are people who need much smaller doses, and there are people who need much larger doses. And the variation in the dose among the extremes could be quite large. Typically it's said that it could be like ten times between the two extremes, and there have been documented cases, as in a very recent paper in the New England Journal of Medicine, where the extremes are very, very far apart.

You see that we -- the common variations in CYP2C9 and VKORC1, they would confer sensitivity to warfarin and it would change the metabolism of warfarin and the anti-coagulation cascade in such a way that the patients would need less drug. So if you

give them the usual dose, you risk over anticoagulating.

So this is just an overview of the CYP2C9 and VKORC1 findings. As I will tell you, we have strong evidence of associations with surrogate outcomes like the maintenance dose of warfarin. We have unclear evidence for associations with bleeding or thromboembolic events.

Now, this talks about associations, about observed relationships of these outcomes with the general types of the patients. They do not talk about whether using or not using the test affects let's say bleeding or thromboembolism.

And we found no study that measured the affects of testing on patients relevant clinical outcomes.

So there is -- I'm going to give you examples now. So there is an abundance of associations with surrogates. This is a good enough example to talk about.

Associations with mean dose. Mean dose is

the mean stable dose of warfarin. When you start a patient on warfarin you start with a starting dose and then you measure their INR, their international normalized ratio, which is a measurement that you can use to titrate your doses to obtain the anti-coagulation status that you want.

And after you titrate your doses, you reach a stable dose, and this is the maintenance dose, crudely speaking.

So there are a lot of studies out there that evaluate whether people who have different genetic background, different genetic profiles for CYP2C9, here is the example for star 2, have differences in their mean maintenance dose.

This is a typical meta analysis plot, and I'm going to briefly guide you through it. Our outcome is differences in the mean maintenance dose, and these are studies that essentially ask this question across a range of populations and across a range of settings.

In a meta analysis, the premise is that if

you have a series of studies that address the same question and they have an epidemiological, biological and clinical cohesion, you can use all of them to gain a better estimate, an overall grand mean that is more informative about the strength of the association.

A meta analysis lines up all the studies and represents each one of them by a square, a point estimate, and a horizontal line. The square is the measurement of the affect, and here it's the difference in the maintenance doses between the two genotypic groups.

The two genotypic groups are carriers of CYP2C9, star 2, versus non-carriers, and these would be an assumption of a dominant inheritance model called dominant recessive, extreme homozygote, other things can be done. This is just an example.

Zero is the line of no effect, and it means no difference between the genotypic groups. Its studies represented, as I said, by the point estimate that says what's the difference in the study and the confidence interval which is a horizontal line. The

confidence interval quantifies the uncertainty that accompanies the studies estimate. Large confidence intervals, you get large uncertainty. Smaller confidence intervals, smaller uncertainty and usually bigger studies.

You can see that overall the meta analysis result is shown in the small diamond in the end, and really there is no doubt in my mind that there is a very, very strong association between the genotypic groups and differences in mean warfarin dose. In fact, we have a lot of studies, and perhaps much more than needed to answer this kind of question.

However, we have less data in associations with clinical outcomes. And again, I'm talking again about associations and not evaluations of applying the tests.

Clinical outcomes in the warfarin case would be bleeding, and this is again an example, associations with major bleeding. We found nine studies that anyhow mentioned bleeding. Bleeding was defined in very different ways across the status, and

five of them somehow defined major bleeding as significant bleeding, and all five used different exact definitions, but you can see here that essentially all of the data that we have are from five small studies. Three of them found no events in either arm so they essentially do not contribute to the odd ratio in the meta analysis. And we've got some results that are in the expected direction. That is that if you have the two allele or the three allele, you are more likely to experience a bleeding event through the mechanism that salt views. But this kind of meta analysis is sparse. And this kind of meta analysis is not based on our best data. And although I am willing to believe this kind of meta analysis, I cannot say that it's based on a lot of data.

However, as I told you before, we did not find studies that informed on testing versus no testing with respect to patients' relevant clinical outcomes.

Another interesting thing is that meta

analysis is a retrospective exercise. You try to analyze published evidence and you try to analyze evidence that have been presented by other people. It's evidence that's out there.

So in fields like genetic epidemiology and pharmacogenetic tests here that are quite prolific, and where you have a plethora of different -- of different genetic variants and a plethora also of associations. There are many, many different combinations that one could test. So you'll see what I mean. This is an example from the VKORC1 systematic review, and this is a sparse matrix of outcomes and genetic factors. And I will guide you through the matrix.

The red boxes define super columns that define different outcomes. This is a different outcome, INR less than 2, INR more than three and so on. You can see that all of these are surrogate outcomes.

The thin columns within each red box, within each super column, are different studies. To be

accurate, we identified 19 studies, but these 19 studies mapped to 23 different ethnic dissent strata. So I'm showing you with the thin columns are the different populations.

When we talked about CYP2C9 we said that we were asking about specific genetic variants, star 2, star 3. Here we did not have a sense of which common genetic variants to study, so we mapped out what had been done in the literature till then.

And you see in the published literature, there are many, many genetic variations, many, many SNPs, and many, many more frankly in VKORC1 or its promoter. Each one of them corresponds to a horizontal line.

Whenever you have a study that assessed the specific bare, the specific association of the genetic factor, let's say -- let me go there. The last genetic factor and INR2 has been studied in the field -- in the stratum that corresponds to the field cell.

The color coding of the cells has to do with the direction of the effects and whether or not they

were statistically significant, or whether or not they are actually described on whether they were significant or not.

What you can see is that this matrix is very sparse. Most of the studies talk about associations with mean dose, and most of the studies have to do with associations with mean dose for a specific set of SNPs that belong to the same LD block, linkage disequilibrium block. They belong to the same genetic region. I mean that they tend to go together. And they have been studied more than the other ones.

There were no studies for the clinical outcome. So if I show you the whole matrix and put the clinical outcomes, you will see that it's completely empty. So you've got a sparse matrix and essentially all the meta analysis that you can do, search under the lamp post. Search wherever you have data and search wherever you have associations, and they can talk about the things that you can see there.

This has been well recognized in prolific fields, and this has been very well recognized in

genetic epidemiology for associations with disease in particular, and this is why we now have movements of a consortia of investigators that tackle specific associations in perspective meta analysis.

So if I tried to map these things onto the ACCE table, it would go like this. We did not address analytic validity. It was not in the key questions, and also it was not in the key questions to address ethicolegal issues.

But analytic validity I don't think that it would be very meaningful to address because all of the studies used pretty standard methodologies to get to genetic variations that are not very difficult to get to. These things typically have accuracy of 99.9 percent.

Clinical validity would get all of the studies that we talked about. Now, this is a wrong -- this had to be six. You see that in total there are 29 different studies for CYP2C9, and here we've got associations with mean dose, associations with INR levels. A usual therapeutic threshold for INR, a

therapeutic target range is to have an INR between two and three. A normal person has between one and one forty. So INR less than two means less coagulation than needed. INR more than three means more coagulation than needed, and this is the kind of outcomes that people have used there, surrogate outcomes.

And also associations with clinical outcomes, expected clinical outcomes. But not really much that assesses how much the test would impact on patient outcomes if it were applied versus not. In the clinical utility section you may map one RCT that was identified during the peer review process, and it was added as an addendum that talked about measured changes -- that measured treatment changes. There is another RCT out and there is another bigger RCT that's being currently conducted on this specific topic, and we are waiting for these examples. But you see how this thing would map on the ACCE table were we to do the ACCE framework.

APOE studies and MTHFR and chemo. APOE

studies, why study APOE studies where there's a lot of complicated mechanisms that I really cannot tell you in a good way, so I'll just leave it like this.

MTHFR, MTHFR is an enzyme that affects the metabolism of folate, and there are some chemotherapeutic agents like methotrexate or 5-fluorouracil that are competitive antagonists of the folate pathway. So folate is very necessary for DNA synthesis and this is the way that these things act. I'll just cut to the chase.

You see that if you try to map the specific pharmacogenetic tests onto the ACCE framework, again, we did not assess the analytic validity and the ethicolegal issues, and again you can see that the associations, whether people who have differences in the genotypes experience an event or not, the associations map like this.

This specific figuring reads a bit funny because you would not really use -- well, who knows -- APOE for a predicted mortality, but this is how it maps under a definition for genetic tests according to

the Secretary -- according to the Secretary's definition, this would go this way.

Again, nothing on clinical utility.

So majority of status, 29, 19, 44, and 11 had to do with clinical validity, and something could be said about clinical utility in one of the topics for the time period that the systematic reviews included.

Now, this is not unique to genetic tests. This is something that is very well appreciated, how different studies map to the different levels, and this is for example evident from a 2005 paper that Athena Tetzioni (phonetic) did in the annals. You see studies about magnetic resonance spectroscopy for brain tumors, how they map in the different levels.

Most of the studies have to do with technical visibility, that's MR produced consistent spectra and whatnot. And you see that very few go to the therapeutic impact, and nothing was identified for the other levels.

And I can tell you that we see the same

thing in a systematic review that we are doing for PETs and lymphomas.

So a bird eye view. The value of every test, and I think the value of pharmacogenetic or genetic tests would measure the same way, is judged by the ability of the test to affect patient relevant outcomes.

This is not easy to diffuse because you have -- you apply the test to a target population, and based on the test, you have to act. If the test does not convey actionable information, then you would not have an impact on patient outcomes in the majority of cases.

So the ability of the test to affect patient relevant outcomes is the real overarching question on which you would judge or on which I would judge the usefulness of a test. However the problem is that different studies evaluate the accuracy of a test. Different studies evaluate the sensitivity, specificity, its operating characteristics, and different studies evaluate the relationship between

the treatments that may affect patient outcomes.

So you have to bring together in a suitable way evidence from different studies, different pieces of the puzzle to answer the overarching question. And I'm not sure that you can always do it in your head.

And finally, there are many, many, many framework-related issues that the report does not inform one and cannot inform one. So these issues have to do with the purpose of the test. When you do a test, you have to define what its purpose is.

Now it has been clarified that we are not talking about screening tests, but in general tests could be put into buckets of screening, diagnostics, prognosis, patient monitoring and treatment guidance. Their support bucket that says imaging and help, for example, for planning a surgery, but it's not of interest here.

Also the role of a new test. If you have a patient management strategy in place and you want to add a new test or you want to add -- you want to insert a test or modify something, you have to realize

is this a replacement test. Is the new test an add-on test. Is the new test a triage test and so on.

Tests focusing on common versus rare variants and tests for common versus rare diseases. Rare diseases and common diseases are two completely different beasts. Common diseases are polygenic diseases, and the common gene common variant hypothesis says that at least for genetic associations of disease, for a common disease like cancer or a cardiovascular disease, you expect very, very small contributions for a large number of genetic factors, rather than a single genetic factor being the case.

In this sense, the cancer or the majority of, let's say, heart disease, the big bulk of heart disease, is not the same as (inaudible) disease or any other disease. That's Mendelian and monogenic.

And there's also a variety of issues that have to do with multipanel versus single panel tests and whatnot.

And I think that this is where I will finish. Thank you.

DR. MC NEIL: Thank you very much. Let's see. Why don't we ask for points of clarification for this particular presentation. I'd actually start with one. Why don't you stay up there?

Did you indicate that the randomized trial on the effect of transfusion -- effect of transfusions for patients on warfarin, those data are not yet in? Is that correct?

DR. TRIKALINOS: Yes.

DR. MC NEIL: When will they be available?

DR. TRIKALINOS: I don't know that.

DR. MC NEIL: Okay. Okay. Thank you.

Yes, Cliff.

DR. GOODMAN: I just want to confirm that the study you presented was based on a selected sample of tests, that you excluded certain tests simply because there was too much evidence to cope with during the time that you were given to do your analysis. Correct?

DR. TRIKALINOS: This is correct. As I said, these tests were selected in iterations and

discussions between AHRQ and CMS, and we helped them vett some topics. So we vetted some topics, and if they were very big, they decided that they were not to be tackled in the specific report.

DR. GOODMAN: So your findings, which are fascinating, are not necessarily applicable to the whole body of tests. These are sort of case examples?

DR. TRIKALINOS: I cannot comment on this from the specific reports, but if you ask my personal opinion, it's very likely that many of these issues are cross cutting for common disease, common variants.

DR. GOODMAN: Okay. Now, toward the end, you had the pieces of the puzzle slide where you said that evidence from different studies has to be brought together to answer the overarching question. That may very well be the case in quite a few instances, but would you allow that it would be possible in some cases, and even desirable, though perhaps costly and time-consuming, to randomize patients to get -- randomize patients in a target population to get a test, follow them longitudinally all the way to

patient outcome. That's possible and it might have to -- the feasibility of doing that might have to do with the natural history of the disease or the episode of care.

So I just wanted to suggest that it's not necessarily the case that evidence would have to be pieced together. It could be, in some instances, done in a single RCT, however difficult or time-consuming.

DR. TRIKALINOS: This is correct. So you touched upon two different issues. Any test, in theory, could be tested with an RCT, and an RCT would give you the desired results. It's just a feasibility, and other issues come into play because if you compare diagnostic strategies in an RCT, you may have to do a lot of -- there are many, many alternatives that change, the timeframe and whatnot.

But I didn't say that RCTs are necessary in each case.

DR. PHURROUGH: Perhaps I need to just clarify. We did not do this well perhaps at the beginning.

We have an arrangement where we are getting on a every-other-year basis an updated technology assessment on the status of genetic testing.

One year it's genetic testing in cancer-related subjects and the next year is genetic testing on non-cancer related issues.

And what we've asked Tufts to do and Tom to do today is from those TAs to draw conclusions around what kind of evidence is out there on genetic testing.

So we did not specifically ask for this particular MEDCAC for someone to review some evidence and say how are these questions answered. We've asked Tufts to give us -- to draw some conclusions around a TA that is not specifically related to this but is on genetic testing in general.

DR. MC NEIL: Mark?

DR. GRANT: Thanks, Tom.

You made a comment on the slide, pieces of the puzzle, if I remember correctly. I'm not sure how often one can do this in your head since you're referring to an indirect evidence.

And my question for you is how often is not often. Do you think people can do that kind of indirect synthesis in their head?

DR. TRIKALINOS: Okay. When evaluating different tests and strategies -- the short answer is that it's likely that it will not be often.

When evaluating different test strategies, you have to take into account a lot of things, and you have to take them into account in a structured framework that would allow you let's say to compare different alternatives, that would allow you to distinguish -- to exemplify your assumptions, that would allow you to distinguish choices from chances, weigh the different likelihoods of risks, benefits, and why not costs, although I should not say this word, and make informed decisions.

And these calculations may be more difficult than one may think.

DR. MC NEIL: All right. I have several people with questions so we'll have, if we could, brief questions and brief responses.

So I have Jim, Eleanor, and Marion. Okay. Neil?

DR. PUKLIN: Yes. I want to thank you for such a lovely presentation, but I just want to make sure I understood the answer to the first question.

The first question I think dealt with the fact that you excluded a large number of published studies in the literature dealing with the same issues simply because time didn't permit this evidence-based review to be completely inclusive.

And your comment and answer was that you would assume that all of these other studies would have the same sort of information in them as you presented on the studies that you were able to evaluate. Is that a correct assumption?

DR. TRIKALINOS: I said that this is my personal -- my personal thought.

DR. PUKLIN: Right. Okay.

DR. MC NEIL: Eleanor?

DR. PERFETTO: Yeah. My question is related. Can you give us some examples of some --

DR. MC NEIL: Louder.

DR. PERFETTO: Can you give us some examples of some of the things that got left out? What had such a large bulk of literature that we might be missing something from that?

DR. TRIKALINOS: For example, say TP53 and head and neck cancers, TP53 and gastrointestinal cancers, HER2neu and all of these things.

There are many, many tests that could be conceived. May I underline this. I use the term tests, according to the Secretary's definition, that says that any genetic variation or any -- I do not remember it verbatim, but these would fall under tests.

Many of these studies are essentially prognostic marker studies, so they use a genetic variation or they use the expression on a slide -- on a pathology slide to associate with disease.

So they would fall into the frameworks of ACCE as studies evaluating tests, specifically diagnostic accuracy.

DR. PERFETTO: But would you say you dropped ten of these or twenty of these? I mean how many of them were there that had that bulk of literature that you put them aside?

DR. TRIKALINOS: I cannot give you an accurate answer to this, but the big pool was this 62 and 91 that were identified. Perhaps Gary Rahman (phonetic) who is the lead of this report could be able to answer this. I am sorry. I cannot give you this answer.

DR. MC NEIL: Maybe you could make a call to him during the break and get an answer?

DR. TRIKALINOS: I could try.

DR. MC NEIL: That would be great. Okay. Let's see. Marion?

DR. DANIS: In the meta-analysis that you illustrated showing the studies of association between genetic variance and INR, whether many of those studies included elderly patients or exclusively had any elderly populations?

DR. TRIKALINOS: If I recall correctly, the

mean ages of the -- the status included elderly patients. The mean ages of these populations were -- if I recall correctly, were not far away from what you would call the Medicare beneficiary population. If you ask about the potential applicability.

DR. MC NEIL: Let's see. Neil?

DR. HOLTZMAN: I'd like to comment about, I guess, your last slide where you talk about the role of tests. For instance, one of the examples on the slide was replacement, that you'd use a test to replace previous efforts to make a diagnosis, if that's what you're doing. And I want to draw an analogy to drug evaluation 'cause essentially, the essence of drug evaluation these days is to compare a new drug, not to placebo, but to the existing or the current means of diagnosis. And it seems to me right now that we're nowhere near from looking at your slides doing that sort of thing. It's the piece that's missing.

And I wonder if you could comment on that and the importance of knowing whether we're really

getting a significant increment in information to make a correct diagnosis or prognosis by the use of genetic tests compared to the current means of evaluation and how we can further evaluate on that level.

DR. TRIKALINOS: Okay. So there are several issues. First, there -- even if you have a simple case of two tests and a simple treatment, there are many, many different combinations that you could study. Test one, then test two, tests one and two at the same time, only one, only the other, one triaging the other, one replicating the other, and so on. All these different combinations may be legit, or some of them may be legit in a specific context. It's very -- it's very difficult in many cases to design a study that would -- that would answer all these questions. That would randomize, let's say, people to all the different combinations and measure their outcomes.

So it may be unavoidable for many cases to have the evidence the way that it is mapped here. So the -- I'm sorry. I lost the second one.

DR. HOLTZMAN: Well, the question is, I

mean, physicians using warfarin for a long time have used various things to titrate their patients to get the appropriate dosage. And my question simply is, how do we find out whether whatever genetic tests are available are going to incrementally improve the ability to place the patient on a correct dose.

DR. TRIKALINOS: Well, really the answer to this question is very complicated, and it's essentially what you would have to find out today. I don't have a good answer. I only have general answers. And I don't think that I can give you a very good one.

DR. HOLTZMAN: Well, it just seems to me that everything that you've said is very important as a preliminary to get to that stage. But for instance, do we randomize trials where you're looking -- and you talked about this in your paper -- standardized means of assessment versus standardized plus genetic. And we're not there yet. I mean, you have a zero in many of those cells.

DR. TRIKALINOS: Okay. So in many cases --

in some selected -- this has been studied. Whether you have to use a randomized trial to evaluate a test, this has been studied. Patrick Pasut's (phonetic) group has written extensively on this and you can track these papers in the literature. The thing is that in most cases -- in very selected cases, you don't really need to do a randomized trial. But this is not going to be the majority. So I'm not discussing this -- when it would be self-evident. When you have to do something to measure it, often it's infeasible to do a randomized trial. And I alluded to the solution of having to make calculations.

And this I can - I can say it perhaps. My personal thinking is that for some cases you have to do what is called a decision analysis. It doesn't mean that a decision analysis is needed always. I believe that it has to be judged on a case by case basis. But now I'm giving you my personal opinions and not something that's in the report.

DR. MC NEIL: Why don't we stick to the

report for this particular part of the discussion. Steve?

DR. PEARSON: Tom, thank you. I thought it was a very good conceptual overview, particularly the linkage that you made between the Fryback and Thornbury levels to the ACCE categories. And I wanted your opinion about one feature of the clinical validity box when you mapped some of the results. What's in that box are associations, not sensitivity or specificity. And in general, associations are an average, obviously, linking, you know, a test finding with an average kind of variation in a certain clinical or biometric outcome.

But it doesn't necessarily indicate that individual patients who might have a, quote, unquote, "positive test," might, in fact, have a physiological or a phenotypic, you know, outcome that doesn't match the test. That's kind of blurred, if you understand my thinking, in this association. So if we're thinking about truly looking at the potential risks and benefits, what would you say about the lack of

sensitivity and specificity data and just these associations when you come to clinical validity considerations?

DR. TRIKALINOS: You would have to tell me sensitivity and specificity to do what. So in the -- let's say in the chemo example, sensitivity and specificity to predict to a year mortality down the road. Instead of quantifying this confusion matrix as a sensitivity, specificity, you can do it as a threshold diagnostic. The threshold would be the equivalent. I think that in many cases, the associations for binary outcomes -- the associations could be tilted towards the sensitivity/specificity paradigm if you were willing to see the whole process of testing someone or genotyping someone as a diagnostic test.

DR. PEARSON: Just real quickly, a specific example. So for the warfarin testing, we can't tell from these associations how many patients out of a hundred who have a positive genetic variant do not show physiologically what we would expect given that

association. Is that correct?

DR. TRIKALINOS: I agree. Absolutely. Absolutely.

DR. MC NEIL: I wonder if I could ask one question. And it may be a follow-up on the question that Neil asked you. All of your studies, the ones that are both included and excluded basically are meta-analyses, are syntheses of data from many studies. So if you look at the recent study on the KRAS wild type and the -- the KRAS mutational study for the effectiveness of treatment for colon cancer.

That was one study in the New England Journal that was a retrospective study in which they dug out the fact that patients did or did not respond depending upon the KRAS mutation. It got a lot of press. And a lot of people, a lot of evidence-based groups have reviewed those data. Would you feel comfortable using the results of that study to say that that particular test had a positive effect on patient -- on physician-directed therapy, which was one of the questions? So it's not a -- it's not a

meta-analysis. It's one study, retrospective, dig out the mutational status of the KRAS gene. How would you feel about our answering a 5 to question 5, I guess -- or 4?

DR. TRIKALINOS: So this would be the kind of studies that would be assembled in a meta-analysis.

DR. MC NEIL: No. I know that.

DR. TRIKALINOS: And I already told you that I would not necessarily interpret the meta-analysis of such studies, as an answer to this question. But this is --

DR. MC NEIL: No. That wasn't -- this is not a meta-analysis. It's one study.

DR. TRIKALINOS: We did not --

DR. MC NEIL: No, no. It's not what you did. I'm asking what you would think about one study. Forget the meta-analysis. One study.

DR. TRIKALINOS: I gave a complicated way to answer. No.

DR. MC NEIL: No, you wouldn't give it a five?

DR. TRIKALINOS: Don't ask me to give you -- to give a rate.

DR. MC NEIL: I can. Why not?

DR. PEARSON: Be strong, Tom.

DR. MC NEIL: Hang in there.

DR. TRIKALINOS: Okay. I will not give it a five.

DR. MC NEIL: You won't? Okay. Because it's one study?

DR. TRIKALINOS: Not because it's one study, but because it -- if I recall correctly, this did not measure patient outcomes in a way that says, do the test, do not do the test.

DR. MC NEIL: Well, it does address one of the questions that we have to answer, which is what is the effect of a genetic test on changes in therapeutic management.

DR. TRIKALINOS: Oh.

DR. MC NEIL: Forgetting the outcomes associated with the management. It is one of the questions. It's question 4. So it's not outcomes,

it's patient directed management. It's 4-A.

DR. PEARSON: No. It's 4-C.

DR. MC NEIL: No. It's 4-A. Physician-directed patient management. They will or will not now get a particular chemotherapeutic agent. And then they may or may not do well.

DR. TRIKALINOS: You're putting me on the spot.

DR. MC NEIL: There you are.

DR. TRIKALINOS: I will -- I will have to review the study and give you a better answer.

DR. MC NEIL: Okay. Actually, I wasn't picking on that particular study. I was trying to get at the concept of one study which had a very clean separation as I recall between patients who did and did not respond to a particular chemotherapeutic agent for a very clear-cut cancer with a very defined genetic mutation. But it was one study, and it was a retrospective study digging out the mutations. Is that -- do I not have that right? Who's that, Maren?

DR. SCHEUNER: Yeah. I mean, it's

retrospective only because they went back to the tumors that were --

DR. MC NEIL: I understand. I understand.

DR. SCHEUNER: Yeah. But I mean, the data were there. And in terms of the biases you might anticipate with retrospective studies with that particular issue. I don't know if they're the same.

DR. MC NEIL: Well, that's what I'm -- I'm just trying to get a sense. And maybe -- because it is going to be something that we're going to have to deal with later. To what extent -- because that looked like a very decent study to me. It was a very good clinical trial. They went back and got all the samples. They did all the genetic tests. They found a clear distinction between those who did and did not respond according to whether they did or did not have a mutation. A ton of publicity. Directing patient management right now is my understanding.

And the question I'm asking you is would you, on the basis of that one study, feel comfortable in having us say, aha, there is an example of one

mutation that gives us a 5 to question 4-A. And I think you said no. Okay. Got it.

DR. PHURROUGH: Perhaps if I could make one other observation, too. This is an example of one type of question that we're asking you to answer. And that is, is a particular trial of sufficient quality to draw conclusions around it -- around that particular question.

Can one trial, even if it's a -- is a single trial quality's sufficient to answer a question, versus the question that Neil was bringing up around the warfarin issue is we have a number of good trials that say that doing the warfarin sensitivity testing allows you to keep INRs within the appropriate range.

Let's make the assumption that the studies do that. There could be some argument about that. But let's make the assumption that we have good studies that say that doing the genetic testing for warfarin sensitivity allows you to maintain your INR in the appropriate range easier than if you did not have those studies.

Well, is that sufficient level of evidence -- even though it may be quality evidence, is it sufficient level of evidence to determine whether patients are better off, whether they have less bleeding or less thrombosis. And so that's the -- that's sort of the one slide that Tom had. You have these studies around tests. And to add to this example, there are additional studies, let's assume, that says if you keep your INR within a particular range -- you have less bleeding and thrombosis.

So can you put those two studies together. That's the question we're asking you. Can you put those two kinds of studies together and draw conclusions around patient care that would result in better outcomes, or do you need to have that specific study that says we're going to test -- we're going to find out what your warfarin sensitivity is genetically. And then we're going to see how much your bleeding and thrombosis is, not what your INR is, though you'll obviously measure. So that's the question we're asking that you'll get to later this

afternoon. What is -- what kinds of evidence do we need to be looking for?

DR. MC NEIL: Let's see. Maren, one quick question and then we're going to move on.

DR. SCHEUNER: I just had a question about the distinction between Mendelian disorders and common multi-factorial disorders. And as we consider the evidence necessary around our questions, do you have any comments? You alluded to the fact that Mendelian disorders were a different kettle of fish or what have you. So could you mention --

DR. TRIKALINOS: So my comment was motivated from observations from genetic epidemiology of common diseases. When you have a common disease like cancer or like heart disease, population genetics theory says -- predicts that there's not a single gene -- not a single variation that confers a lot of risk to have this disease. But there are many, many different variations, and each one of them has a different number of them in our genetic profile. And these build up our genetic profile that confers our overall

susceptibility.

And this is why common diseases are so difficult to -- such a difficult case to pinpoint specific genetic factors that are associated with them. Mind you, I'm not talking about diagnostic -- about pharmacogenetic tests now. You see genome-wide association studies, very, very large experiments that scan the whole genome with gazillions of markers and try to find an effect. Our latest results show effect sizes of 1.15, 1.20. Very, very small effects that are really, really, really small. And they're not easy to grasp. And it's not easy to see how you would utilize the scan of information.

In contrary, if you have a Mendelian disease that's monogenic, if you have the gene, it has a very, very high penetrance. It will lead to the disease. So you're more close to the usual paradigm of tests. You do a CT scan. You see something in the brain. It's something that's -- that has a high likelihood. It's something that informs you a lot on your downstream decisions. Does this answer?

DR. MC NEIL: It will have to answer for now. And maybe you two -- Maren, you can have a word with Tom during the coffee break. But I don't want to shortchange our other speakers. So if we could go on to Dr. Coates who's going to talk about EGAPP.

DR. COATES: Okay. I'd like to first thank you for this opportunity to speak about EGAPP in methods for assessing evidence for health improvements from the use of genomic tests. When we saw this announcement for the purpose of this meeting, we thought that it would be valuable potentially to share information on EGAPP because we think it offers a method to assess the evidence for determining the requirements with specific genomic tests.

EGAPP was developed specifically to address this question. Are genomic tests ready for use in clinical practice and in public health? EGAPP stands for Evaluation of Genomic Applications in Practice and Prevention. And its purpose was to establish and test a systematic evidence-based process for evaluating genetic tests and other applications by genomic

technology and the translation -- transition from research into practice. And to just say a point about Maren's question, this project is focusing not on single-gene disorders, not on rare diseases, but more trying to get into the role of and evaluate the role of genetic testing for common disorders, including common chronic diseases.

EGAPP evolved and was developed and shaped by a number of different international and U.S. meetings, including a 1994 Institute of Medicine report, several advisory panels that have been put together to advise the Secretary for Health and Human Services on genomics and genetics, including most recently, the Secretary's Advisory Committee on Genetics, Health, and Society. And the folks who were developing EGAPP also participated in international meetings and took into consideration reports on evaluation and quality assurance and quality assessment from those meetings as well.

EGAPP is a non-regulatory CDC-supported initiative. It's not in any way designed to do

oversight or regulation or take any functions from other parts of the federal government. It's put together specifically to develop a process for evaluation that's evidence-based, transparent, and publicly accountable, and to integrate and build on, and use existing processes for evaluation that have been developed by others. EGAPP works in part through the activities of three different groups that have been established and are supported by CDC. The most important is the independent multidisciplinary work group composed of non-federal experts that have actually developed the methods for EGAPP, oversee evidence reviews, and shape those, and then make recommendations based on those evidence reviews.

There's a steering committee of federal agencies to offer oversight for this activity and advise. And we have a stakeholder group that includes test developers and practitioners and others with an interest and stake in genomic testing for common disorders. The steering committee as I said represents -- has representatives from a number of

different federal agencies, including the critical players in these issues from CMS, FDA, and AHRQ.

The EGAPP approach uses common processes for developing evidence-based guidelines, published and transparent methods, systematic standardized evidence reviews using and evaluating a range of information.

One way that EGAPP is somewhat different than some other systematic review is that there is a lot of attention paid and attempt to find gray literature and to access information on tests that, often for laboratory-developed tests, is only available from the marketers. And there's -- for example, information that's available on the FDA website that may or may not have been incorporated into other kinds of reviews. It uses technical experts and stakeholders and consultants and reviewers, but not as decision-makers. There's peer review of the evidence reviews, the systematic evidence reviews, and recommendations by experts, agencies, and stakeholders. And there's a final evaluation and recommendations from this independent

work group, the independent panel, that's primarily from academia in an attempt to minimize conflicts of interest.

EGAPP's approach builds on methods from other processes. And I think you can see features that are common from AHRQ and FDA approaches and others, including the ACCE framework which was just presented. The methods for this process have just recently been published. This article was published in January. Steve Teutsch is the lead author. He's been a member of the U.S. Preventative Services Task Force for many years and was the chair of the Secretary's Advisory Committee on Genetic, Health, and Society. And we submitted a copy of this paper to you and the panel for your consideration in terms of looking at different methods to evaluate genomic tests.

The EGAPP evaluation method involves a careful, explicit definition of the disorder, the test, and the setting. The working group felt that it's very difficult to make generic statements about

genetic tests. They always -- the outcomes or the evaluations always going to depend on the specific uses of the test, which specific type of test is used, and the clinical setting. Is it in primary care, is it something that's marketed directly to the public over the internet, that sort of thing.

There's an evaluation of the accuracy and reliability in detecting genomic markers of interest, a focus on analytic validity, that's somewhat different from some other approaches to these reviews which sometimes will assume analytic validity. And actually, there in that current set of evidence reviews there are examples of tests for which there are reports on clinical validity, but there -- the information on analytic validity simply has not been published. It's not available.

There's an evaluation of accuracy and reliability in predicting the disorder or phenotype of interest, including the drug response. That's the clinical validity issue. And to follow up on some of the discussion that was raised by Tom's presentation,

this -- the EGAPP methods are evolving, and the work group is developing their approach to how to evaluate clinical validity based on the topic. And they selected a variety of different kinds of types of test to try to test and develop their methods.

But they're getting increasingly into looking for validation. Has a study finding been replicated? They're looking at calibration. If something's been found in one population, does it actually work prospectively in other populations? Attempts to synthesize the literature and information on sensitivity and specificity there under the receiving operator characteristic curve is not available, they ask the -- the working group asks the evidence reviewers to try to estimate that based on published information.

And more recently, they're getting into these issues of classification and risk prediction which have been raised by methodologists as important components in evaluation of tests. So there is an attempt to incorporate that kind of information when

it's available and to model it if it's not available. There's also an evaluation of the evidence of improved health outcomes and the utility in decision-making. And in the absence of information on health outcomes, they do focus a lot on it used in decision-making.

There's an assessment of contextual factors. And that again addresses one of the questions that's been raised in discussion so far. If the test is being proposed and it's proposed to replace or supplement some other kind of technology that's already available, does it really add value, or does it just add cost. And there's an overall assessment of the benefits and harms, taking into consideration all of these issues. The ethical, legal, and social issues are incorporated into this framework primarily in terms of assessing clinical utility and then the overall assessment of benefits and harms.

The first EGAPP recommendation that came from the working group was published a year ago. And it was on cytochrome P450 polymorphisms in adults with nonpsychotic depression treated with selected

serotonin reuptake inhibitors. And I want to give a little bit of an example for this. This topic was taken up because they thought -- the EGAPP working group was anticipating they'd make a positive recommendation for this test because there was published literature.

They also took it up because it's a common disorder. It certainly would be common potentially in a Medicare population. And there's a lot of interest in the test, and potentially, it could be widely used. And it's a condition that's -- an issue that's faced. Dosing and selection of the medications is a topic that's faced by primary care clinicians in most practices.

The recommendations were based on an evidence review that was done through the evidence based practice center program from AHRQ. So this is an example of the scenario. Patients newly diagnosed with depression are possibly treated with SSRIs. For example, Zoloft. The marketing of the genomic test says it may help with selection of particular

medications and in determining the doses of those medications to improve the effectiveness of it and to reduce side effects.

Patients and clinicians might be aware that there's an FDA-approved CYP450 genetic test available. And the question is should they use this test and how. This is an excerpt from an internet report. And let's see if this pointer works. So the idea of the test is that if you can -- to use this genetic to categorize people into -- by their metabolism status into poor or rapid or intermediate metabolizers, then you can make beginning dose recommendations based on the use of that test.

The EGAPP review developed this analytic framework. And this is the way all of the EGAPP reviews are done. And it addresses again another question that was raised in the discussion, that it's often not the case that there's going to be one source of information that answers the direct question that was posed originally. That information is going to have to be pieced together from a lot of different

sources to try to determine whether or not the test actually works.

So we begin with adults with nonpsychotic depression and then try to piece together a chain of evidence to determine whether or not there's improvement in terms of depression, quality of life, et cetera. So going through genotype testing, looking at harms and benefits of that metabolizer status treatment decisions. The evidence review found in terms of analytic validity that sensitivity and specificity appeared to be high. However, in looking at clinical validity, there was no consistent association between CYP450 genotype and drug levels, clinical response to SSRI treatment or adverse side effects.

It's one of those situations in which early research was quite promising. The patients were given large doses of some of the medications. And there was some segregation in terms of poor and rapid metabolizes. But follow-up studies using regular dosing found inconsistent results. With regard to

clinical utility, there were no studies that used genotyping to guide choice or dose and then studied subsequently the patient outcomes. So there -- and this again was a study where this topic was taken up because the working group thought that there was a lot of literature and that this might result in a positive recommendation.

Given the evidence review, the recommendation statement isn't unanticipated. There's insufficient evidence for a recommendation for or against use of testing. However, the working group said that in the absence of supporting evidence and with consideration of contextual issues, EGAPP discourages use until there's better information available, including clinical trials. And the reason -- the contextual information that they took into consideration are things like using these tests adds a cost to treatment that's going to be borne by patients or their insurance or whatever. And given the inability of the test to actually predict response to treatment and drug doses and that sort of thing, that

it might actually result in harms to patients.

Stakeholder response to EGAPP has generally been positive from the stakeholder group that has been assembled. And this is an example. The Secretary's Advisory Committee on Genetics, Health, and Society which represents a lot of different stakeholders said that HHS should create and fund a sustainable public/private entity of stakeholders to assess the clinical utility of genetic tests. For example, building on the EGAPP process.

There have been subsequently several additional EGAPP recommendations. These were just published in Genetics in Medicine, and we forwarded some of those to CMS for consideration. There are recommendations from the working on tumor gene expression profiling to improve outcomes in patients with breast cancer.

This is an example of a prognostic test that might help inform an approach to evaluations of prognostic tests. And both an evidence review and the recommendations, the overall evaluation of the

evidence by the EGAPP working group were available for this. This one, again, was a recommendation for insufficient evidence. The working group took up the question of UGT1A1 genotyping to reduce morbidity and mortality in patients with metastatic colorectal cancer treated with Irinotecan. And that's another example of a pharmacogenomic test. Again, there was insufficient evidence.

It was unclear in the review whether even in the subgroup of the population that had genomic testing that suggested that -- indicated that they would be at higher risk of side effects, that reducing the delivered dose of the chemotherapy would result in net benefits because there's greater risk of recurrence of the cancer. The recommendation on genetic testing strategies in newly diagnosed individuals with colorectal cancer aimed at reducing morbidity and mortality from Lynch syndrome in relatives I think illustrates an example of some -- of where genomic testing, genetic testing is somewhat different from other diagnostic tests that might be

considered by CMS -- is that there are implications for relatives with inherited disorders.

And in this case, the evaluation and recommendation did consider benefits to the patients as well. Patients with Lynch syndrome are at higher risk of having a second primary or recurrence. And the working group found that there was insufficient evidence on that, although there are other groups that dispute that kind of a finding. So information on these are available at this EGAPP reviews website. The reports can be downloaded. There's an arrangement with Genetics in Medicine that these are made available free of charge. So in terms of considering EGAPP, we think these methods can potentially be used to assist in determining the requirement for a given test by CMS. The EGAPP methods were developed from -- based on various advisory group recommendations.

They used evidence based medicine kinds of approaches which are standard. The method was developed by an experienced panel that was expert in developing these kinds of methods, including

geneticists and others of a very diverse set of disciplines in that working group. They've been peer reviewed. The methods are published. There's been generally a positive response from stakeholders and their peer reviewed published reviews and recommendation statements on diagnostic and pharmacogenomic tests and the prognostic one as well are available as examples that could be considered.

So I want to end the talk. I just had prepared a brief presentation. And to say that -- I want to include that these findings and conclusions are mine and do not necessarily represent those of CDC as a whole. Thank you.

DR. MC NEIL: Thank you very much. Are there questions for Dr. Coates? Yes, Neil?

DR. HOLTZMAN: Well, it's very gratifying to see CDC has taken the ball that was recommended by a number of task forces and to have made this kind of progress. So I'm very congratulatory of (unintelligible) and the whole CDC for undertaking this effort. In our handout we received the three

recommendations that appeared in Genetics in Medicine in January. And I want to ask you a question about that because it seems to be it greatly complicates the task that you have to perform.

DR. MC NEIL: Neil, do you want to just point to what you're talking about right now?

DR. HOLTZMAN: Yes. I will. It's on page 71.

DR. MC NEIL: Okay.

DR. HOLTZMAN: At the bottom of the page under the heading, "Contextual issues important to the recommendation." And what I'm referring to is actually the fifth bullet. "A future scenario with a proliferation of competing licensed products without comparative effectiveness data has potential to confuse patients and clinicians." And not to --

DR. MC NEIL: I'm sorry, Neil. I don't mean to -- I'm sorry. I don't mean to interrupt you. I think people, including myself, are having a hard time finding out where you are.

DR. HOLTZMAN: Okay. It's under the first

paper that follows Dr. Coates talk. "Recommendations from the EGAPP working group, can tumor gene expression profiling improve outcomes in patients with breast cancer."

DR. MC NEIL: Oh, okay. In the original packet that we sent out. Okay.

DR. HOLTZMAN: And if you look at the bottom of page 71 of that reprint, that's what I'm referring.

DR. MC NEIL: Got it.

DR. HOLTZMAN: So essentially what I'm asking, as you note there, is the problem that you're often dealing with proprietary tests, some of which -- and there are examples in this paper -- the trials of have been supported by the test manufacturer. And you have a difficult time in terms of the limitations of that data in terms of what comparisons they've made of deciding whether they're beneficial. And we're going to probably see more of that. How are you going to deal with that? How are we, the public, going to get unbiased information that really tells us whether any given proprietary test adds something to what we've

already got?

DR. COATES: Well, I think that the Secretary's Advisory Committee on Genetics, Health, and Society has made recommendations around that, that more information be made available. I think they specifically requested that information be provided on an easily accessible website. How to get that done, I can't say. I'm not -- don't have insight into what would be needed to get that information out. I think it needs to be -- needs to be made publicly available. But I can't say how to get that done.

DR. MC NEIL: Can I follow up with another question? On two of your slides regarding the SSRIs, on one slide you say there's no consistent association between blah and blah. And then in your recommendation you say there's insufficient evidence for or against. And the question I have is, does no consistent data always mean insufficient evidence, or does it sometimes mean forget it?

DR. COATES: As I understand it, the working group is -- it does take into consideration other

pieces of information for their -- there are going to be different categories of insufficient information. I think when there's inconsistent evidence, it's always going to be an insufficient evidence.

But they may take other pieces of information like the contextual information to -- then go the next step and say it's probably not a good idea. We can't -- there's not sufficient evidence to make -- information to make a recommendation for or against. But we can make a statement about use of the test now based on what information that is available that's broader than just that particular piece of evidence.

DR. MC NEIL: So in other words, in your two slides where you go from no consistent association to insufficient evidence, there are several paragraphs in between there --

DR. COATES: Yes.

DR. MC NEIL: -- that make it impossible for us to justify the second conclusion on the basis of the first statement.

DR. COATES: Right. Right. Yeah. Then there were those contextual issues around the fact that it's very difficult for a clinician already to select a medication and to identify a dose. And adding uncertainty with that, with a costly test could result in harms. And that was their -- that was their conclusion. And that's why they ended up making the recommendation of not using it.

DR. MC NEIL: Marion?

DR. DANIS: When the EGAPP group was deciding about how to come to their conclusions, the -- I want to ask how you weighed -- which way you leaned on the tendency to want very well controlled studies versus the reality that often those can't be done but you have a preponderance of evidence -- bits of evidence that make for a coherent story tending to show something's effective, and our understanding that with time you may -- you know, initial studies show a positive finding. And subsequently somebody repeats it and does a randomized controlled trial. What was the general philosophical approach of the group about

moving to conclusions when you've got a moving target about evidence?

DR. COATES: Well, I think the group -- the general approach is to recognize that there may not be single studies that are going to provide all of the information. And that's why they work on these analytic frameworks and try to assemble the chain of evidence to get from A to Z. So their approach is to try to use what's out there, what's available to -- you know, to get to a recommendation. And they don't -- they consider all kinds of evidence, all types of evidence. It's graded with, of course, clinical trials being ranked highest. But the way their method is put together and in that article that Steve Teutsch is the lead author on, it is possible to get to a conclusion without clinical trials data. So -- and that is their approach.

I think in terms of -- just to give some context to that, three of the members of the working group have been on the U.S. Preventative Services Task Force and have experience with making these kinds of

decisions. So for example, that task force made recommendations for the use of sigmoidoscopy based on observational studies. It didn't require a clinical trial. There were two case controlled studies. And they made recommendations for the use of colonoscopy based on biologic plausibility based on the evidence for sigmoidoscopy.

So I think there's a willingness to try to give -- to take the whole body of information that's available to make a determination. And that's why it's so critical that the test, the use of the test, the purpose of the test, be so clearly specified up front. And you know, the overall harms and benefits issue, you know, comes into play in making that final decision. I think they would require more information for which something is more likely to cause harm.

DR. MC NEIL: Maren, did you have a question?

DR. SCHEUNER: Hi, Ralph. Thank you very much. I was just going to ask my question again about Mendelian disorders. Do you feel that the EGAPP

methods could apply to evaluation of genetic tests for Mendelian disorders? And why -- when you introduced the topic, you indicated that the EGAPP was set up mostly for the common diseases. Although you did look at HMPCC.

So could you just -- I'm just trying to get this group to start thinking about whether or not we should consider Mendelian disorders differently from common multifactorial disorders as we consider what evidence is necessary to make our decisions.

DR. COATES: Right. I think that's a pretty complex question. The reason that the EGAPP working group decided to take up the common diseases because other groups were already -- there are groups making recommendations and doing evaluations for, say, newborn screening. And from the genomics community for rare disorders, I think one difference between the EGAPP approach and the approach for many of these single gene disorders is that in the genetics community, it's more commonly thought that information is clinical utility. And so one doesn't have to

necessarily have an intervention that's going to provide health benefit based on that genetic test. That just providing the information may be useful in and of itself to the patient and to the -- and potentially to family members.

So I think the difference is -- I think this framework could inform the approach to rare single gene disorders. But one might have to broaden the -- in getting to clinical utility, the issue there. The EGAPP working group has more of a focus on clinical outcomes and medical decision-making and less on say, provisional decision-making that would inform a patient whether or not to have an additional child or you know, that kind of thing. Those are -- that's not so heavily weighed in the EGAPP process.

DR. MC NEIL: Neil, final question.

DR. HOLTZMAN: Yeah. I want to come back again to that paper that followed your presentation, the one on EGAPP recommendation on tumor gene expression. And it relates to the question of clinical validity. For instance, table 2 on that

paper, page 70, looking at the relation between specific genotypes and risk of distant metastases.

And you give the sensitivity, specificity from the various studies and odds ratio. But what's missing there is predictive value. And it seems to me that odds ratio is a very poor substitute for negative or positive predictive value. And I wish you'd comment on your or CDC's feeling about what is sufficient evidence for clinical validity beyond sensitivity and specificity.

DR. COATES: I personally think predictive values are very important. And I think that would have added a lot of information to that review.

DR. MC NEIL: Thank you. I have the sense that we're on a roll right now with the two talks that we've just had and the multiple questions. And I'm going to make a proposition which you are free to reject. And that is that we keep going with public input right now, and individuals as they wish to take a break.

So if you -- I will take a vote on that. We

will either take a 15-minute break right now, and you can go do whatever. Or we will keep going and you can in an ad hoc way get up. So how many would like to take a break right now? Steve would like to take a break. Okay. Steve, take a break. You were on the Beltway or on some road or other. Okay. So why don't we ask our -- seriously, I'm just getting a sense that the group is getting wound up and getting the questions. And I'm just afraid that we might -- I'm sorry?

So let's see. We have Mitchell Burken. And I'm told by Maria, our very precise executive secretary that I have a timer here. And the minute you open your mouth, I press start.

DR. BURKEN: Hi. My name is Dr. Mitch Burken. I'm representing the contractor medical director new technology work group. We're an informal discussion group without any designated Medicare program responsibilities. We're comprised of CMDs from AB MAC, a program safeguard and appeals or redeterminations contractors.

The CMD new technology work group acting within its purely advisory role, which is to emphasize the critical importance of crafting detailed guidance on personalized medicine, in this case, molecular diagnostic testing evidence construction. It is hoped that this level of detail will allow all relevant stakeholders, CMS, private insurers, test assay developers, academic reference laboratorians, clinicians, et cetera, to clearly understand both the type and quality of evidence which is needed to support Medicare's reasonable and necessary coverage criteria.

We believe this guidance will establish a more consistent, rational, and fair coverage process. We hope that this guidance will include specific acceptable study designs such as RTCs, comparative cohort studies, nonconcurrent archive specimen testing, and acceptable metrics, for example, by a marker driven RCT arms, high negative predictive values to avoid unneeded therapy.

And we really hope that this MEDCAC guidance

will include at least clear specification of whether reasonable and necessary can be met with relatively strong clinical validation studies. If so, what are acceptable design metrics for such studies and/or outcomes driven clinical utility studies? And if so, what are feasible robust study design metrics for all stakeholders to embrace?

And on this note, we'll hope that you can reference or at least consider relevant methodologic dialogue from the recent 12/16 FDA ODAC regarding outcomes and prospective retrospective biomarker studies even after acknowledging that they're really different mandates of FDA and CMS. And what I'm going to do is, since I have a couple of minutes left, go off script a little bit since Dr. McNeil did touch upon the KRAS mutation study and asking the panel about the relevance or robustness of a single study.

And as I said, as one gets into the details of the FDA ODAC, there were discussions on, you know, what really makes one of these prospective/ retrospective studies viable. You know, what types of

inclusion rates of archive specimens are valid versus what kinds of inclusion rates are not valid.

And again, this is a little bit beyond what may be the MEDCAC discuss. But I do -- or we do believe that there is, you know, an overlap in what goes on. And finally what I'm going to do is just kind of go off script one more time for a moment and just say, in the questions, question number three talks about a prognostic assessment. And certainly, you know, there are a lot of terms being used today, clinical validity, clinical utility, and there's just a whole compendium of different words that are used.

But I would contend that the term predictive assessment is certainly relevant to this group. And I would just ask that the panel during the later discussions perhaps clarify and contrast a predictive test versus a prognostic test 'cause I do believe that's an important point. Thank you.

DR. MC NEIL: Thank you very much. Okay. Let's see. We have -- who is it -- Dr. Fowkes from the ACP.

DR. FOWKES: Good morning. My name is Dr. Mary Fowkes. I'm a professor of pathology at the Mount Sinai Hospital and School of Medicine in Manhattan. I'm here today on behalf of the College of American Pathologists where I'm a member of the Federal and State Affairs Committee as well as a member of the College's Molecular Pathology working group.

The College appreciates the opportunity to appear before you today and provide our perspective on evidence requirements for genetic and genomic tests. I have no other financial incentive aside from my involvement with the CAP. The College of American Pathologists is a national medical specialty society representing more than 17,000 pathologists who practice anatomic pathology and laboratory medicine in laboratories worldwide. The College's commission on laboratory accreditation is responsible for accrediting more than 6,000 laboratories here and abroad. Our members have extensive expertise providing and directing laboratory services and

participate as peer inspectors in laboratory accreditation programs.

I would first like to highlight the role of pathologists in developing, delivering, and interpreting genomic tests for patients. Laboratory directors are responsible for ensuring that test methodologies selected have the capability of providing the quality or results required for patient care. A pathologist interprets the data produced in the laboratory in the context of the patient's personal and family health information. In addition, many test results require interdisciplinary discussion between primary care clinicians, radiologists, oncologists, and the pathologist where the pathologist provides clarification regarding the significance of test results, unusual or unexpected results, and recommendations for additional testing in the context of the specific patient.

This process ensures the right test for the right patient at the right time. College members feel that genomic tests are not unlike numerous other

laboratory tests that they have successfully introduced into medical practice. Indeed, the development of molecular tests such as BCR/ABL and HER2neu testing had allowed the diagnostic assessment of disease entities with specific associated prognostic and pharmacogenomic implications in both hematologic and solid tissue tumors.

This targeted approach is transforming our understanding of tumors, allowing access to targeted therapy irrespective of gender, ethnicity, and age. Although the use of BCR/ABL and hertonew testing is similar to prior diagnostic tests with both laboratory validation and accreditation, additional genomic tests may be at different stages in the evolution of -- in the accumulation of evidence.

There is no uniform checklist that can adequately represent the pathologist's complete evaluation for each patient. Therefore, the clinical decisions and use of a test is -- however, the clinical decisions and use of a test is still guided by well-established performance characteristics,

including appropriate patient population, establishment of normal values and reference ranges, clinical specificity and sensitivity different from the analytical specificity and sensitivity, as had been discussed here today, positive and negative predictive value, and documented correlation of laboratory findings with other studies from the literature.

These performance characteristics are described in detail in an Archives of Pathology and Laboratory Medicine publication that is currently in press which we can provide the reference to. I would like to discuss two examples of recently introduced tests that are improving patient care, but where the level of evidence for clinical use is different.

KRAS, as we've discussed, is an oncogene frequently mutated in several types of cancer. Testing for KRAS mutations has guided chemotherapy decisions for patients with metastatic colorectal cancer and has recently garnered a lot of attention.

The presence of mutations has shown in

multiple studies to correlate strongly with the -- with whether a patient will respond to treatment in the use of specific chemotherapy agents. This genetic assay selects patients most likely to benefit from therapy, protects those unlikely to benefit from harmful side effects, and saves money for the health care system.

Gliomas are primary tumors of the brain. Gliomas provide a second example where a genomic test provides personalized benefits for patient management since chemotherapy would not be initiated without a clear end point in therapy. Oligodendroglioma is a particularly important type of glioma with distinctive histologic, molecular, and clinical features. Molecular testing has revealed losses of portions from two chromosomes, 1 and 19, correlate to a diagnosis of oligodendroglioma.

Importantly, loss of regions from the short arm of chromosome 1 and long arm of chromosome 19, 1-P and 19-Q respectively, correlate with higher sensitivity to specific chemotherapy and better

prognosis with a five year survival rate 50 percent higher than in tumors without 1-P/19-Q loss. Prolonged survival in response to chemotherapy has recently been found in additional tumors, brain tumors, with 1-P/19-Q loss such as tumors with intermediate oligodendroglia features or mixed astrocytic tumors.

Because there are no alternatives for this test, and the test may be used to guide decisions about the length of therapy, evidence requirements may be different than those for KRAS testing. In summary, the CAP believe pathologists and other laboratory professionals are key sources of knowledge and experience in the development and delivery of high quality cost-effective laboratory services. And the CAP is willing and eager to contribute to discussions with clinicians, regulators, payors, and others sharing common interests.

As medical specialists in the diagnosis of disease, pathologists have a long track record of participating in -- practicing -- sorry -- evidence

based medicine through the development of appropriate laboratory tests and selection of alternative diagnostic methods.

As you consider the recommendations on evidence requirements, we ask the MEDCAC to keep an open mind, recognizing that one set of criteria may not be appropriate for all testing situations and carefully review the impact of your recommendations on the ability of pathologists to provide diagnostics in the best interest of our patients.

Thank you very much.

DR. MC NEIL: Thank you very much. Bruce Quinn from Foley Hoag.

DR. QUINN: Hi. It's a pleasure to be here. My name's Bruce Quinn. I worked for four years as a regional Medicare medical director, and I'm currently full-time staff with Foley Hoag which is a law firm. I'm an internal sort of thought capital expert. And I'm not here representing any particular client or viewpoint but my own.

So I'm going to make three points in my five

minutes. One, as Mitch said, the legacy guidance for genetic test coverage is limited. Second, local carriers do have difficulties with these coverage policies as you saw a few minutes ago. And some experience we had with frameworks to build that -- bridge that cap.

In Medicare, the statute says we can't pay for care that's not reasonable and necessary. And there are several places over the years Medicare has amplified that with additional phrases. Appropriate in duration, accepted practice meets but does not exceed the patient's need.

If you compare the FDA with Medicare which is interesting on many levels -- more than we have time to go into -- Medicare looks at published medical literature. The FDA can have potentially cartons of data.

Often the criteria for Medicare and the FDA are very different. For example, the duration and frequency are very important for an insurer, but usually not mentioned on an FDA label.

Some examples with local coverage and these policies, there's an LCD, a local coverage policy in Medicare for the BRCA test which is performed in the states only in Utah. And that policy was developed in the early 2000s by Utah Medicare, at that time a one-state Medicare program. As Medicare has been recontracted, there's now a group of states in the same jurisdiction that have the same policy specifying if you ran a lab in North Dakota or Montana exactly how you would get paid for the BRCA test, although no labs do it there.

And the carrier Noridian spread the LCD to several other states with a couple of caveats. It says we cover the BRCA test, the colon cancer test, and no other molecular testing in the policy, although you can track down some footnotes on the website that say they don't apply it to infectious disease, and they only apply in part B, not in part A. But that's not in the actual policy. Anyway, as there is further recontracting, Trailblazer took over Colorado and immediately dropped the policy. Palmetto took over

Nevada and California and imported the policy into California, and after six months dropped it because it only covered BRCA testing. And if you read it literally, it excluded any other kind of molecular testing.

And no other states have anything similar to that policy. Now the whole eastern two-thirds of the U.S. there's nothing comparable. I found that frameworks can help bridge the gap. And the dialogue here is not just within the insurer community, but with manufacturers as well. Something that Mitch referred to, looking at the larger group of stakeholders. I'm going to use an example of the oncotype DX test. I easily could have used a fictitious test, an 10-G pancreatic cancer or a test for cancer X with N genes. But people probably in their heads would have been translating it to something they knew.

So the question is, how do you think about this as a pair where you have much -- at least a local pair where you have much less time than a, you know, a

three- or six-month tech review. So early on we figured there were probably -- we should break reasonable and necessary into four questions. And it turns out we were actually echoing the Fryback and Thornberry in a somewhat simplified form. Although no one had mentioned that to me at the time, and I was on a lot of conference calls in Medicare, and no one ever mentioned that particular schema.

But we said is the test valid. You know, is it reasonably accurate and valid? Is it incremental accuracy over existing tools? How much better is it than other tools? And then for which patients is the test actually necessary? Another thing I noticed and was actually after hearing Dr. Gutman speak at a conference is how important the label is at the FDA. Very specific statement tied to scrupulous data analysis. And across the development of a product, there are a wide range of different value propositions.

Early on, you say what is the IP, what are the barriers to entry? Then it's safe and effective.

Then it's clinical utility. And then in the marketplace, there are different value propositions again. I'm happy to send anyone slides that would like copies of these. So I said, well, let's treat these just like an FDA label and have the exact wording for different value propositions. You know, we can measure 20 genes accurately. We can predict recurrence of breast cancer. We can improve the clinical decision or improve the outcomes in breast cancer patients. And each one of those has a different kind of data behind it.

And what you find very quickly is just measuring RNA accurately is pointless. Obviously, that's enough. The bottom level, this test improves survival over ten years in breast cancer patients would actually be very hard to prove, not only from the time table and the dollars, but even from the ethics. And I think the ethics of some of these trials with diagnostic test are often different than for therapeutic.

So if you can determine that here are women

with a five percent chance of breast cancer recurrence, and here are other women with a fifty percent chance, and you know that, you've shown that five times in large studies. You can't now randomize those women at the tails of the curve to getting or not getting chemotherapy. It wouldn't be ethical even if you wanted to wait fifteen years.

So that means that middle step, you can improve the clinical decision, is the decision point. You can't get in less than 10 or 15 years to the bottom one. And it might not even be out there to do an RTC. And then if your question is improving the clinical decision point, it's comparative. What are you comparing it to, traditional pathology, immunomarkers, other approaches?

I'm happy to take questions. Thank you very much.

DR. MC NEIL: Thank you very much, Mr. Quinn. What we'll do is hold questions until we've heard from all of the outside speakers. We now have Russell Teagarden from Medco.

MR. TEAGARDEN: Hi. Good morning. Thank you for the opportunity. That's not me. Am I forwarding here? That's me. Okay. Thank you.

Medco is a pharmacy benefit management firm. I'm sure all you know, it's big. The point of listing some of this is to indicate that we have a lot of information about utilization of drugs and so forth. We're pushing about 70 million covered lives now with lots and lots of different kinds of plans. So there's just an awful lot of information available from this environment and ones like it.

Our interest in genomic testing is driven by our current research. Our efforts are focused on that because that drives a lot of our clinical programming that is primarily aimed at drug selection, drug dosing and administration cost efficiency. Genomics has a promise there to help that. So we're very interested in it. And it's also becoming a more important element in the health care reform debates. And we are involved in that, want to be involved in that. We have something to say about it. And indeed in this

area, we are currently talking to federal officials about our interest in Medicare demonstration projects involving genomics. And we're willing to use our therapeutic resource centers that are based in our pharmacies as laboratory environments. So we've a big interest in this and put a lot of resources on it.

Now, in case I run out of time here, I want to make sure I get my two main points in. And one is, as far as we're concerned, the requirements for evidence should permit these two items. One is that data can be drawn from a broad range of health care settings and administrative claims data bases so long as they're analyzed by the appropriate methodologies.

Secondly, the strictness or the way the evidence is used should be flexible so we can take into account certain things that are more or less urgent. It got to the point early this morning would one study that's retrospective be enough for you, and this would say maybe, given the situation, how dire it may be, and so forth.

We have done a lot of this type of work with

the kind of information that we get from our environments. We recently showed that PPRs reduce the effectiveness of clopidogrel. Others have as well lately. Up to one in five people might wind up in hospitals in six months after starting warfarin from a bleed or a clot. That's a catastrophe in my view. 25 percent of the U.S. population is on one or more drugs with biomarker information on the product label. In other words, a lot of people are on drugs that there may be some genomic information that's relevant.

I'm not going to go through the rest of this. You can see this. But we've just used this kind of data from natural settings with the appropriate methods to test certain hypotheses to get information that's important to individual patients as well as various policy approaches. Basically, what we're saying is that RCTs are great. They're important. They're necessary. But they mainly just focus on efficacy. In our kinds of environment, we're interested -- the kinds of decisions we make and our

payors make and so forth, we need to know effectiveness. So we can draw a lot of information that can be used to get at the effectiveness question.

There's huge amounts of this information. We're going to be getting more on the genomics as patients give us that kind of information. They do now. We have clinical programs now involving over a hundred of our plan sponsors who testing their members on warfarin and Tamoxifen. The plans enrolled in these programs cover over six million people. And so we're getting a lot of this information in.

Furthermore, you know about the direct to consumer. We can expect a lot of consumers to have this information and turning it over to the health care system, us, you. And so there's going to be a lot to develop there. And it would be a shame to let all this go to waste when we can derive really good information about it.

Flexibility too is important so that we don't let certain opportunities go by just because of perhaps some arbitrary strictness of evidence

requirements. We should take certain situations into account. Let's just take warfarin. Again, here's a situation. The drug's been around for decades. It is a catastrophe in the harm it causes patients. Many, many good intentions and efforts have been put towards making it safer. It's still a disaster. We have some very promising genomic information that would suggest that maybe we could do a little bit better.

And the testing is not a risk in and of itself. And it's getting really cheap. So we should take these kinds of things into account when we decide on the evidence that's going to be used is point of that. So I'll -- I won't recapitulate here because we're quick. I'll stop there. Thank you.

DR. MC NEIL: Oh, thank you very much. Penny Mohr.

MS. MOHR: Thank you. The Center for Medical Technology Policy on whose behalf I am presenting is a non-profit organization that provides a neutral forum for improving the value of comparative effectiveness research so that it's more informative

to patients, payors, clinicians.

One of the major strategies that we are pursuing is to improve the quality of evidence for decision-making collaboratively and to develop specific recommendations for the design of clinical research that reflects the information sought by health care decision-makers. I have no conflicts of interest, and CMTP is funded by a balanced mix of payors, product developers, foundations, and government grants.

A primary goal of today's MEDCAC meeting is to define the desirable characteristics of evidence that could be used by Medicare to determine that a genetic test improves health outcomes in Medicare beneficiaries. We've been looking to answer the same question for a specific test, gene expression profiling for breast cancer treatment through developing what we call effectiveness guidance documents.

In developing our effectiveness guidance documents, we have concretely addressed some of the

questions raised for today's meeting today. I have time to only highlight a few of those. What are the desirable characteristics of the analytic validity of gene expression profiling, and what is the level of evidence that will give decision-makers reasonable confidence that this specific genomic test improves health outcomes.

Before I get into the specifics of the gene expression profiling document, I wanted to say a bit more about our effectiveness guidance documents and the process we use to develop them. Effective guidance documents provide technology specific recommendations targeted to clinical researchers and product developers regarding how to design clinical studies that will be informative to patients, clinicians, and payors.

They're analogous to FDA guidance documents which provide guidance to product developers on evidence needed to obtain regulatory approval. But in contrast, they aim to describe the evidence that's needed for coverage decisions. The study designed

recommendations in an EGD aim to define a level of evidence that will give decision-makers reasonable confidence that a technology improves health outcomes.

I want to emphasize two terms in that phrase. First, decision-makers, meaning health plans and patients and clinicians. Second, reasonable confidence. We're not trying to define a gold standard here for a threshold of evidence.

We recognize that the gold standard may be presented too high a bar, given the pace of innovation and other feasibility issues. Evidentiary standards need to also take into account practicality and financial barriers if we want patients to benefit from this innovation. This brings me to another fundamental aspect of our approach, which is multi-stakeholder and collaborative in nature. Effectiveness guidance documents are drafted and are revised through a transparent and interactive process involving decision-makers, technology developers, clinicians, consumers, and clinical researchers, and others.

For our gene expression profiling document we commissioned Johns Hopkins University to write the effectiveness guidance documents because they had recently completed a systematic review of the evidence for ARC. And they were most familiar with ways in which the existing trials failed to produce the evidence desired by decision-makers. As noted, the initial draft recommendations were crafted with input from a broad range of stakeholders, including CMS, private payors, the CDC, and other members of the EGAPP initiative.

Regarding MEDCAC's questions about the desirable features to establish analytic validity, again, I do not have time to go into very much detail. But just a few new aspects that the report said the test product developers need to report within patient reproducibility the percent of successful assays and the variability of risk classification.

Regarding reasonable assurance that it improves health outcomes, our guidance document currently states that an RCT either concurrent or non-

concurrent, is the only design that provides an unbiased assessment of treatment effects. A non-concurrent assessment from an RCT that has banked analyzable specimens and sufficient patient follow-up can provide very high quality evidence such as was done with the oncotype DX test.

As Janet Woodcock noted in her Health Affairs article in November last year, it is unlikely that one size fits all standards of evidence will be appropriate. Not every diagnostic test needs to be supported with an RCT. But diagnostics that are intended for a large population and have a large effect on treatment patients need to have a robust level of evidence. As this is true with the gene expression profiling test for breast cancer, our effectiveness guidance document has aimed closer to the gold standard.

We recognize that these recommendations are the beginning and not the end of the dialogue. We plan to post our draft effectiveness guidance document on gene expression profiling on breast cancer

treatment on a website within a couple of weeks and welcome your input.

We feel that the important contribution here is to be clear and precise about a set of recommendations to be modified based on feedback from the interested parties. We expect that many will believe that our study -- such studies are not feasible. And we look forward to hearing specific recommendations for study designs that may be more practical while also having sufficient validity.

The approach we are taking is consistent with the recommendations of SAC-GHS and the CDC, both of whom have highlighted the need for public/private collaboration to specify evidentiary requirements needed for genetic tests. Hopefully, these initial efforts will stimulate a constructive dialogue as well as creative ideas on how to improve this initial effort.

Thank you very much.

DR. MC NEIL: Thank you very much as well. Okay. Let's move on to David Mongillo from American

Clinical Laboratory Association.

MR. MONGILLO: I'm David Mongillo with the American Clinical Laboratory Association. I'm vice-president for policy and medical affairs. ACLA represents local, regional, national hospital and independent clinical laboratories throughout the United States. Most, if not all, of our members perform genetic testing. Thus, we have a very significant interest in the issue.

I want to begin by drawing your attention to an editorial that was published last week in the New England Journal of Medicine that was actually noted by Penny Mohr. The title was, "Pharmacogenetics, Tailoring Treatment for the Outylers." It was written by Dr. Janet Woodcock and Dr. Larry Lesko.

It stated that in some cases randomized controlled trials will be needed to determine whether pharmacogenetic testing is worthwhile. In others, less rigorous approaches will suffice. The editorial was written in conjunction with the publication of results from the International Observational

Retrospective Study demonstrating the utility of genetic information for improved determination of the dose of warfarin treatment.

In many ways, the editorial and the study captures ACLA views on the subject. Specifically, genetic testing is a vital tool in helping determine treatment for individuals in the area of genomics. And that such tests can be validated with a range of scientific methods on a case by case basis to demonstrate usefulness in improving care.

The difference between drugs and diagnostics should not be overlooked. Diagnostics can lead to the selection of therapy. They can help manage disease and are used as measurement tools in outcomes research. Thus, there's a broad spectrum of evidence that should be considered in evaluating in diagnostic tests. While randomized clinical trials may be the gold standard for some procedures and therapies, they have significant limitations when applied to many diagnostic situations.

For example, in what we heard earlier by Dr.

Quinn, to determine whether a given test can predict recurrence of a condition, a randomized clinical trial would have to track patients prospectively over a long period of time to determine whether that condition returned. It could be many years before the results are known. We're also faced with problems with the lack of study participants in rare diseases, or, as was pointed out in the New England Journal article about idlers, how we would deal with that in terms of study populations. All of those significant hurdles to performing randomized clinical trial.

The impact would be on patients. That's unnecessary delay in access to diagnostic tests that could immediately help patients, reduce side effects, and help control health care costs, which we all recognize is a very important concern. There are other alternative scientific methods for determining the genetic test is reasonable and necessary. Carefully constructed retrospective studies involving patient data with surrogate markers can yield scientifically valid, clinically meaningful results

quicker. That's not significantly different than that study that was published in New England Journal of Medicine.

Genetic test validation can be performed using archived specimens that have been stored and cataloged. It is often possible to use such samples consistent with principles of informed consent and appropriate treatment of patients and patient specimens to determine whether an individual with a given genetic profile ultimately had a recurrence or responded to a particular drug or therapy.

It is possible to analyze such findings to determine whether it adequately supports the conclusions being drawn. Utilizing such retrospective reviews of archived specimens in lieu of prospective clinical trials can result in more rapid determination of the utility of a diagnostic procedure without adversely affecting incentives to develop beneficial new tests.

Differences in the type and uses of diagnostic genetic tests strongly suggest it would be

both difficult and inappropriate to develop a single national coverage decision that can apply to all genetic testing. ACLA encourages MEDCAC to recommend flexibility and alternative options that are based fundamentally upon the notion of balance. A balance on the need for sound evidence and the need for patient access.

Thank you.

DR. MC NEIL: Thank you very much as well. Okay. Our last scheduled speaker is Roger Klein from the Blood Center of Wisconsin.

DR. KLEIN: Hi. I'm Roger Klein. I'm here on behalf of the Association for Molecular Pathology, a medical and professional society comprised of approximately 1700 physicians, doctoral scientists, and medical technologists. I have no conflicts to report.

The desirable characteristics of evidence for most genetic and genomic tests should not differ from those associated with diagnostic testing or other diagnostic medical procedures generally. However, DNA

and RNA-based tests are heterogeneous in their methodologies and wide ranging in their clinical applications. Moreover, genetic and genomic tests are integral to the concept of personalized medicine.

An extraordinary volume of new discoveries is combined with evidence-related issues not necessarily specific to genetic testing to present historically unique challenges in evidence evaluation. Available studies may be limited in size, number, and scope. And study subjects may vary in disease course and presentation. Novel information from genome-wide association studies may present unique statistical challenges.

Importantly, both the current and previous administrations have given furtherance of personalized medicine a prominent place among their health care policy goals. The evidence standards on which CMS coverage decisions are based will play a major role in the extent to which progress in personalized medicine is made and the speed with which genuine advancements are introduced into medical practice.

Clear and professional society laboratory accreditation programs help ensure analytic and clinical validity of genetic and genomic tests within individual laboratories. Analytic validity encompasses analytic sensitivity, specificity, assay reproducibility, linearity for quantitative tests, and consistency in response to limited changes in pre-analytic and analytic variables.

Yet the meaning of sensitivity and specificity may vary with the assay under review and the diagnostic question posed. Some genetic tests lack a gold standard for comparison on results. In these cases, collaborative studies using a single large representative panel with well characterized samples that are tested and reported blindly under routine laboratory conditions are desirable.

However, such ideal studies are rarely performed. Fortunately, molecular diagnostic methods tend to excel analytically. Although assays differ in technical features and clinical applications among laboratories and institutions, the CLIA regulations

and the laboratory accreditation program of the College of American Pathologists help ensure their analytic and clinical validity.

Clinical and analytical validity should be prerequisites to the non-investigational use of genetic and genomic tests. However, those of us who are proponents of evidence-based medicine must recognize that it has inherent limitations when applied to genetic and genomic assays. For these tests, evidence of analytic and clinical validity may be adequate. But clear demonstrations of clinical utility may be lacking, even for tests widely believed to have medical value.

In the absence of direct proof of clinical validity, an important role for physicians' experience and judgement remains. Medical assessments are rarely based on a single test alone, genetic or otherwise. But instead, consider patient history, physical signs, and the results of other diagnostic modalities.

Detection of genetic variance, KRAS being one example, supplements the pathologic examination of

tumors. DNA-based testing is combined with clinical and other laboratory data in the diagnosis of inherited disorders. The multiplicity of factors contributing to drug metabolism renders clinical judgement essential for the use of pharmacogenetic testing in medical practice.

The absence of high quality direct outcome based data often necessitates reliance on surrogate markers. Although changes in physician-directed patient management may indicate a consensus within the medical community about the value of a particular test, they do not necessarily ensure that the clinical utility of the test has truly been demonstrated. In some instances, knowledge later acquired will cause rethinking or refinement of practice changes.

Intermediate -- indirect or intermediate outcomes can be helpful in assessing the clinical utility of a test. Yet surrogate markers can be misleading because they may overlook the effects of variables not considered.

Again, direct outcomes-based data is rarely

present for genetic or genomic tests -- because direct outcomes-based data is rarely present for genetic or genomic tests, clinical judgement, context, and expert opinion remain necessary to assess utility and argue against rigidity in CMS's approach to coverage. Ethical issues ordinarily should not adversely impact the rigor of clinical studies of genetic testing. However, there many be areas for which ethical issues could potentially affect study quality.

Diagnostic testing for inheritable diseases has implications for a patient's family members. Studies of cancer patients address diseases that are often fatal and for which therapies may be highly toxic. Concerns about the implications of genetic information or the apportionment of potential useful diagnostic approaches among terminally ill patients could potentially hinder the recruitment of study subject or bias results in unforeseen ways.

The age of the Medicare population usually should not adversely impact the generation or interpretation of clinical studies of genetic testing.

However, it is possible that age-related attitudinal or demographic characteristics or a greater overall likelihood of death could potentially with study recruitment and/or bias results.

Moreover, as test validation has not been performed on significant numbers of older patients or is unreflective of their disease status, generalization of results to Medicare patients may not be appropriate. Lastly, in disorders characterized by age-related expressivity, disease features that could impact assay performance may be different in Medicare patients than in the larger affected population. As experts in the clinical use and technical aspects of genetic and genomic testing, the Association for Molecular Pathology stands ready to assist the Committee, CMS, and its contractor medical directors to help address the complex evidence-related issues associated with genetic and genomic testing.

DR. MC NEIL: Thank you very much, Dr. Klein. Is there anybody in the audience who like to make some comments? We have time for some unscheduled

comments. Otherwise, we'll go on and address questions to our previous commentators. No? Okay. We'll move on. So at this point, it would be useful to have the panel, that's this group, raise any questions that they have to those individuals who have made presentations all morning, either the previous six speakers or the prior two.

Okay. Deborah?

DR. SHATIN: I have a question for Roger Coates?

DR. MC NEIL: Roger, are you here?

UNKNOWN MALE VOICE: Yeah. He's here.

DR. MC NEIL: Oh, yeah. To the right.

DR. SHATIN: My question had to do with the study of SSRIs. And I think it raises the broader question of evidence-based for testing in relationship to pharmacogenomics. My question is, in this particular study -- and we just have the summary here -- that the information wasn't really sufficient to be able to determine whether these classes of drugs were effective for patients.

My question is, in this particular instance, the effectiveness of particular drugs within the class may vary by the individual patient. So I think it raises the question, we've spoken globally about the complexity of genetic testing that you can't have one set of standards across all different tests. In this instance, just within a class of drugs, there may be specific information that's important to look at that is not addressed in doing this type of an analysis. And I raise that as a question to see whether the results of this particular working group, whether they looked at differentials across drugs within the class.

DR. MC NEIL: Whether they look at what, Deborah? I'm sorry.

DR. SHATIN: Different adverse reactions --

DR. MC NEIL: Oh. Adverse reactions.

DR. SHATIN: -- or efficacy for drugs within that class of the SSRIs.

DR. MC NEIL: By different kinds of patients, is that what you were inferring?

DR. SHATIN: Right. Right.

DR. COATES: If I understand the question correctly, it's is there variability among the different SSRIs --

DR. SHATIN: Right.

DR. COATES: -- in the ability of the test to predict response.

DR. SHATIN: Right.

DR. COATES: And the answer is that there were small numbers of patients in these studies. The studies are small to begin with. And then when you try to separate effects by different medications, there's even less information available. But it appeared, just based on the variability that was observed, is that there probably is variability in the metabolism among the SSRIs and which variance would be associated with that. So there might, in fact, need to be tests for each specific SSRI.

And the PROG (phonetic) drugs were metabolized differently than the SSRIs, and there's probably variability among the SSRIs. That was part of the discussion in the evidence review.

DR. SHATIN: Okay. Thank you.

DR. COATES: But there's limited evidence about that.

DR. MC NEIL: Steve?

DR. GUTMAN: Ralph, while you're up there, I have a question as well. Because my personal view of the EGAPP process is, of course, that it's sort of like driving a Cadillac. It's the gold standard. It's wonderfully done. The flip side of that is that it's the gold standard, and it's wonderfully done, and it takes a lot of time to do. And this morning we've heard certain themes about flexibility and about contingency. It's my understanding that no one in your group is, in fact, looking at ways of dealing with data that are incomplete in a more facile manner. Is that correct, and if so, can you elucidate what your thoughts are on lessons learned from EGAPP that would allow interim decision-making?

DR. COATES: Sure. The process has taken a long time. This project's been in existence for several years. And as you've seen, there are only

four recommendation statements. And that's in large part because the methods were being developed as the working group put together the strategy for how to deal with these issues. But there is recognition that the evidence reviews are very complex and take a long time.

And there's a plan now to try to develop a model for what are called rapid reviews that would essentially address the same set of six issues that I outlined, going from the -- what is the specific test, the specific clinical scenario, the specific disorder through the issue of clinical validity and overall balance of harms and benefits. But it would be done in a more rapid fashion.

Those methods haven't been clearly worked out yet. There are a couple of examples that have been called rapid reviews. And a rapid ACE review was done for warfarin. And that was part of the basis for the American College of Medical Genetics recommendations around warfarin use.

So there's an -- the -- we recognize that

the process is slow and limited. And I think it's still a difficult issue because all the evidence needs to -- all of what's available needs to be taken into account to make a judgement. And it has to be for a specific test for a specific use in a particular scenario. And so there's need for information that's very specific.

DR. MC NEIL: Thank you very much. My apologies to Dr. Fowkes. I wasn't looking out at 180 degrees, so I missed her hand. So?

DR. FOWKES: I just had one question or one comment to the group. I know that there's been a lot of emphasis on molecular testing. I think there's a lot of validity in research work in molecular testing and certainly, some possibility of value for molecular testing in diagnostic work.

But I have a little bit of a concern on some of the evidence being presented in literature on molecular testing. And the only -- I'm going to present one example because it's the only one that I'm very -- more familiar with. And it's been discussed a

little bit here today.

The Oncotype DX testing that has the 21 gene assay, if you go to their website to look at their literature related to the validity of their testing, they reference a paper in the New England Journal of Medicine which is from 2004 and makes the comment that the breast test for serving the American Joint Council on Cancer did not add tumor grade to its staging criteria because of the sparseness and variability of the data.

However, the current standards for staging based on the Journal of Clinical Oncology and the Association of Directors of Anatomic and Surgical Pathology from 2004 and the Journal of Clinical Oncology from 2009 both state that the histologic grade is assessed by the Nottingham grading system, which is a more current grading system being used today by breast pathologists, provides a strong predictor of outcome in patients with invasive breast cancer, and should be incorporated into breast cancer staging systems, and is currently the recommendation

for staging of all breast cancers.

And so I -- my concern with some of this genetic testing is the thought that genetic testing is going to give something that is more or better than the pathology evaluation of the specimen. And I'm not sure -- you have to be careful when you evaluate this literature that you're comparing the same things.

DR. MC NEIL: If I understood you, the Onco-DX article was published in 2000. Is that what you said?

DR. FOWKES: No. The Onco-DX literature was published in 2008. But its prospective validity study was published in 2004 and references a different method of grading --

DR. MC NEIL: Right. So I guess the question --

DR. FOWKES: -- which was not --

DR. MC NEIL: Right.

DR. FOWKES: -- which was not thought to have as much validity in staging. The current recommendation for breast cancer staging uses the

Nottingham grading system which is a very strong and robust method for grading.

DR. MC NEIL: So I think what you're raising is an issue. I don't think we want to pick on that one test in particular. But --

DR. FOWKES: No. No.

DR. MC NEIL: -- to use that as an example that would get at what's the incremental value of a test.

DR. FOWKES: Yes.

DR. MC NEIL: And in fact, you're saying that the Nottingham data may provide just as much prognostic information as the Onco-DX test now. But I thought -- I guess that will be determined if the Onco-DX -- is that the TAILORx study that is being done for that? If that study shows the same predictive powers for its mid grade. But your point is well taken.

DR. FOWKES: I think it's important to look -- make sure you're testing the same -- or looking at the same things.

DR. MC NEIL: Well, we may not be looking -- I think we're always -- aren't we always going to be a situation where it's possible that we're going to have temporal differences between what was collected in an archive -- in a randomized clinical trial and the data that were used at that time compared to what we have currently.

If we lock ourselves into what is available in 2009 and look for impact on outcomes for tumors that have any kind of differences in survival after, you know, a year, we're dead. So we can't be quite that rigid.

DR. FOWKES: Right. But the criteria for pathologic diagnosis is something that's continually changing as well. It's not going to change every year. But certainly over a five-year period, the diagnostic criteria for certain tumors does change. And that should be taken into consideration.

DR. MC NEIL: That's a very good point. Thank you for making that.

DR. SCHEUNER: So let me just ask again, I

mean, how that relates is that the oncotype DX test as I understand is appropriate for a certain stage of breast cancer, stage I, lymph node negative, ER positive. And what you're saying is that with this Nottingham system of staging was different from the staging that was used for the original?

I'm just trying to understand. Is it really a matter of a comparative effectiveness, or is it really a matter of getting the right disorder for the test?

DR. FOWKES: Okay. So when there's a breast cancer being evaluated in the hospital, and the tissue is removed, the tissue is evaluated by the pathologist, and a specific tumor stage is designated for that tumor. They do PR -- they do a variety of different tests including ER positivity and evaluate whether the nodes are positive or negative.

In instances where the tumor is ER positive and node negatives, the oncologists tend to request the oncotype DX be done -- testing be done, the 21 assay array. However, the pathologist evaluating the

tumor, if you look at the grade of the tumor histologically and use the Nottingham grading criteria to evaluate those tumors, the probability of recurrence correlates to that grading and gives similar results and similar findings to what the oncotype testing does.

DR. MC NEIL: Could you make those data available to us or to CMS, the published article, because that's actually what we have to go on.

DR. FOWKES: I can give you what I have here.

DR. MC NEIL: No. You can send it to them later.

DR. FOWKES: Okay. Sure.

DR. MC NEIL: Great. Thanks.

DR. FOWKES: Thank you.

DR. RADENSKY: Thank you, Dr. McNeil. I'm Paul Radensky with McDermott, Will, and Emery. And we are outside counsel to Genomic Health with oncotype DX. Just two points that I wanted to make. One is with respect to the study that was published in the

New England Journal in December 2004. That was based on the National Adjuvant Bowel, Breast Cancer Study B-14 which was conducted from the late 80s on, so that the staging that was there was different data that were available at that time.

DR. MC NEIL: Right.

DR. RADENSKY: Which I think is reflecting, Dr. McNeil, your point that the pathological evaluation -- and when you're looking, you're always looking at something that is a moving baseline, which I think is an important point when you're looking at any of these tests.

Another point that I think is important is that there are data beyond the 2004 publication. There are studies looking at chemotherapy benefit that were published in the Journal of Clinical Oncology and then other data that I know were presented this morning from Kaiser.

The last thing that I think is also very important is the TAILORx is not a study of looking at oncotype DX versus management without oncotype DX.

It's a study where the National Cancer Institute has determined that all of the patients who potentially would be eligible for the study will get the assay.

Those who have low risk -- and it's a slightly different assessment of low risk that the NCI is using from what Genomic Health has published. That those that are low risk would be treated without chemotherapy consistent with the treating physician's determination. Those that are at highest risk would get chemotherapy consistent with the physician's determination.

And really the clinical question is the value of chemotherapy in the middle range. It is really not -- the primary endpoints are really not Oncotype DX endpoints.

DR. MC NEIL: Well, thank you very much. I do think, as I said, it's not our job. We're not evaluating Onco-DX here. And I want to make sure we keep that in mind. I think the purpose of your discussion and Dr. Fowkes as well emphasizes the need to realize that we're talking about issues that relate

to the consideration of genomic tests.

And the one that's at hand right now is the incremental value, and the new one that is at hand is the fact that these tests are frequently going to be done with historical clinical data. I mean, medicine does move on. So we have to realize we're not always going to have everything lined up like peas in a pod for a particular test at a particular time compared to everything else. We'll do the best we can. So I think we'll just put that. But other comments? Yes, Mina?

DR. CHUNG: Thanks. I am in the context of trying to establish the level of evidence required to establish clinical validity. I'm interested in Dr. Coates' comments regarding that same issue we've brought up over and over again about this CYP450 SSRI where there's no -- where you found no consistent association in terms of clinical validity. But that led to an insufficient evidence recommendation just based on that.

Given that many -- for many of these tests,

we're looking at single genetic tests. And these don't account for gene-gene, gene-environmental interactions which are very under-studied. So at this point, what kind -- is that the type of information that you include in your contextually considered recommendation, number one?

And number two, what kind of evidence would you then have used to make a negative recommendation as opposed to the insufficient evidence recommendation?

DR. COATES: First, I should say, these wouldn't be my recommendations. But they're recommendations of the working group. And my understanding of the way the working group works is that the overall -- to make a negative recommendation, they would have to determine that the balance -- potential balance of benefits and harms was one of harm overall.

And that would include the assessment of all the issues, not just clinical validity. And the question specifically about clinical validity that you

were asking is -- could you just repeat that?

DR. CHUNG: Well, given that, you know, we are -- most of the studies may be focused on a single genetic test and doesn't account for unknown variables that may account for some of the variability we're seeing, the inconsistency. Then how do you -- how does EGAPP judge that?

DR. COATES: Right. I do believe that part of the evidence review looked at factors that may modify the relationship between the genetic variant and the -- and the outcome. And so for example, differences by race, ethnicity, or in terms of other medications that were being taken, that sort of thing.

Unfortunately, those pieces of information are commonly not included in any of the studies. And so -- and there were small samples in the studies as well. So I'm not sure they're actually -- in the evidence review, they're able to break down or look at whether or not these relationships, the predictiveness, was varied according to different categorizations that one could make in terms of

environmental factors.

DR. MC NEIL: Let's see. I saw somebody else. Cliff? Cliff and then Steve.

DR. GOODMAN: A question for Dr. Coates and --

DR. MC NEIL: Dr. Coates, why don't you make yourself comfortable right in the front row?

DR. GOODMAN: And Dr. Quinn, I think, as well. I want to ask a question that doesn't apply to any particular test. And that has to do with the type of basic requirement of study design.

In looking at the evidence requirements that EGAPP has developed and related frameworks What, if any, circumstances for establishing analytical validity or clinical validity -- clinical utility -- analytic validity and clinical validity would require an RCT? Are there any circumstances that would require an RCT to establish either analytical validity or clinical validity? And I will have the obvious follow-up question.

DR. COATES: I think there's a specific

table that -- in the Teutsch methods paper, the methods paper that was referenced there -- that describes the levels of evidence and categorizes the level of evidence for each of the -- each of the issues, analytic validity. And then there's another table that says how they put all of that together.

So I can't specifically recall whether there's -- clinical trial is required for analytic validity. I don't believe so. Or for clinical utility -- for clinical validity.

DR. GOODMAN: I would suspect the same. So can I just ask the follow-up question. If we knew from peer review of the literature that a group of well-defined patients, defined through their phenotype and/or their genotype, for that set of patients we have evidence from an RCT about what treatment works best, then insofar as your analytical framework is concerned in piecing together the evidence that we've heard, as long as we could get to that point going through the routes for analytical validity and clinical validity, we could get to that point without

RCTs, as it sounds like we may very well be able to do.

Then the only RCT evidence we might need is for clinical utility in a group that is already well-defined by non-RCT approaches. And we know that a drug will work or not work or a therapy will work or not work for that well-defined patient.

So in this analytic framework, I'm asking that if we can use -- if it's possible to use non-RCT efforts to get that far from your analytical framework and then show a good piece of RCT evidence that a well-defined patient group does better or not with therapy or therapy B, is that not a useful construct and a solid evidence chain?

DR. COATES: I think that would be considered by the working group to be good evidence. Yes.

DR. GOODMAN: Okay. Thank you.

DR. MC NEIL: Would that be the KRAS study that I keep coming back to?

DR. COATES: I'm not familiar enough with

the KRAS study to comment on that.

DR. MC NEIL: Okay.

DR. GOODMAN: That's helpful. Thank you.

DR. MC NEIL: Let's see. Steve and then Neil.

DR. PEARSON: Dr. Coates.

DR. MC NEIL: Maybe you want to be comfortably standing by the microphone.

DR. PEARSON: Put the chair right there. Also two questions. Sorry. One is about analytic validity again. The evidence review community that's used to thinking of diagnostic tests, they know how to think about concerns about interpreter training, the generalizability of academic centers to communities in practice, different patient spectrum.

When it comes to the analytic validity of these kinds of tests, even reading through the EGAPP material, can you help synthesize for us what are some of the evidentiary concerns that a review group should look at? I'm thinking about reproducibility, for instance. I mean, what are the standards for that

kind of evidence that CMS should look for?

DR. COATES: Well, I think many of these are lab-developed tests. And there's only one source of the -- of getting it done. And so I think one of the issues that CMS might look at is availability of evidence on the performance of that assay and that lab. And that's often not available. And so -- or sometimes not available. Sometimes it's published or not. And I think maybe going to the test providers and requesting that kind of information may be important.

With regard to -- and I think that would be validity and reliability. And the characteristics such as what percent of the assays provide no information. You know, they're essentially thrown out. Those kinds of issues. Since a lot of the assays are not done by several labs, then issues of how well they perform in different settings with different laboratories may be less important. But one issue might be how well they perform in that specific population for which they're being proposed.

DR. PEARSON: I would just suggest that to a certain extent, the credentialling process of laboratories is black box to clinician evidence reviewer-types. And so it's just very hard to know how much to be concerned about the actual, you know, result itself, how much trust we can have that it is a, quote, unquote, "true" result.

The other question I just wanted to bring back was this, again, kind of the somewhat surprising, as you said, result of your own evidence review on the SSRI dosing information. Since there was no even association of the genotype and the drug levels, let's assume that there were. That there had been evidence that the gene test did show definite correlation or association with drug levels. The next level of evidence would have been to ask, well, did that change physician behavior? Did it change the dosing of the drug? But I want to ask if that's kind of a cytology because if clinicians -- how would a manufacturer set up a trial to get that kind of information when clinicians aren't really going to know what to do with

this information unless they're told how to us it.

The internet algorithm that you showed said if you get X test result, you should reduce your dose by 20 percent. So in a sense, you have to set up a trial and give the information on how to use the test result in order to see if clinicians just follow your algorithm. It just seems a little bit circular. Do you have any comments about, if we're thinking that we would really like to see that this information changes clinician behavior, how do we disentangle that from the fact that we have to tell them how to use it in the first place?

DR. COATES: Well, I think that's a good question. I think what EGAPP was looking for is just observational information. Are there changes in practice based on using this test? Is it -- is there any information out there in the literature?

DR. MC NEIL: Could I just clarify, Steve, what you were asking?

DR. PEARSON: Sure.

DR. MC NEIL: And then Neil, I understand

you're next. When you said whether it changes physician behavior, do we want to think about it in the context of whether it changes physician behavior or whether it should change physician behavior?

DR. PEARSON: Well, that's an open question. I mean --

DR. MC NEIL: Do we want to address that at some point today? We don't have to do it right now. But is that an issue that should be on the table?

DR. PEARSON: Well, the SSRI is an interesting example. It's hypothetical obviously. But assuming that clinicians did receive a test result that they believed was true, would -- I mean, it's just hard to know what we're really interested in seeing.

DR. MC NEIL: Or the warfarin for example. That would be another one that might be a little cleaner since we know the results on that one.

DR. PEARSON: Well, the warfarin, too. Well, we don't know --

DR. MC NEIL: I mean, there's less ambiguity

in how the test works. Right?

DR. PEARSON: There is. Although, again, if clinicians receive a test result but are not guided towards how to use it by some --

DR. MC NEIL: Pretend they are. Pretend they are.

DR. PEARSON: Pretend they are by?

DR. MC NEIL: Pretend. A first year medical student in medical school, they learn that.

DR. PEARSON: It looks like Louizs wants to jump in.

DR. JACQUES: There may be of assistance -- sorry -- assistance to be provided here by looking at CMS regulation and the regulations surrounding reasonableness and necessity around diagnostic tests. Say it's a test that the treating physician uses in the management of the patient.

So if it's any help, it doesn't say that the physician ought to use if prudent or ought to use if doesn't want to get sued or something like that. There's a presumption that a physician would make a

reasonable use of the test. So if that helps somehow in the construct of your thinking around this.

DR. MC NEIL: Okay. So we have Neil, Mark.

DR. GRANT: I was just going to comment. I would say to change physician -- change physician behavior in the manner that has been demonstrably shown to improve outcomes. Whether it should or does. But however it does, has been shown to improve outcomes. That would be my --

DR. MC NEIL: Neil and then Jim.

DR. HOLTZMAN: This is a question or comment on two presentations, one by Dr. Quinn and the other by Dr. Mongillo.

Dr. Quinn, first of all, I noticed that you did cite a paper -- it's actually a set of three papers that appeared in JAMA in January of this year by Attia and Ioannides on how to interpret genetic association studies, a very valuable set of papers.

But the question I want to make is somewhat related to that, as you'll see -- is on your presentation. You talk about clinical decision point.

Now, there seems to be sort of general acceptance that although they're loosely defined, analytical and clinical validity are critical. But when we come to clinical utility, that's where there seems to be a lot of confusion or the term flexibility has been introduced. Ioannides who's a second author of those series of papers has done a number of studies on the bias that's introduced by single studies, particularly those that are the first to report an association.

And it seems to me -- and this came up, I guess, in the KRAS paper that Dr. McNeil had mentioned, that how do you guard against that? I mean, how long do you have to wait before you accept a single study that may be remarkable in its claim for benefit based on particular population studies but where we know now that very often when that happens, that the efforts to replicate that study are not successful? Maybe I should stop there and let you respond to that?

DR. QUINN: Yeah. I'm familiar with the work you're talking about that Dr. Ioannides looked at

dramatic first studies and then found how often they were replicated or found they were sometimes replicated more weakly. I think that -- you know, I actually just finished a 300-page Ph.D. thesis by Donna Messner which you get on the internet about the history of the FDA making safe and effective decisions, which was really interesting for me, having worked in the Medicare system.

Even after decades of refining the regulations and experience and special cases like accelerated approval, every case is different. You know, every ODAC is different. Every new drug or every major new drug is different. So I think you end up with people -- there are things that only happen in the human brain. And the wonderful example that's easy to stick in your mind is the Watson and Crick paper from 1952 where they had five lines of evidence suggesting DNA was a double helix. And I read a paper that said if you looked at any one of those lines of evidence, it was not definitive or credible in itself. But when you put the five of them together, and you

put it inside a human brain, everyone that read that paper believed DNA was a double helix.

So I think the reference here is there are some things, like there's actually been more than one KRAS study. But like -- the KRAS studies are so tightly linked to the drug never working if you've got a certain mutation, that people make that connection pretty quickly. And then there are other things like the warfarin genetics where there are so many things going on, different disorders that lead to warfarin, different genotypes, different height and weight, different diet, different conflicting drugs, different use of INR, different compliance. There's so many things going on that that step is a lot fuzzier, and people want to see more, you know, clinical applications. So I think my answer is I still case to case. You can't just say well, it's one study, or it's two studies.

DR. HOLTZMAN: Well, the point I was trying to make is that there does seem to be some bias based on how our journals accept and reject papers. That

papers with positive associations are more likely to be published than papers with negative associations.

So they get to be in there first, and of course, that stimulates efforts of other people to try and replicate those studies. And the more that happens -- and this is particularly true in the area that Ioannides looked at, which were genetic association studies -- that they fail.

So it's not a matter of what you may take in any particular study, but the fact is that maybe we have to sort of wait a while and see how these claims for associations work out. And they don't always work out.

DR. QUINN: That's hard to disagree with. I mean, this idea of overfitting curves is something you have to be sure there's one trial set and then one confirmation set or more than one confirmation set. When you violate that, then you get into overfitting problems that are legion.

I mean, this whole area, it's not that easy to track the logic behind these tests. I mean, Dr.

Gutman would be a world expert at that. But you know, sensitivity and specificity are sliding scales, they're spectrum effect. I mean, we could go on for 20 minutes about how much more complicated these are than the way we, you know, teach a first-year medical student.

DR. MC NEIL: Okay. We have Jim. Thank you very much.

DR. HOLTZMAN: I have a question for Dr. Mongillo, too. May I go on?

DR. MC NEIL: Oh, I'm sorry. You had another question? I'm sorry. Okay. Sure. For whom? Who's it for? I'm sorry. I missed it.

DR. HOLTZMAN: Mongillo. Is that how I pronounce it?

DR. MC NEIL: Oh, yes. Mongillo.

DR. HOLTZMAN: One of the things that you talked about was the difficulties of long term studies. You put it in the context of randomized clinical trials. I'm not sure it has to be. But if we're waiting to see whether a predictive test will

have a long term benefit in terms of the particular types of therapy that you might apply based on that genetic test, doesn't it seem incumbent to arrange some method of following the people -- it doesn't have to necessarily have to be randomized. But how can we be sure today to say, okay, here's a test that is going to predict the likelihood of metastases or recurrences in a 10 or 15 year time period without attempting to see whether that actually happens?

DR. MONGILLO: Absolutely. I mean, who could argue with the need to follow therapies, procedures, services, medical services, over time to determine the kinds of things you're talking about.

I think what we struggle with -- and the question came up when Dr. Pearson said it's a black box for physicians to know analytic validity for diagnostics or clinical validity. It's not. It shouldn't be. It may be. But it shouldn't be.

There are extremely well documented regulatory oversight procedures that have to do with CLIA, certainly for analytic validity. We think it

has clear direction for clinical validity. And it's a jumping-off point. I mean, that's really what we're talking about is making sure that there's strong analytic validity, clinical effectiveness, the term you want to use. And then of course, there's a need to follow. But that's what happens in medicine. I mean, I'm trying to think about other aspects of medicine, stents, back surgery, all sorts of things that are introduced into health care delivery with reasonable assumptions.

And then as time goes on, people say, well, maybe that wasn't exactly as effective as we thought it was going to be. So that's my response. Yes. Of course, we should look at over time. But it's a jumping-off point, and we think, you know, analytic validity is strong. Clinical validity should be strong. And then you can move on from that point.

DR. MC NEIL: Okay. Are we ready for Jim? No? You pass? Okay. Cliff and then Marion.

DR. GOODMAN: Yes. This maybe another Dr. Coates question. I'm not sure. It had to do with --

DR. MC NEIL: He's still there. Don't worry.

DR. GOODMAN: -- comments about the lack of -- the lack of relevant evidence in the peer reviewed literature. And I think you may have mentioned -- someone mentioned laboratory developed tests and so forth. And the concern is this, that our job, I think, is to help CMS provide some road map or ground rules for the kinds of evidence that will be expected for entry into national coverage determinations or other Medicare coverage decisions.

And so are we headed for trouble insofar as the kind of evidence that we'd like to get which has traditionally been peer reviewed versus where most of the evidence is -- much of the evidence is for laboratory developed tests that may not be in the public domain. To the extent that LDTs don't go through the premarket approval clearance processes that other tests go through, test kits and systems, are we going to be asking for some evidence that we're not going to get or that LDTs are going to have to

have developed for them in a non-traditional fashion, being put out into the public domain? Is that going to be a gap in the evidence that we're anticipating?

DR. COATES: Yes.

DR. MC NEIL: Well, that was short.

DR. GOODMAN: Thank you.

DR. MC NEIL: Marion, do you have a question that has a longer answer?

DR. DANIS: I wanted to ask a question of some of the clinical pathology folks who have presented because I'm thinking about, in the long range, the need to generate large data sets to help us move along in the collection of evidence. And it seems to me that the sort of thing that the pharmacogenetics arena makes possibly easy by the sort of data collection that Russ Teagarden talks about with huge numbers of pharmacy benefit programs collecting data for us, I don't know whether there's any kind of analogous kind of large scale cooperative collection in the pathology community.

And I'd like to hear some folks talk about

whether there is that capacity 'cause as we think about what levels of evidence we require, it seems to me important to know what levels of evidence we can acquire.

DR. MC NEIL: Well, that's a little bit -- that's an add-on to an earlier question. Okay. Who's going to -- is that Dr. Coates again?

DR. DANIS: No.

DR. MC NEIL: No. I'm sorry. It's the pathologists. So there we go.

DR. KLEIN: So as clinical pathologists, as molecular pathologists, we work very hard to try to acquire these types of evidence. But typically, we're not dealing with drugs. We're dealing with diagnostic tests that have somewhat limited financial implications. And so it can be challenging to get funding in order to do clinical translational related work. I mean, the NIH likes to fund basic science. But they don't like to fund us to tell how well our tests are working. And so in that respect, it's a funding gap to some extent.

And we do try to pull resources. We do attempt to assess these things. Quite frankly, the medical community does a pretty good job in terms of filtering out tests that haven't been at least demonstrated somewhat in the literature to have efficacy. I mean, clinicians don't like to order tests that don't help them. And so you could set up all the tests you wanted. If people don't think it's useful, they won't order them.

DR. DANIS: Yes. But in the absence of -- I mean, if we don't have enough evidence yet -- I mean, for instance, would coverage with evidence collection required by something that the pathology community could find -- I'm trying to sort through here.

DR. KLEIN: No. I think it's a reasonable suggestion. I mean, the problem is evidence. I guess what I would say is I think if it's funded, it's great. The problem is is that a typical molecular diagnostics lab in a hospital is not a money-making enterprise. And so we're happy just to get by. And so we don't have people to necessarily fill out all

kinds of paperwork and do all kinds of -- you know, it's hard enough to comply with the accreditation and the CLIA related regulatory burdens. If it's funded, I think everybody would love it.

But it's just -- you know, it's hard to place that burden on individual laboratories. And we have to -- you know, when you're looking at this area, you know, these high profile tests in Genomic Health and other assays, you know, Myriad Genetics, large companies, they get all the attention, but they're not the -- they're not the vast majority of the test menu.

The test menu is done in academia and in hospitals. And we're running small laboratories that usually are borderline staffed and don't make a lot of money. And so you have these -- you have these different competing concerns. I mean, a large public or venture funded enterprise maybe can afford to engage in the type of data collection and analysis that certainly, you know, a molecular pathologist would love to do but is going to have trouble getting funding and time to do.

DR. MC NEIL: Dr. Fowkes, do you want to respond to that?

DR. FOWKES: I just wanted to comment as far as some of the molecular testing that's done in our institution, when it comes to disease entities like some of the brain tumors. Oligodendrogliomas are not a very common tumor. The only large numbers that you accumulate over time are in institutions that have enough neuropathology, enough neurosurgeons, to have a large enough population to even have any numbers of oligodendroglioma. We have some in our institution, but there are not that many. Maybe 10 or 15 a year.

DR. DANIS: But I think that that's exactly the point I'm trying to ask you to help us think about in terms of what is the capacity to cooperatively collect data to get -- you know, you just don't see the kind of numbers you see when you're talking about everybody taking warfarin and getting their medicines from a --

DR. FOWKES: And yet the significance of 1-P, 19-Q co-deletion is enormous because you're talking

about potentially having just a simple biopsy with a little burr hole where you take a little bit of the tumor and you find out that this patient's going to respond to a chemotherapy drug, and prevent the open cranial procedure that has a lot of morbidity and mortality associated with it and hospital costs.

And yet the numbers are so small, the only way that you're going to be able to accumulate numbers that are going to be big enough for a true evidence-based heavily weighted paper is if you could combine information from all of the different institutions throughout the country that have large tumor banks or tumor -- brain tumor surgeries. And that isn't possible yet.

DR. MC NEIL: Now, why?

DR. DANIS: Well, why isn't that possible? That's what I'm asking about. That's the point of the question.

DR. MC NEIL: That's the question.

DR. FOWKES: Part of the problem is that each of the different individual institutions has --

there's no uniformity in pathology. There is, and there isn't. Neuropathologist on the west coast may have slightly different interpretations of a tumor than neuropathologist on the west coast (sic). It's a very sub -- it's a very small subpopulation of pathology.

A lot of the time, if there's any question about the diagnosis, the tissue gets sent to other pathologists in the country that have expertise in the field. They may agree, they may not agree. But the numbers, even throughout the entire country, are still small. Even if you pooled everything, it's small. And to be able to have everyone agree would be very difficult.

DR. MC NEIL: Well, maybe I could just interrupt because I think this is an extremely important point. And we understand your position, and I think we understand the position, I can just see by the nodding heads here, that that's probably not an acceptable answer in 2009.

DR. FOWKES: A lot of the -- a lot of --

what happens in an individual institution is that when you have a pathologist that makes a diagnosis, that pathologist is working in concert with the oncologist and the neurosurgeon and the clinician and evaluating all of the information. And that oncologist and surgeon and clinician knows what the pathologist means by that diagnosis.

DR. MC NEIL: No. We understand that. We understand that. We don't want an abstract from your hospital and your patients because that's not going to get us anywhere. I think the real issue is, if we're talking about genetic tests that apply to presumably smaller numbers of patients and require maybe ten institutions to get a few hundred patients to have any kind of the validity, whatever it is that we're talking about by this group, the issue is not that it's hard because we know it's hard.

The issue is going forward, how do we do it so that, for example, when this panel or an analogous panel meets in the future to discuss a particular genetic test, or when one of the evidence-based groups

evaluates data for a particular genetic test, they don't automatically say insufficient data. That's what we're getting at. And I know that that's not --

DR. DANIS: So what I'm asking is, if you got reimbursed for sending that sample and having, you know, a genetic test done with it, you know, so that there is -- you get reimbursed, but you also are collecting data at the same time. Is that something that would be manageable?

DR. FOWKES: It's -- the reimbursement and -- the numbers are still small for an individual institution. I think the only way to end up accumulating the data that we need is to have information technology where we can pool information from other institutions throughout the country and gather that data.

DR. MC NEIL: Just to follow this for a second, and then Mina.

DR. FOWKES: I mean even if there's a little bit of variability between different institutions, I think that the data collection with the genetic

information is still going to have some validity.

DR. MC NEIL: It strikes me that this is where some group, whether it's the NIH or another group, has got to figure out how these samples get put together and analyzed. And it certainly can't be a burden to your one institution, because frankly if you come up and say the data are compelling, that this particular test for a brain tumor are there on the basis of your five years' worth of experience on ten patients a year, you're not going to get a big hello. So that is the reality today.

What we have to do is think about how to proceed or how to develop an infrastructure. And it's probably none too early to be thinking about that in terms of the pathology groups, the various institutes at the NIH, the various genomic testing groups. We really have to start getting there. But we do need to move on.

So is this in direct response to this question?

MR. TEAGARDEN: Yes.

DR. MC NEIL: Because I have Jim and then I have Mina. Go ahead, please.

MR. TEAGARDEN: I think a lot of this information is going to turn up in places it hasn't been typically. And, I mean, the tumor tissue stuff is a little different story. But I can tell you that on the information that's relevant to drugs, we're going to see it turn up where it's collected as part of the benefit administration now and pharmacy practice generally.

So, for example, as I mentioned, we have MEDCO commercial programs now where we facilitate the testing of people for their warfarin and tamoxifen drug therapy. So we're collecting, you know, 2C9, 2C19 in some cases, VKORC1 and so forth. We're collecting that for individual people. And that is then reported to physicians. And then they're given various, you know, information on how to use it and so forth.

And Dr. Pearson, we are able to follow just whether they change doses within a particular period

of time. We can follow that.

Nonetheless, that information goes into the patient's profile, and it can be adjudicated and evaluated down the road. So if somebody tells us in the course of their tamoxifen therapy, we learn through a later 2D6, extensive metabolizer.

We know that now, and it's available there to evaluate, you know, with the appropriate --

DR. MC NEIL: How do you know that?

MR. TEAGARDEN: Because we get the test results.

DR. MC NEIL: At MEDCO?

MR. TEAGARDEN: Yeah.

DR. MC NEIL: Because you've asked for it before you fill a prescription?

MR. TEAGARDEN: No. It doesn't have to be before we fill a prescription. But we can contact the physician and the patient and see if they're interested in this testing. If they are, we can facilitate that through our partner lab.

The lab gets us, MEDCO, the test results as

well as the doc. MEDCO may or may not elaborate on test results depending on what the findings are.

DR. MC NEIL: So we probably should talk about this in greater detail later. But what percent -- pretend you ask patients to participate in this genetic test for --

MR. TEAGARDEN: It's a pretty high percent.

DR. MC NEIL: It's a very high percentage?

MR. TEAGARDEN: Yeah. You would be surprised.

DR. DANIS: A random sample or --

MR. TEAGARDEN: Well, there's -- we have a study with Mayo that's more of a study thing. We have commercial programs for the payers. We have over a hundred, like I said. Six million people have said we buy into the tamoxifen and warfarin testing. And they contract with us to do it. That's where get a lot of information.

But we do put it in the patient's profile, just like we put in allergy information. As a matter of fact, if we get information that's relevant to say

a phenotype on another drug, we shoot that information out to the doctor. So anyway, we have -- we are able to collect. There are other ways of getting it.

DR. MC NEIL: I think what she was asking is not what you're doing with the information. Are you getting a random sample of the information, or are you close?

MR. TEAGARDEN: Well, we're getting a sample of whoever is in these programs. It's not -- it's not -- we just take it all.

DR. DANIS: So it could be that it's people who are particularly having trouble with control of their INR. Right? I mean, or is it a --

MR. TEAGARDEN: That's what you need to know.

DR. DANIS: Yeah. We'll ask you about that later.

DR. MC NEIL: All right. Let's move on.

MR. TEAGARDEN: There's information available in more places. There's going to be more of it.

DR. MC NEIL: That would be great. So we have Jim, Mina, and I thought I said Catherine.

DR. PUKLIN: So I'd like to make a comment, which I hope is not irrelevant, as to why the clinical relevancy of all of these tests seems to be the field where there's no useful data. In other words there's no randomized control clinical trial.

And it has to do with the nature of the subject that we're discussing. If you're doing research, you will investigate in a randomized prospect of trial, something like a new drug or a surgical procedure or a new drug for cancer. And it will be done with a randomization protocol, where patients will be randomized between one test or another -- not a test, but one procedure or another procedure or another drug.

The tests that you're talking about don't get studied in that fashion. They're not randomized in any clinical trial. These are viewed by institutions as validating clinical tests. So there's nothing in the treatment of the patient that is at

stake when they're submitted as part of any protocol.

Validating clinical tests are done as an after thought on the patients that are being treated. And you've heard that discussed already this morning. Patients with cancer are going to be treated by their oncologists by the parameters of the clinical trials that they're participating in or by the standard of care. And somebody doing a clinical test, such as we've heard about, could be -- they could have varying tests for each institution studying the same things and competitive gene markers. But there's no randomization. It's what you get.

So the best way to get this information -- and I think this information is going to be coming because there's a whole host of trials sponsored by the National Cancer Institute, for example, wherein harvesting blood and tissue samples are part of the informed consent that the patient commits to regardless of the clinical protocol.

So now specimens are being stored in tissue banks, I suspect by the National Cancer Institute or

SWAG or COG, for genetic testing after the fact, after the patient's already been treated, which is why you end up having a lot of their anecdotal medicine and will never probably have randomized control clinical trial information, which is what we're actually looking for.

So I just wanted to bring that up. That's how the information is going to be obtained. That's how it's currently being obtained. But when we have a problem, for example, in our laboratory at the Detroit Medical Center, we have a large HIV population. And the director of the molecular biology lab found that the test -- the test he was doing to determine the viral loads, he was using a new product that he was sold, was showing that the viral load copies were much higher than they had been previously with the other test that he was using. So he wanted to investigate this.

But this wasn't a research project. This was a validated process. And so he simply ran his own trials, much as you would run trials on warfarin

treatment, to find out what genes are important.

Translating it from the validation of the test into a clinical setting is always going to require the jumping of that gap.

DR. MC NEIL: Okay. I'm looking at my watch here. Let's have quick questions or comments from Mina and Catherine, and then we'll decide what we're doing for the rest of the day.

DR. CHUNG: Just a comment about pathology not necessarily being able to gather repository data.

As you mentioned, there are large precedents already for that infrastructure on a national basis, through NIH, with national repositories already in place, data repository as well as bile repositories. And we have cardio thoracic surgery repositories.

I think the link -- the pathologists may not be the proper link in which to get the tissue because you have to consent the patient. But I think that coverage -- we should consider coverage issues for diagnostic tests, genetic tests, particularly in those who are part of a clinical trial, looking at long-term

outcomes or who maybe will contribute their data to some long term clinical trial outcomes trial.

Otherwise, we're not going to get that kind of data, as you already pointed out.

DR. MC NEIL: Catherine, did you have a final comment?

DR. ENG: We've heard a lot today about the potential benefits of genomic testing, but I'd like to hear some of the potential harms because I think that particularly in terms of diagnostic testing in a mass testing or without the input of physicians and then -- and even in the EGAPP, you know, framework, what part does harm, potential harm, play in recommendations?

DR. MC NEIL: Okay. Is that a question?

DR. ENG: Yes.

DR. MC NEIL: You're on again.

DR. COATES: I could just illustrate the kinds of harms from the reviews and recommendations that EGAPP has done so far. On the question of UGT 1A1, pharmacogenomic testing for the use of irinotecan with advanced colorectal cancer patients, the harm

that was identified there was that even though it appeared that one of the specific genomic tests could predict which patients might have more adverse side effects, it was unclear even in that group whether or not reducing the dose of the medication, chemotherapy, was beneficial to the patient because there was evidence of reduced efficacy of the chemotherapy in preventing recurrence of the cancer.

And they felt like there needed to be more information on the balance of benefits, that is reduced adverse events, versus harms, which were a greater likelihood of the recurrence of the cancer. And that wasn't clear from the kind of information that was available to them then.

On the issue of using the pharmacogenomic testing for SSRIs, one of the concerns was in a complex situation, where clinicians and the patients are trying to choose a particular test, and then doing dosing, that the literature there suggests that's a very complex and difficult process. That adding a test in that situation where there wasn't good

prediction, one would add cost to people obtaining, you know, who are being treated for common disorder. And that there could be harms because the prediction wasn't good. The prediction could result in giving the wrong dose of the wrong medication.

I'm trying to think now of the other one.

DR. MC NEIL: I think I've got the drift. I think we've got the drift. Good. Okay. Let's decide what we want to do. I think people are probably hungry. And Steve's the only one who took a break. So we have to respect the interest of others in taking a break. It's almost quarter of twelve. How long do you want to eat? (Unintelligible.)

We're going to go right now. We're going to go right now. The question is, what time do you want to come back?

UNKNOWN FEMALE VOICE: 12:30.

DR. MC NEIL: 12:30?

UNKNOWN FEMALE VOICE: 12:30

DR. MC NEIL: Let's aim for 12:30. And if there are stragglers, we'll go for 12:45.

(Whereupon, a luncheon recess was taken.)

DR. MC NEIL: Right now, on something called the initial open panel discussion. And just to clarify what that is, that's a discussion among the panel members. If a panel member wants to ask a question of a particular person in the audience, he or she is able to do that. However, it is not a time for individuals in the audience to make ad hoc comments unless asked. Their chance was already, and they gave us very, very good points about a number of different areas.

So I was trying to think about how to begin this discussion. And probably we should begin it by going right to the questions. There are a number of other issues that came up this morning regarding randomized trials are not in one study versus many studies and archived versus non-archived and inconsistent results and the use of judgement and long term versus short term effects.

All those issues came up this morning. And we addressed them a little bit. But I think it might

be better to talk right now about the questions. And then to the extent that we want to bring up other issues, we can.

So let's talk about the first one. Remember all of this is with an aim to answering questions four - four is the big one. Five and six are also important, but they may be a little bit more straight forward.

So the first question is, are the desirable characteristics of evidence for diagnostic genetic testing different from the desirable characteristics of diagnostic testing in general? That's the question, diagnostic testing. Yes, Cliff?

DR. GOODMAN: I was trying to think about this later this morning. And I sort of have a null hypothesis. And here's the null hypothesis.

DR. MC NEIL: Okay.

DR. GOODMAN: The Frybeck and Thornbury paper has been cited a few times this morning. It is probably a good, though imperfect, it's a good proxy for the desirable characteristics of diagnostic

testing in general.

So I would pose that we have not heard anything this morning that would push us off the Frybeck and Thornbury framework. And I'd love to hear from anyone what would push us off that framework that would tell us that there are different characteristics of evidence that we're seeking now.

DR. MC NEIL: That's a good way of starting. Thank you. Let's see.

DR. GUTMAN: Yeah. I also like that framework a lot. And I would argue that a test is a test is a test. And that where there might be a deviation, if there is any at all, would be in the highest hierarchy, the ethical issues because you are not only marking the patient, you have a potential to mark the patient in permanent ways that might damage their insurability or their employability. And you also might affect the patient's family. So I would argue a test is a test is a test. And maybe you need to be a little bit more sensitive to ethical issues.

DR. MC NEIL: Okay. Steve?

DR. PEARSON: I guess I would just point out I agree that schema makes sense. But it's not a list of desirable characteristics. It's just a list of characteristics of categorizing what we can look. So it doesn't really tell us whether we should expect level two or level three or level four.

And as -- I forget. Oh, it was Dr. Trikalinos whose presentation pointed out that when you do go looking for what evidence is available in radiology, a diagnostic test, it's almost all one and two. So part of our charge is obviously to say, is that desirable enough? I mean if that's desirable, if that's what there is for radiology, is that what we would expect for --

DR. GOODMAN: No. I meant to suggest I thought that they all six were desirable.

DR. PEARSON: Okay.

DR. MC NEIL: I assumed that what you meant, Cliff.

DR. GOODMAN: Yes, as a starting point.

DR. MC NEIL: As a starting point. Yes.

Neil?

DR. HOLTZMAN: Well, maybe the question of desirable doesn't really fit with what I'm about to say. I think there are differences that have to be taken into consideration.

One of the things that surprised me is there seems to be very little difference here between germ line genetic testing and somatic cell genetic testing. And the difference is from other tests I think come up much more in germ line testing because one is dealing with inherited characteristics. And the fact that whether for common or Mendelian diseases, one is making a discovery of a person's genotype as relevance for relatives.

And that -- Steve briefly mentioned that. But I think that has major concerns that raise issues of informed consent, before one does germ line genetic testing, such as in predispositions to cancer.

The other thing which is much more nebulous is that in many genetic tests for common diseases, one is dealing with relatively low predictivity. I mean,

we're not -- we talked about this a little bit this morning. We're not generally talking about an all or none or yes or no phenomenon from the test result.

Of course that applies to many other tests as well. I think as we approach the area of testing for common diseases, we have to recognize that seldom are we going to have a perfect fit between the test result and an outcome.

DR. MC NEIL: Okay.

DR. PHURROUGH: I'd like to ask Tom, if he doesn't mind, to spend a minute discussing why he was suggesting that ACCE has some benefits over Thorn -- Thorn -- yes, whoever -- Thornbury and Fryback.

DR. TRIKALINOS: So the ACCE framework specifically addresses all -- my personal opinion, specifically addresses all the details that have been brought up. It specifically addresses what happens to the patients. It specifically addresses what happens to the families of the patients and tries to put them all into a framework.

I did not show you all the questions that

are in the ACCE framework. I told you only that the ACCE framework has four components. And these four components map very well with the Frybeck framework. If you go through these questions, the 44 questions that ACCE has, you will see that they systematically went through the whole process.

I would not say that the Frybeck framework is not good enough. It's just that ACCE is specifically tailored to static genetic tests. And they have specific questions for many of the things that have been stated.

DR. DANIS: I was struck by your - you don't have to come up here. But just to say that it seemed to me that the translation or the ability to translate ACCE to the other framework was very helpful and that the questions that you're raising about germ line really reflect many of the ethical issues and don't necessarily undermine the extent to which you could argue that the desirable characteristics of a genetic test are like the others, like other tests in general.

DR. MC NEIL: So what I'm hearing is that, in general, a framework for testing is a framework for testing, with perhaps greater emphasis on ethical issues for some of the genetic tests, particularly if they're germ line mutations.

But that in general, Frybeck and Thornbury are good. And that the ACCE may make it a little bit more specific in some areas. But by and large we can - actually it's probably hard to quibble with this hierarchy. The details within the hierarchy might need more discussion.

DR. BERGTHOLD: Barbara, I thought we were somewhere saying that ACCE framework was better than the --

DR. MC NEIL: Well, I think that's what Tom was saying, that it provided more specificity.

DR. BERGTHOLD: Yeah. I mean I liked it. I was wondering if you thought if it would be helpful for us to make, you know, some kind of -- see the sense of the panel in terms of a recommendation or something.

DR. MC NEIL: Certainly. Can we do that without - what I guess what I was muttering about, Linda, was the fact that when I looked at the ACCE, A-C-C-E, they mapped in many ways to the components of Thornberry.

DR. BERGTHOLD: Yeah.

DR. MC NEIL: And there were different words with different subsections under them. So I was viewing them as essentially the same.

DR. SCHEUNER: I just wonder --

DR. MC NEIL: But maybe not. I'm sorry.

DR. SCHEUNER: I just wonder if the societal efficacy, which is the level six, really captures all of the ethical issues around the highly penetrant germ line mutation analysis issues.

DR. MC NEIL: Well, we haven't really talked about that. That's correct.

DR. SCHEUNER: That would be my only -- I mean, that's where I would prefer --

DR. MC NEIL: You would. Okay.

DR. SCHEUNER: -- the ACCE framework because

it's just broader in scope, I think, from what I can tell by looking at this.

DR. PEARSON: One thing that I view as a potential benefit of kind of keeping both frameworks in mind is that the ACCE evaluation doesn't really -- I mean one of the benefits that again we can do is to help manufacturers and clinical researchers understand what types of study designs will be viewed as most contributory towards our understanding of the clinical validity and clinical utility.

And so when you keep the different categories of the Frybeck and Thornberry, I think it just lends a certain specificity to that part of the discussion that's useful.

DR. MC NEIL: I actually agree. That's why I prefer the - but are we quibbling here? I think we are. No, you don't agree, Linda?

DR. BERGTHOLD: No. We're finished.

DR. MC NEIL: No. I mean I get the sense that we really think that they are pretty much the same. But, Eleanor, do you want to have a different

view of things?

DR. PERFETTO: My only comment would be not so much that any of the frameworks are better than any of the others because I think that we could use them in a way and any of them would work. But I guess I'm sitting here looking at them thinking if our question is whether any one of these frameworks has a characteristic that would make it apply only to genetic diagnostic testing and not to any other diagnostic testing, then we ought to point that out. But I don't see it.

I think it appears to apply to any kind of diagnostic testing. So our answer ends up being no, there isn't any specific characteristic that separates them.

DR. MC NEIL: Well, the only one was when Maren mentioned that maybe the ethical component is a little bit stronger in the ACCE.

DR. PERFETTO: Could be strengthened. But it's there.

DR. MC NEIL: But we haven't seen the

differences in either one of those. Yes, Cliff?

DR. GOODMAN: Just one that's especially relevant to Medicare, though, is, for the record, the Frybeck and Thornbury level six, societal impact does say cost effectiveness. I don't know that we're in a position now to say that CMS for Medicare coverage purposes is going to incorporate that element.

DR. MC NEIL: Well, that's true. That's a good point.

DR. PHURROUGH: Nothing prevents you from recommending to us that we, in fact, consider that. And whether that's different for genetic testing or diagnostic testing, you could opine on also. Whether we in fact take that recommendation or not is a different question.

UNKNOWN MALE VOICE: I was just cognizant of there have been some direction from Congress about when and when not to consider cost thus far. But we're going --

DR. PHURROUGH: Your job is to tell us what you think the evidentiary standards ought to be. And

we apply that to all the laws and regulations that we have to follow.

UNKNOWN MALE VOICE: Fair enough.

DR. MC NEIL: So hold on. Hold everything. Wait a second, Deborah. I've got you.

It strikes me as a little bit beyond the scope of this panel to start thinking about whether Medicare should be introducing cost effectiveness, isn't it, for diagnostic genetic tests, or is it?

UNKNOWN FEMALE VOICE: Probably.

DR. MC NEIL: No? I mean --

DR. PEARSON: If it's an element of evidence that should be viewed as important to the judgement of reasonable and necessary, I'm not sure why it would be outside our scope to comment or to suggest, certainly not to interpret.

DR. MC NEIL: Okay. Eleanor and then Teresa, I think -- I mean, Deborah. Sorry.

DR. SHATIN: That raises a question of what are we really saying here because we're looking at multiple, multiple items. Are we saying that for a

genetic test to have sufficient evidence, it has to meet all of level six through here?

DR. MC NEIL: No, no, no.

DR. SHATIN: Or are we saying that this is a reasonable standard to use? And then for specific tests, it might be a specific level of evidence?

DR. MC NEIL: The latter.

MS. SCHROEDER: I think that was kind of what I was getting to as well, is that if we've got a diabetic genetic test, it certainly doesn't need the RCTs that we might need. That as long as we have a framework, that subsets in that framework would be more equivalent to specific tests over others.

Some tests aren't going to need the cost effectiveness, which is a huge and also prohibitory cost for a lot of companies, when you look at the price of running a trial. But I think as long as we keep it narrowed, follow the same framework, but certain areas of the trial or the subject under genetic testing, whether it's diabetes or it's cancer, it would fit into certain of these categories.

DR. MC NEIL: Right.

MS. SCHROEDER: And then guide whether you need to do an RCT or a retrospective.

DR. MC NEIL: I wonder if you'd be willing for me to make a suggestion? Some of these questions are a little heavier duty than others. And this first one strikes me among the lightest. And I wonder if we've had enough discussion on it for the moment. Because we could spend a lot more time on this, but I'm not sure if it's the most fruitful use of the remaining couple of hours that we have, unless Mina, you have a compelling point.

DR. CHUNG: Well, I have a question in terms of whether or not it is allowable for us to consider the societal issue, especially in germ line tests that may have a larger impact on family than on the beneficiary.

DR. MC NEIL: So that's not this question yet. Right?

DR. CHUNG: But are we allowed to address that issue?

DR. PHURROUGH: You can make the recommendation that if we consider genetic tests, if you believe that all of those considerations should include this societal impact, that is an appropriate recommendation to make.

Whether we can do that or not is a separate question. If you think it's important, then it's certainly an appropriate recommendation.

DR. MC NEIL: All right. Moving up the scale of difficulty. Number two, what are the desirable characteristics of evidence for determining the analytic validity of genetic diagnostic tests? This is probably tougher. I know Maren has some feelings about this, and others do as well, so -- no comments? Yes?

DR. GUTMAN: Well, having just left FDA, I can tell you --

DR. MC NEIL: Yeah. That's good.

DR. GUTMAN: -- what we used to think over there, which is you want at least four things. You want some measure of accuracy or what's now referred

to by CLSI as trueness. You don't always get that. That's what you want in some comparison to a traceable reference material or method. When that's gone, then you want something that's a traceable tool working method. When that's not available, sometimes you make do with what you can. But you want some measure of trueness.

You always want -- you can sometimes scrimp on that. You always want robustness, precision, a measure of imprecision. You want proper stresses in the study of imprecision. So if it's a home test, you might not need to see it at three different sites. But you would still need to see different operators, different lots, different time environments to stress it, so you understand whether you're getting a consistent signal.

You want to know the specificity, and you want to expect that you'll miss it, and it will deteriorate with broader use. So you want to know the specificity of testing, how often interfering materials or substances or effects will cause false-

positive or false-negative results.

And then depending on how low you need to go, you might want to know the level of quantitation or the level of measurement. So you, at a minimum, want to know those four things.

I would actually add because of Carolyn Comp (phonetic), the troublemaker at NCI, you probably want to know something about the pre-analytical variables it might impact and analytical testing.

DR. MC NEIL: Okay. That was one of Maren's concerns.

DR. SCHEUNER: Right. And I just don't know to what extent I might -- it reflects more on the clinical validity. So some of those pre-analytic factors influence your interpretation of the test, more than what's happening in the laboratory.

DR. MC NEIL: Okay. Good point.

DR. SCHEUNER: And then I guess I would just say, you know, maybe when I brought this up earlier on the phone was the issues of again germ line mutation analysis versus somatic mutation analysis and getting

at that tissue. And if it's cancer, making sure there is enough tumor in that tissue to do the analysis.

And we even saw that with the oncotype DX example, where, I don't know, 15 percent of the time, they just didn't have enough there to do that.

DR. MC NEIL: So is that characterized by sample prep or something?

DR. SCHEUNER: So that would be a pre-analytic factor in terms of getting the right sample, how it's fixed, you know, if it's paraffin or fresh frozen. But I don't know that it's any different from anything else, but --

DR. PHURROUGH: Is the maturity of genetic analysis such that we could be comfortable that labs in general recognize -- and there are outliers -- labs in general can be assumed to have similar characteristics, have the same characteristics in their lab as we think are important across the field as a whole? Or is there some immaturity that would mean that some labs are challenged in coming up with accurate results?

DR. MC NEIL: Neil, do you have a comment to that or a thought?

DR. HOLTZMAN: To answer your question in one word, no. And I think this raises a very important point because particularly many of the tests that are developed today are in single labs. And the evidence that may be reviewed -- and I'm talking about analytical validity here -- maybe can be based on one or a small number of laboratories.

And they may meet all of the criteria that Steve talked about. And yet when the tests get out there, and these would be more diagnostic kits, that FDA would review and clear, where a test might stay in a single lab.

But the sort of gold standard of approval of testing in a clinical setting is proficiency testing. So that one would like to see any clinical test meet the standards that are set by some independent, outside, proficiency testing program.

So I think until we see that, we can't be confident that labs doing genetic tests, let alone

other genetic tests where proficiency testing is part of it - part of it's done by College of American Pathologists -- is met -- is satisfied.

DR. MC NEIL: Could I clarify, ask a point in that? So does that mean if hospital X develops its own test for something or other, that we should be reluctant to consider the validity of that test in the absence of some blessing by the American College of Pathologists or some other group?

DR. HOLTZMAN: Yes.

DR. MC NEIL: Yes? Okay. Cliff?

DR. GOODMAN: Yeah. A couple of points. The first thing is this question that we're dealing with is using this term analytic validity. And this, in fact, is one of the distinctions between Frybeck and Thornberry and the ACCE. I don't want to lose this point. Frybeck and Thornbury is very good at a high level with its six levels. Excuse me.

ACCE does maps to it pretty well. It's more detailed. And what ACCE introduces are, in fact, these three terms: analytic validity, clinical

validity, and clinical utility. Those three terms, while we might think they're applicable in other kinds of diagnostic modalities, had been defined, in fact, in terms of tests.

So we may want to state or consider stating or accepting the use of the terminology analytic validity, clinical validity, and clinical utility as important descriptors of evidence that CMS might require. Because those have, in fact, been defined for laboratory tests. Important point.

Now, when you get to each of these three points, in this case analytical validity, the points that Steve Gutman made were very good. And what this suggests is the following: as our friends at Blue Cross Blue Shield Association Tech do, one of their criteria for their evidence requirements is that the technology has been subject to or pass muster, if you will, with the applicable, regulatory authorities, which is typically the FDA. That's one of your five criteria.

And CMS might want to consider, when

describing evidence requirements for tests, that they have passed muster with the applicable regulatory authorities, which in this case is a little more complicated. It could be FDA and/or CLIA and/or a couple states. New York is one of them.

So I think that we want to think about aligning with some of the things Steve said, but we have to be a little more precise about our terminology here.

DR. GRANT: Can I answer?

DR. MC NEIL: Go.

DR. GRANT: I don't necessarily disagree, but I'm not sure that in this case that conforming to CLIA is probably -- it's good, but we've had many discussions where it really doesn't meet the evidentiary needs for somebody making decisions about the benefit or harms of a potential test.

DR. MC NEIL: So you like the FDA better, Mark?

DR. GOODMAN: I'm not vouching for FDA versus CLIA.

DR. MC NEIL: Well, Mark was.

DR. GOODMAN: Mark was. But I am saying that there is regulatory oversight of tests in the form of FDA CLIA in certain states. And we might want to take a lead from other technologists, some efforts that site a minimum threshold for passing regulatory requirements. Then we can get into the details about whether we think CLIA suffices or not. That's a separate issue.

DR. MC NEIL: Could I just ask Neil -- I want to ask you, does that differ from your proficiency issue? It does. You first have to have the regulatory oversight. Then you have to have measured proficiency in the sites for the test for which regulatory authority has been obtained. Is that correct?

DR. HOLTZMAN: Yes and no.

DR. MC NEIL: Oh, okay.

DR. HOLTZMAN: I'll get to it.

I think there is a very important distinction between CLIA and FDA. Because the tests

that FDA must see -- Steve can amplify or correct me on this -- are tests that are essentially being marketed as kits, where they will be used in a wide variety of laboratories and, therefore, must meet FDA approval. FDA approval includes analytical and clinical validity.

Now, the tests that CLIA deals with are tests that are provided by a single laboratory and marketed to the public from that laboratory. One of the things that CLIA will expect is proficiency testing. FDA may not deal with that on an ongoing basis because although they look at clinical and analytical validity, they are looking at it only as a single kit at one point in time, when they give pre-market approval, usually.

So I think that's why I say yes and no. If it's a CLIA test, proficiency testing and other aspects of analytical validity become much more important. The great weakness of clinical -- of CLIA's involvement with test is that they don't look at clinical validity.

DR. MC NEIL: Okay. Yes, Steve?

DR. GUTMAN: Well, I'm not representing FDA. I'm an ex-employee. But I do know and love the place. And the deal here is the analytical expectation between FDA and between CLIA are extremely similar. What is different is process and transparency of the two -- of what's going on.

So FDA will get a submission and will review it from soup to nuts. Some people will say we're really nuts. Some people will say we're toast. But the deal is, we'll review it from soup to nuts. We have independent review of data, independent review of claims, independent review of labeling. And then that's all posted in a very transparent way on our web page.

So all our reviews are public. You can swear at us. You can swear by us. But you can't say we do anything in the cover of night.

Under CLIA you get an inspector, whether it's a CAP or a CLIA or a COLA. An inspector comes in, they review the lab safety manual, the

documentation showing education. They review the menu of tests. They review specimen requirements. They review temperature charts. They review procedure manuals. And at the same time, they review the analytical performance of the test. And then it's non-public.

So there are striking differences in process and in transparency. There is theoretically, at least on paper, no difference in performance requirements.

DR. MC NEIL: Okay. Marion, and then I'd like to make a summary and see if I can catch where we are.

DR. DANIS: I'm struck, in looking at the wording of the question, that we're not supposed to just talk about what parameters we're concerned about with regard to analytical validity, but the characteristics of the evidence that substantiate that. And I was going to suggest that maybe some of the tables -- EGAPP criteria for degrees of evidence would be something we might want to adopt, rather than reinventing the wheel. Because if you look at the

tables in the Teutsch -- the article again --

DR. MC NEIL: The Teutsch article?

DR. DANIS: Yeah.

DR. MC NEIL: That we don't have?

DR. DANIS: No. We have it.

DR. MC NEIL: Oh, okay.

DR. DANIS: It's the general overview one on the evaluation of genomic applications and practice.

DR. MC NEIL: Yeah.

DR. DANIS: And it strikes me that there is there a whole table with levels of the hierarchy of data sources and whether they are considered, you know, for analytic validity, whether they're adequate for adoption of the test or - it just strikes me that we could build on that very well. And it would serve the purpose for addressing this question.

DR. MC NEIL: So then Marion, you would be talking about table three, four, and five. Right?

DR. DANIS: Tables three, four -- yeah.

DR. MC NEIL: Okay.

DR. DANIS: And whether the information is

convincing, adequate, or inadequate, or judging the analytic validity, and the clinical validity, and anything further.

DR. MC NEIL: So how should we proceed? Should we - we've mentioned some generic items, which dealt with accuracy, precision, specificity, level of measurement.

DR. DANIS: Yeah. So for instance --

DR. MC NEIL: I thought regulation, proficiency, that kind of thing.

DR. DANIS: Yeah. So whether the data comes from, you know, a collaborative study using a large panel of well characterized samples and that it has to have summary data from well designed external proficiency testing. I mean it seems to me that those are the sort of things we would want to articulate.

DR. MC NEIL: Right. Right. Well, is there any reason to think that if EGAPP has spent months and years developing these criteria that we could improve them?

DR. ENG: I would second the recommendation

to follow these tables. With the caveat that I think it was mentioned today that the field is moving very fast, and the question is whether there is flexibility in that process. They've spent a lot of time and effort in making sure that it's a good platform.

DR. MC NEIL: No. Actually, that's a good point. We heard that from several speakers. And that's definitely not one of the -- that's not written here. Well, pretend we agree to flexibility. Neil, did you want to --

DR. HOLTZMAN: I think these tables three, four and five are excellent tables, not specifically related to alternating issues. Or if they are, there's one that's missing. And that is the Hardy-Weinberg equilibrium. If you're doing a germ line study -- and this is widely accepted in the Teutsch Study -- that one wants to see whether the data fit with Hardy-Weinberg equilibrium. So that's just another point. If you're dealing with genetics, that's where there's a difference.

DR. MC NEIL: So this is supposed to be for

genetic tests. Right?

DR. HOLTZMAN: Yeah.

DR. MC NEIL: So how did they miss that? Unless it's implicit.

DR. HOLTZMAN: Quickly, I don't see it.

DR. MC NEIL: All right. We take the point that Hardy-Weinberg should be --

DR. PHURROUGH: (Inaudible.)

DR. MC NEIL: Right. Right. That's true. That's true.

Steve was just saying if that's not considered, then the study can't be well designed. And one of their considerations is a well designed study. So we'll put Hardy-Weinberg in there.

Are there other issues then? Thank you, Marion, for pointing this out. This is a good place to --

DR. DANIS: Just one other twist. I think it was mentioned this morning of the recent articles in JAMA that were really terrific.

DR. MC NEIL: Right.

DR. DANIS: And they have some critical appraisal guidelines that I think we might as well adopt too.

DR. MC NEIL: They do.

DR. DANIS: And they refer to --

DR. MC NEIL: Steve, did you have a comment?

DR. PHURROUGH: Well, I just want to clarify. We're just talking about the analytic validity part of these tables right now.

DR. MC NEIL: Correct. We're on question two.

DR. PHURROUGH: Good.

DR. MC NEIL: We're on question two, left hand column. Okay.

Why don't we, for the sake of argument, I think we may want to iterate on some of these questions after we've considered the next more difficult one. But let's for the moment assume that we know what we're doing with question one, we know what we're doing with question two. That's not to say we won't refine them. But let's just take as a given

that that's our baseline and move on to what I think is a much tougher one, which is question three. And tougher still is question four.

So you can read three yourself. But going beyond the analytic variability, what are the differences for each of the three considerations below: diagnosis, prognosis, or pharmacogenetic assessment?

So why don't we do the one by one? Why don't we think about what we should be considering beyond analytic validity for diagnostic testing, using genetic tests?

DR. PHURROUGH: And I think you could read that to say you are likely, as you get to the final on question two, to suggest these three tables. Question three says, are these three tables sufficient for more than analytic validity? Are they sufficient for clinical validity? And do you apply them differently for each of these three categories?

DR. MC NEIL: Oh, good way. Okay. Because we have clinical validity here. So maybe that would

be the place to start. Clinical validity for diagnostic test, do you think, Steve?

DR. PHURROUGH: When those tables apply --

DR. MC NEIL: Right.

DR. PHURROUGH: -- the same as they did for analytic validity and clinical validity of these three categories.

DR. MC NEIL: Yes, Neil?

DR. HOLTZMAN: Well, the first thing in looking at table four under clinical validity, a clear description into the disorder phenotype and outcomes and interests. Well, so far as the description of the disorder --

DR. MC NEIL: I'm sorry. Where are you?

UNKNOWN FEMALE VOICE: Table four.

UNKNOWN FEMALE VOICE: Page nine.

DR. MC NEIL: Oh, that's the second part of table four.

DR. HOLTZMAN: Yeah, under clinical validity.

DR. MC NEIL: Yeah.

DR. HOLTZMAN: Well, I think this matter of -- and it's not going to be a set answer to it -- but the matter of disease that originated is critical.

To give one classical example, there was a paper in nature in the 1990s that claimed a strong association between bipolar affective disorder and a specific gene type. And then as more patients were studied, or the patients were followed up longer, some of those who were alleged to have bipolar, did not have it. And others in those families did have it. And it rendered the original association of non specificity significance.

So this is a critical issue and not always easy to resolve. If you look at Diabetes Type II now, it's now becoming a set of diseases for which the genotypes may apply to one, but not to the other and how one makes that distinction. And, in fact, if one could not make that distinction, there seems to be heterogenia out there, one has to be very careful accepting evidence of association or non-association.

DR. MC NEIL: Okay.

DR. DANIS: Can I just clarify what you were saying? From what I hear you saying, that underscores the importance of the criterion that's listed here in table four, i.e. that you have a clear description of the disease, and that you have a clear idea about who you're going to include and exclude. Right? Is that right? You're supporting this?

DR. HOLTZMAN: Definitely. Oh, yeah. And I think to make it clearer, one of the sort of things that might lead to homogenic, you're dealing with common diseases. So one might not necessarily want to include all age groups. So in other words, if the age of onset of the disease is early, that tends to be, very loosely, more likely to have genetic components inherited components in late onset disease.

There may be gender differences. There may be specific responses to therapy already that make distinctions. But there are a number of parameters that one can look at to subdivide a broad category of disease into smaller categories of diseases in trying to look for gene associations. And I'm talking here

about germ line.

DR. MC NEIL: So other comments on clinical validity, ala Steve Teutsch, and diagnostic assessment?

DR. GOODMAN: I don't see it under the column of clinical validity, but are we not interested in understanding the risk of being wrong here? Depending upon what kind of condition you've got, you maybe want to be extra sure, based on stronger evidence, that the test is going to be right. There's a price to be paid for obviously false-positive negatives.

DR. MC NEIL: They don't have that here. Do they?

DR. GOODMAN: So I'm just wondering if we need to reflect potential benefit and risk for a patient with a given set of problems.

DR. MC NEIL: I think it may be here.

UNKNOWN FEMALE VOICE: Wouldn't that be utility?

DR. GRANT: That would be clinical utility.

I mean the cost of false-positive and false-negative really falls under weighing risks and harms.

DR. MC NEIL: Well, no. There's a specific specificity to a test.

DR. GRANT: Well, it's here. It's here. That's listed.

DR. MC NEIL: From the second part of table four, second one from the bottom.

UNKNOWN FEMALE VOICE: Yes.

DR. GOODMAN: But I'm not talking about what happens, you know, whether you treat the patient or not.

UNKNOWN MALE VOICE: Right.

DR. GOODMAN: I mean if you get wrong information at this stage, given certain problems, the patient could be in more trouble.

DR. SCHEUNER: That's an outcome.

DR. MC NEIL: But doesn't that move on to the next column?

DR. SCHEUNER: Yes. That's an outcome. Well, alternatively --

DR. PEARSON: Sorry. I agree with Cliff, that there's some difficulty in the fact. I think of sensitivity and specificity as related directly to clinical validity and not to utility. Utility would be how you use the information to guide patient care, but the actual true positive, true negative, false-negative, those in my mind would be validity.

DR. SCHEUNER: Absolutely. I think we're agreeing with you that those are measures of -- the clinical sensitivity and specificity are measures of how well the test is going to predict your phenotype or outcome. But if you -- how you use that information, if wrong decisions are made as a result of a test that has poor specificity or sensitivity, that's the utility of it.

DR. MC NEIL: Right. So the specificity properly, I agree, belongs with the validity component. Do you disagree with that, Cliff?

DR. GOODMAN: That's what I was --

DR. MC NEIL: Oh, but it's here. It's under validity.

UNKNOWN FEMALE VOICE: It is.

DR. PEARSON: What's interesting, though, is in table three, the types of studies and the levels of evidence seem less tied to, in a sense, test accuracy of that type. We're talking about case control studies and other things. It's not really liked.

DR. GOODMAN: What I'm saying, Steve, is that you may want to roll up higher in the hierarchy under the clinical validity study types depending upon the patient's situation. And we heard some discussion about that this morning, about maybe having sliding requirements, depending upon the inherent risks.

DR. MC NEIL: So you think, Cliff, though, when we move into clinical utility that that will roll it up enough for you? No?

DR. GOODMAN: I'm not yet persuaded. But I'd be glad to be.

DR. MC NEIL: All right. Maren, you looked like you were nodding against that idea first, and then Marion.

DR. SCHEUNER: I'm not quite sure because

I'm not sure I'm following the conversation. But I just think for me the distinction between analytic validity, clinical validity, and clinical utility is rather clear. But I guess is what I'm hearing at this end of the table is that issues around clinical validity -- I don't know. I don't know.

DR. MC NEIL: Marion -- I'm not sure, I guess I'm a little -- and then Mark.

DR. DANIS: I think sensitivity and specificity are relevant for assessing clinical validity. It seems to me, just as you were saying, that we should take the concern about risks associated with a diagnosis and use that as the criterion for how far up or down you go on what level of certainty you demand, you know, or how --

DR. MC NEIL: Sure. Right. Mark?

DR. GRANT: Just a couple comments. I think test accuracy, you know you view it in and of itself, according to the characteristics of the test. How you weigh the consequences of the test has to do with the benefit or utility, which is I think more what you're

getting at. So I would tend strongly to take that view.

The other piece that's not here, and maybe it's a little quibbled, but just from reading a lot of diagnostic studies papers, is that the way sensitivities and specificities are set oftentimes varies. And I have particular difficulty when it's data driven. And the manner in which it's --

DR. MC NEIL: What do you mean when it's data driven?

DR. GRANT: The point estimates of clinical sensitivity and specificity in table four, yeah, that's good to have. But how you get there, you know, when they pick optimal cut offs, that's difficult -- and how uncertainty is stated. So I would add that. It's a little piece, but --

DR. MC NEIL: That's a good point, Mark. It might be reasonable for us to say that point estimates of sensitivity and specificity are useful. But more useful would be ROC curves, if available.

DR. GRANT: I like classification tables.

DR. MC NEIL: All right. Whatever. I mean it's the same.

DR. GRANT: Yeah. Whatever. But that's -- yeah. But go ahead.

DR. SCHEUNER: Isn't that particularly important when we talk about any kind of result that has some kind of prediction?

DR. MC NEIL: Well, we're not there yet. It is, but we're not on prediction yet. We're still on diagnosis.

DR. SCHEUNER: We're still on diagnosis.

DR. MC NEIL: We're going to get to prediction because it's a much bigger deal there.

Could I make this suggestion that for the first -- unless there's something that people don't agree with on this table, that we move on to prognosis, because I think that this table doesn't exactly apply in the same ways. And we're going to want to make some suggestions here, at least I think we are. That is genetic tests for prognosis.

I mean, I see one thing that I think is

missing is follow-up. We've talked a lot about follow-up, short term, long term. And depending upon what we're looking for, that's probably required when we're looking at prognostic tests and the length the follow-up would be.

DR. DANIS: Well, it does say well designed. Don't you assume that longitudinal --

DR. MC NEIL: It depends upon how long, how long.

DR. DANIS: Yeah.

DR. MC NEIL: I guess I was asking for more detail on longitudinal.

DR. GRANT: I would agree, but I was just going to add, if I could?

DR. MC NEIL: Sure.

DR. GRANT: I mean in general for prognostic studies, one would always like to have not just a measure of test accuracy or discrimination, discriminatory ability, but also how well it's calibrated. I mentioned that before. I think those are absolute requirements for prognostic data.

DR. MC NEIL: Now, that's not here.

DR. GRANT: And that's not there yet.

DR. MC NEIL: Calibration is excellent -- excellent. But that's not for just genetic tests. That would be for anything.

Now, one of the commentators this morning raised the issue of prognosis versus prediction. I don't remember who. You, Mitchell. Does anybody have any feelings about that here? Because I'm not exactly sure how to respond to that.

UNKNOWN MALE VOICE: (Inaudible.)

DR. MC NEIL: Pardon?

DR. GRANT: The difference between prognosis and prediction --

DR. MC NEIL: In this particular context. Yeah.

DR. GRANT: -- in this context is prediction is predicting response to therapy. I think generally it's accepted in the usage here, as the pharmacogenomic assessment. Whereas prognosis is determining what the future outcomes are likely to be.

DR. MC NEIL: Is that what you meant, Mitchell?

DR. BURKEN: Yes. The hallmark of a predicted test again, as Dr. Grant said, is one where you would be able to see if the by marker would influence some outcome in therapy.

DR. GRANT: I just want to add one other piece that's not explicit here, but it has been discussed along the way, too. And that's a distinction between when the test is to add incremental value, or there is a substitute and how that evidence is portrayed and how that allows one to interpret a decision in an informative manner.

I think that is very important to be explicit about because obviously, it's not just sensitivity or specificity. These are the incremental values.

DR. MC NEIL: You know, that's probably one of the most complicated - you and I know that from Blue Cross.

DR. GRANT: Yeah.

DR. MC NEIL: But that's been one of the most complicated areas to tease out, looking at the incremental value of a test in terms of prognostic ability. And that should definitely be here. Recognizing that is an important consideration, and should definitely be here, because it's not well understood.

DR. GRANT: For example, one of the things -- I think it's been applied to a lot of the tests of late to determine incremental value, you know, for clinical validity -- would demand classification and reclassification to allow interpretability and to translate them to the likely consequences on benefit or harm. And that isn't here.

DR. MC NEIL: Right. I mean, in that particular case, we're always looking at patients who change they're --

DR. GRANT: Right.

DR. MC NEIL: -- who change the cell they're in as a result of having this new information. And that's not always possible from a lot of the studies.

That's not always possible to obtain that from the studies that are published.

DR. DANIS: So we're saying that if someone wants to get coverage for a test they're claiming is of incremental value, they'd have to show evidence that some reclassification happened?

DR. MC NEIL: Yes. Yes.

DR. GRANT: And not just that reclassification in terms of clinical validity that will get there, but also how the consequences for benefit and harms. So it's not just that you can reclassify correctly two percent or four percent, but in fact, they have clinical consequences that are meaningful on whatever metric one chooses to measure.

DR. MC NEIL: So today, actually, the warfarin would have been an example where - would that have been an example where --

DR. SCHEUNER: If you had an algorithm --

DR. MC NEIL: Right.

DR. SCHEUNER: -- that didn't have the genetic markers, and you were trying to predict dose,

and then you added the genetic markers, were you able to reclassify subjects enough --

DR. MC NEIL: Right.

DR. SCHEUNER: -- to make a difference?

DR. MC NEIL: Right.

DR. SCHEUNER: But we didn't really have that pre - I think with the oncotype DX, you could say this is the historic, clinical way of classifying subjects to be at low, intermediate, or high risk. And now we're going to throw on top of that these genetic markers. And do we reclassify enough subjects to make a difference?

DR. MC NEIL: And they did that is my understanding.

DR. SCHEUNER: Yes. They did that.

DR. GRANT: Ultimately.

DR. CHUNG: But is this holding genetic tests to a higher standard than other diagnostic tests, where we don't require incremental information and maybe except equivalence?

DR. MC NEIL: Maybe. But that's probably

the way we should be going.

DR. PHURROUGH: And I guess maybe the CMS party line here is we should not attempt to compare what kinds of recommendations you're making here today on the diagnostic genetic testing to other kinds of diagnostic testing, where we, in fact, have not established standards. So we may be saying we're applying higher standards, but because we don't have standards currently, you should not attempt to determine whether these are too high compared to others.

DR. GRANT: I would make a very strong case that in fact these standards are not so high. In fact, they're minimal. And there really is no other means to evaluate benefit and risk without knowing how patient's treatments are subsequently changed. So the traditional approach is -- and there was a comment before about -- in the EGAPP odds ratio -- is that they're not really informative. They're not informative about classification, like the diagnostic likely ratio. There might be other ones.

But there are other ways to do it. But I would argue strongly that this is no different than what we would like to see across the board for any diagnostic test.

DR. MC NEIL: And if we had a clean slate, we'd apply it for any diagnostic tests. Neil and then Steve.

DR. HOLTZMAN: Well, in the context in which we've been told to deal this distinction between diagnosis and prognosis is a little problematic because you had told this morning that we're dealing with people who have already symptoms and signs of disease. So in diagnosis, we're certainly not predicting. We're asking the question of whether given a person with a constellation of signs and symptoms, does the diagnostic test help us in any way. And then that seems to lead directly into prognosis.

If you have a person with symptoms and signs, and diagnostic genetic test tells you they've got a certain genotype, then the value of that information is not saying whether or not the patient

has the disease, we already established that, but whether it will help us in terms of designing the therapeutic regimen for that person. And that's prognosis.

In other words, if you have the genotype and the disease, will having that genotype influence your decision of what kind of therapy to use? And that's where questions of prediction --

DR. MC NEIL: Right.

DR. HOLTZMAN: -- and clinical validity come up.

DR. PHURROUGH: Neil, won't there be -- for a diagnostic test, I think the assumption in most cases would be there is some prognostic value to that diagnosis. But there may be some other test which have no diagnostic value that have prognostic value.

DR. HOLTZMAN: Correct. The first category you're talking about is not a clean line.

DR. PHURROUGH: Yes.

DR. MC NEIL: Steve, I think, was next and then Maren.

DR. GUTMAN: Yeah. Well, again from the FDA perspective, diagnostic actually refers to all uses of the tests. So screening would be diagnostic. Actually making a specific diagnosis would be a diagnostic. Prognosis would be a diagnostic. Predictive testing would be a diagnostic.

So I'm assuming here that diagnostic is intended to make a specific diagnosis in this case based on the ground rules --

DR. MC NEIL: We're not talking about screening tests here.

DR. GUTMAN: No. No. In the ground rules in the symptomatic patient.

DR. MC NEIL: Ground rules diagnosis is diagnosis.

DR. GUTMAN: But I guess I'd like to point out that either prognosis or diagnostic tests could be viewed as more adjunctive or as more stand alone. It speaks to the issue of reclassification and understanding how.

So while I certainly have no argument

against looking at a prognostic test in terms of its ability to correctly reclassify a patient, I would argue that the same should then apply to the original diagnostic test that's creating the diagnosis at hand. The classification, how that classifies patients, might also be interesting.

DR. MC NEIL: Oh, I think that's correct. Maren?

DR. SCHEUNER: I guess the only other thing as a medical geneticist ordering genetic tests, often it's to arrive at a specific Mendelian disorder diagnosis. So I don't know to what extent that's maybe -- I mean here it's not so much about -- I guess it influences prognosis because now you have a better sense of the natural history because you've been able to pinpoint what's going on in that patient.

Whether it's familial amyloidosis or porphyria or hemochromatosis, there are six thousand single genus orders, some of which affect the Medicare population. So I guess I --

DR. MC NEIL: So what's your point?

DR. SCHEUNER: I don't know how you feel down at that end of the table about what I just said.

DR. MC NEIL: I'm not sure, Maren, I get your point.

DR. SCHEUNER: Well, I don't know if -- I think here it's -- I don't know if you could talk about reclassification, for example, when you're ordering a test to know is this Von Hippel-Lindau disease or not.

DR. MC NEIL: Well, we're in prognosis right now.

DR. SCHEUNER: He just mentioned diagnosis and reclassification as well. And I wasn't quite sure that the two always go hand in hand. Steve just mentioned that, you know, maybe even with the diagnostic tests, we should apply the same reclassification.

DR. MC NEIL: Right. He mentioned that diagnostic test should be associated with the same kind of reclassification issue where you're doing a diagnostic test, and you're wondering whether you're

reclassifying patients differently compared to the diagnosis you had when you just had clinical signs and symptoms, for example.

DR. SCHEUNER: Right. Where you have like, you know, a handful of --

DR. MC NEIL: Right.

DR. SCHEUNER: -- signs and symptoms, and you're wondering could this all be attributable to this one diagnosis, or am I really dealing with all these different things just by chance.

DR. MC NEIL: But I just want to keep our eye on the ball here. Is this relevant to our discussion right now on prognosis?

DR. SCHEUNER: Perhaps not to prognosis. I'm sorry. I was thinking more of diagnosis.

DR. MC NEIL: So let's stick with prognosis. If we want to go back to diagnosis, we can. But I really want to keep - we can get confused enough with these categories. So let's stick with prognosis. And then we can go back and rewind, if you want. So Mina and then Catherine.

DR. CHUNG: So if I'm understanding these points, reclassification would have something to do with prognosis.

DR. MC NEIL: Absolutely.

DR. CHUNG: And I would like to bring up one other recent example of a bio marker that lost its coverage, ultra sensitive C reactive protein, where it was probably felt to be equivalent to conventional lipid testing. And subsequent data comes out supporting that there is some incremental benefit even in low LDL subjects to actually treat based on ultra sensitive C reactive protein levels.

Now, genetic testing, as I said earlier, and we've pointed out earlier, with some of the inconsistent data and all the unknowns that we have, a genetic test may have looked very promising in some earlier studies. Then you get these inconsistent results. But it is not clear with the current data why there are inconsistencies.

And that's going to happen with any diagnostic test, but I would not like us to refuse or

advocate not covering that test because we have so many unknowns that perhaps we will be actually inhibiting the progress of personalized medicine if we don't cover. And you then squelch a lot of research that may go on to try and determine why there are inconsistencies.

DR. MC NEIL: So just to be clear, we're not making coverage decisions here.

DR. CHUNG: Right. I understand. Or recommendations.

DR. MC NEIL: Or recommendations. We're trying to identify the criteria that should be used for prognostic or diagnostic tests. What you just said, in some sense, contradicts most of what we've said so far, where the emphasis has been on having several studies as a good measure of the value of a test from diagnostic and prognostic viewpoints.

I inferred from your remarks that a couple of one good test, you shouldn't throw the baby out with the bath water because you'll then preclude other good research studies validating that.

DR. CHUNG: Well, I think we all desire multiple tests to confirm. So I don't want to imply that at all.

DR. MC NEIL: Okay. Good. Catherine, did you have something?

DR. ENG: I wanted to comment about the necessary characteristics of the evidence. This morning I heard there are common diseases, and then there are rarer conditions. And I think that there should be a higher level of evidence in the clinical testing for more common conditions --

DR. MC NEIL: More demanding.

DR. ENG: -- more studies and randomized control. And I'm thinking about a very common condition in the Medicare beneficiary population, which is dementia, Alzheimer's, dementia.

And so I think the more common condition in terms of prognostic, I think we have to have a higher level of evidence. Now, if there is an uncommon condition, and somebody finds a genetic association, there may be a reason to have less. In other words

cover that, without going through hundreds of randomized control studies and so forth.

DR. MC NEIL: Mark?

DR. GRANT: If I could just sort of amplify some of those comments? I think that every decision has a different degree of uncertainty accompanying it. I mean a level of uncertainty of which one adopts or does not. And it will vary according to multiple factors.

And, in fact, the one thing that's missing from here, although it weaves throughout all the criteria, is that a lot of this is all about quantifying uncertainty to provide information for the decision makers. It's whether how much bias is there and then the generalized ability, that's uncertainty when this is disseminated.

And the only piece here really has to do with point estimates, which really is just a sampling issue. But I think that the thresholds for decision making are going to vary according to multiple factors. And they should. They should.

DR. MC NEIL: Can I just regroup for a second to make sure I know where we are?

We're on prognosis, and we're talking about clinical validity. And so far, I've heard everybody say they like the Teutsch middle column. And I've heard, I think -- but I want to make sure I have this right, it's why I'm summarizing where we are -- so far two additions to that middle column. One is, the one that was just mentioned or mentioned a while ago. And that is when we're thinking about prognosis, we should be emphasizing -- I think I'm interpreting our remarks correctly -- the incremental value. And if that's the case, we definitely need reclassification matrices.

The comment was also made that we probably need that for diagnostic tests, when we're going from zero to something, but for sure in a prognostic test. The comment was made we need to know the marginal increment, and we can do that only if we know how many patients moved up or down the diagnostics -- the prognostic spectrum.

The other new point or emphasized point in

this middle column was, I think, the fact that Catherine just raised, which was that for common diseases, when we're looking for prognostic factors or prognostic tests, we probably are going to want more studies than for rarer diseases where we might be satisfied with fewer studies. Was that your point, Catherine?

DR. ENG: (Nodding head.)

DR. MC NEIL: Now, I don't know if I've captured everything that we've said that differs from this middle column for prognosis. But I want to say that that's where I think we are. And if you think we're someplace different, say so. Yes, Deborah?

DR. SHATIN: I would just like to make a comment. I agree with what Catherine said, and it raises other questions, too, which is, are we saying, is this the floor or the ceiling? For a new genetic test, are we saying it has to meet every single item in the clinical validity column? Or are we saying that where it's reasonable, it would meet those particular items? And I think that's very important

to the discussion today.

DR. ENG: I was thinking more about the reasonable, not meeting every single point.

DR. MC NEIL: I mean Steve has actually put a little hierarchy within each one of these on the first part of the column, going from good to not so good. So I assume we would have the same.

DR. SHATIN: I'd like that documented.

DR. MC NEIL: Yes. I wish I had read it more thoroughly before we came. I read it, but I didn't memorize it. Should we move on? And we can come back to pharmacogenomics with regard to - yes, Marion?

DR. DANIS: Can I just ask one thing? I'm not sure I see any prompt in the subsequent questions about dealing with the clinical utility piece. And I was just wondering when we think about prognostic assessment, and I'm prompted to think about this because of the Alzheimer's issue that was just raised.

DR. MC NEIL: Isn't that question four?

DR. DANIS: Okay. Sorry.

DR. MC NEIL: Okay. Pharmacogenomic assessment, who wants to take a crack at that, in terms of analytic validity? That may be a little bit tougher. But we're strong. I think. All right, Cliff.

DR. GOODMAN: You're introducing the decision to use a drug or not, a therapy. So if you're going to start introducing another technology, you've got to be concerned about the impact on the benefits and risks inherent in that. It's a more involved question.

DR. MC NEIL: Moving up the spectrum. So that would mean -- what does that mean in terms of analytic validity?

DR. GRANT: I would say, you know, in general the same principles apply. It's how accurate is the test to be able to discern whether patients accrue benefit or not or harm from therapy?

And complicating it a little bit more, you can always say that tests can be predictive and prognostic at the same time. So by that logic, I

would say you have to assess them by the same means. I might be wrong.

I would say that in general the same -- as I said, the same principles apply.

DR. SCHEUNER: So the clinical validity issue is how well does it predict the response? That's the phenotype. Is that that response variable? And then in terms of outcomes related to that, health outcomes, what have you, that would be the utility.

DR. MC NEIL: Uh-huh. Can I ask Maren a question or maybe Marion? Are case controlled studies good for pharmacogenomic studies?

DR. GOODMAN: What question are you asking about then? If you're asking whether or not treatment based pharmacogenomic treatment improves effectiveness?

DR. MC NEIL: I'm asking whether you can trust results from case control studies as a guide to whether a genetic test is good for guiding therapy from a pharmacogenomic perspective.

DR. GOODMAN: But are you also asking

whether the potentially indicated drug works or not? Or are you stopping just to say how good is the test in telling you what drug to use?

DR. MC NEIL: I guess I should be asking the latter. Just how --

DR. GOODMAN: Well, I'm not going to answer your question, But I'll muddy it by suggesting I think when you link a drug to a diagnostic or, I guess, when you link anything to a diagnostic, you certain immensely -- I would agree, you raise the stakes because the drug becomes the slave to the diagnostic, in that if you've chosen the wrong patient, you've actually had an impact on the efficacy of the drug. You've complicated the design.

And in some ways you've telescoped the design because the clinical sensitivity and specificity are measured in outcomes of drug response. So in some ways, when you have pharmacogenomics, you merge clinical validity with clinical utility in an odd way.

So again within the bio market community

that I know and love, the big fight in the pharmacogenomics arena relates to whether you need to look at the entire population, whether you need to do all-comer studies or whether you can, in fact, use feasibility data and mechanistic data and background data and study only bio marker relevant patients.

The advantage to the second is that it's very directed and very facile and very fast. The disadvantage is at the end of the study, the only thing you know about the test itself is the predicted value of a positive. You don't actually know the sensitivity of the test. You don't know the specificity of the test. You don't know the predicted value of a negative. And about the drug, you know that the drug worked in bio marker positive patients or bio marker relevant patients.

You don't know what the performance of the drug is like in those excluded. So it's a much more complicated mix.

DR. MC NEIL: Okay. So I still don't know the answer to my question. Does that mean -- maybe

you could -- I'm still trying to decide whether the kinds of clinical studies that you would do to look at a genetic test in terms of pharmacogenomics, a pharmacogenomic genetic test, where obviously we're now closely linking the test the choice of a drug and the outcomes of a drug.

Does the type of clinical trial we do, is it the same as for everything else? Do we accept all kinds of clinical trials for that? And Steve, you just gave the article.

DR. GUTMAN: Yeah. I know.

DR. MC NEIL: And if you go in the bio marker positive only, then you're stuck with predicted value positive.

DR. GUTMAN: Yeah. I know.

DR. MC NEIL: But you're very efficient at getting that number.

DR. GUTMAN: And FDA has cleared and approved it. And I believe many third parties have paid for at least some bio markers that have been studied in that manner.

DR. MC NEIL: Okay.

DR. GOODMAN: FDA --

DR. GUTMAN: That's not to suggest -- I think it's the standard you should aim at. It's just an observation.

DR. GOODMAN: FDA still has to make a decision about whether to approve the therapeutic side of this pharmacogenomic intervention. And in virtually all cases, I would hope that it would be done with an RCT. There are some few cases where that might not be appropriate for a person.

DR. MC NEIL: That's true.

DR. GOODMAN: Sample size too small, other ethical problems. So that evidence needs to be developed with regard to the level of evidence required for the test that might trigger the use of this drug.

DR. MC NEIL: Right. That's what I'm asking.

DR. GOODMAN: I would suggest that you have to crawl up the evidence hierarchy some because you

are triggering a decision that introduces benefits and harms.

DR. MC NEIL: Okay. That's what I needed --

DR. GOODMAN: And one more factor, Barbara. And that is sometimes it's really important to be exactly right about the decision to use the drug or not. And sometimes you don't have to be so precise. And so you could maybe go further up or down the scale depending upon how important it is to be precise in your finding. Sometimes it isn't so important. And in those cases, you might be able to down the hierarchy a notch.

DR. MC NEIL: So what you're saying is you may need to go up or down depending upon the benefits and risks of the drug.

DR. GOODMAN: Yes.

DR. MC NEIL: That's a good -- good way of doing it. Excellent. Yes, Theresa -- I'm sorry. Eleanor?

DR. PERFETTO: But I think one of the things that is important to remember is that in terms of the

FDA's assessment of this, the test and the drug are going to be evaluated separately in terms of whether or not the test works to do what it's supposed to do and whether or not the drug works to do what it's supposed to do. But there will be consideration of the use of the test when the drug is evaluated. The decision's not made in one decision. It's made in two separate decisions.

DR. GUTMAN: Well, actually, it's contingent. There have been some circumstances where actually there have been -- herceptin is an example where the biologic and the tests were -- they were in separate decisions, but they were administratively linked.

DR. MC NEIL: Are we ready to go on to question four, with the idea that we can go back? I mean I get the sense that we are refining our thinking. All right.

So this is really tough. So question four says that if we think there are different criteria, depending upon whether it's diagnosis, prognosis, or

pharmacogenomic assessment, we have to answer four a, b, and c separately, which would mean nine responses, right, versus three. Now Maria is prepared --

DR. PHURROUGH: Don't let that enter into your thinking, to do the correct thing --

DR. MC NEIL: Of course.

DR. PHURROUGH: -- not the efficient thing.

DR. MC NEIL: No, no, no. We're always correct, Steve. Maria, by the way, is equipped to handle either option with her spreadsheet. Is that correct?

MS. ELLIS: Yes.

DR. MC NEIL: So I'm trying to figure out, how do we feel about this? I mean in general, we've said that we like table four, but we've tinkered a little bit with it. We've tinkered with it on prognosis, primarily. And for pharmacogenomics, we raised the issue that Cliff raised about moving up and down the hierarchy, depending upon the cost and benefits of the subsequent therapy.

DR. GOODMAN: I didn't say costs.

DR. MC NEIL: I didn't mean costs. I mean benefits and risks. I'm sorry. Correct. I understood what you meant. I said it wrong. So how do you want to do it, gang?

DR. PEARSON: I think we should probably just - again, table four is really about individual studies. And it really doesn't drive to the issue of a body of evidence and what kinds of levels of evidence we would assume are desirable. And in my mind, we haven't really talked that much about clinical utility yet. So I would suggest we need to spend some time doing that.

DR. MC NEIL: And where is that? You don't think that's question four?

DR. PEARSON: Well, no. I said I think we need to discuss that now.

DR. MC NEIL: Yeah. That's what we're doing.

DR. PEARSON: But it's not table four.

DR. MC NEIL: Oh, no, no, no. It's not. We're going to question four. But the question is, do

we answer question four separately for each of the three paradigms, or do we consider them the same?

DR. PEARSON: I would suggest we just launch into it and see if we end up with three different pockets --

DR. MC NEIL: Okay.

DR. PEARSON: -- or not. My guess is that we may not feel that it needs to be split up so much.

DR. MC NEIL: How do people want to do it? Do they want to go that route? I see no movement, up or down or sideways. Mark, say something.

DR. GRANT: I want to do it once.

DR. MC NEIL: What?

DR. GRANT: I want to do it once.

DR. MC NEIL: You want to do it once?

DR. GRANT: Yeah.

DR. MC NEIL: Well, let's try. Okay. So let's just pretend we're answering this question right now, Maria. We're not doing it yet. We're just pretending.

So for each of the outcomes, how confident

are you that methodologically -- I can't say that either -- rigorous evidence on the outcome is sufficient to infer whether the test improves patient centered outcomes? And then there were three of them.

So physician-directed patient management. Steve, do you have any advice about how we approach this question?

DR. PHURROUGH: I think you can just have a general discussion around whether you think that good evidence changes in physician-directed patient management, in fact, is a good indicator of improved clinical utility, improved patient outcomes. And if you think so, then discuss whether it differs by diagnostic, prognostic, or pharmacogenetic.

DR. MC NEIL: Okay. Got it? We have a lot of discussion. As much as you want. Let's see. Jim was first.

DR. PUKLIN: I would just say that's one of the great problems with American medicine is that evidence based findings, based upon randomized clinical trials that are clearly beneficial to a whole

segment of the population, still can't be introduced properly to the American public.

I think metabolic control of diabetes is a perfect example. The majority of the people in this country still, ten years after the DCCT, are not being properly managed.

So how can you expect that anything is going to be influential in a private practice or in the practice of medicine. The fallout is incredible after you get outside of academic centers in all areas.

DR. MC NEIL: Okay, Mark?

DR. GRANT: I just wanted to make a sort of broad comment about this whole picture of evidence and as it relates to clinical utility. I think that, you know, ultimately what decision makers need is evidence presented in some way so that they know whether they're going to be right or wrong with some relative degrees. Right?

Now, it would be wonderful if you had randomized control trials or direct evidence to support every decision. That doesn't happen, for the

obvious reasons. Although sometimes it has to happen because the risks and the benefits are so tightly wound and balanced that you need to know.

On the other hand, we're dealing here with a body of evidence that -- what I mentioned before -- is indirect. So the EGAPP has a model. But ultimately, it can be explicit or implicit. And I think we do it informally. And implicitly we have our accuracy data. We say, well how is that going to inform whatever these -- in this case we're talking some intermediate outcomes and the outcomes or the ultimate clinical outcomes. And how you piece it all together really is the crux of this and sort of where do you stop along the way.

But the other parts that I think we're not grappling with -- and this isn't the place to grapple with it -- but I'd just like to lay it out there, is that although it's nice to lay out analytic validity, clinical validity, and clinical utility. You look at them and say, well can we put them together in our head. And that's what I asked Tom before.

That's very difficult to do, except in the clearest of circumstances, such as KRAS for metastatic colorectal cancer is very hard for oncotype. It was hard. And I think that a lot of the process here demands a greater degree of explicitness, not just in linking those pieces of the puzzle, in terms of what the outcome is, but also the degree of uncertainty. Because a lot of this has to do with uncertainty, and we pay not enough attention.

Every time you stop patient centered management, you have more uncertainty, surrogate measures, more uncertainty. And the more you can tie it in into that hard outcomes, the better off you are.

But I think that it might be worthwhile to consider taking a more explicit approach. It allows, I think, for more transparency and, in some cases, more defensibility about the process.

DR. MC NEIL: So say what you mean by more explicit. Express the way --

DR. GRANT: Well, if you look at the EGAPP and the other evidence review of oncotype or the gene

expression profiling, I mean, for example, they reference in cost effectiveness studies, but they're modeling. Whatever.

But one of the interesting findings of an older study was well, in fact, you do improve quality adjusted life years, but at the expense of life expectancy. That's important to know for a decision maker. And unless you put it together in that fashion, you'll never be able to cull out those pieces of the puzzle.

You know, what are the harms of unnecessary chemotherapy versus a potential benefit? And how do people weigh that? Now, we don't tend to be explicit about that and say qualities, or we can do it in life years. However you want to do it.

But these pieces can be linked together. And I think we ought to think about that because it's very difficult to do in our heads. They're just criteria.

DR. MC NEIL: Right. Steve, you had a comment.

DR. GUTMAN: Well, I agree with just what's been said. I think the theme that was coming in this morning about flexibility differences resonates with me. So I think that this is both easy and hard. If I were rating this, I'd say a is one. I have very little confidence in it. C is a five. I have great confidence in it. That's the good news. So those two are easy.

And the problem is b. And unfortunately b, depending on the circumstances, at least in my book, it ranges from minus one to six.

DR. MC NEIL: Those aren't allowed values.

DR. GUTMAN: I work at the University of Central Florida, so they allow that.

DR. MC NEIL: Oh, okay.

DR. GUTMAN: So the problem is that you might not need c in every case or any of the three markers for diagnostic or prognostic or pharmacogenomic because there might be power of evidence that gets you the outcome without going all the way to the clinical trial.

I would argue that probably there will never be a circumstance when just changing physician directed-patient management by itself would get you very far. And I don't know this afternoon with what's on the table, that we can actually answer for CMS all the possible choices for b.

DR. MC NEIL: Neil? Neil, then Steve.

DR. HOLTZMAN: Part of the problem for me is this distinction between diagnostic, prognostic, and pharmacogenomic. Because in each case, and I said this before -- and remember we're talking about diagnosis. We're talking about confirming or ruling out a specific diagnosis in somebody with symptoms and signs.

So within that category of diagnostic tests, there are others. There are some that also have prognostic significance. But there are some diagnostic tests that may not have prognosis diagnostic implications.

Within prognostic tests, there are some that will influence therapy. And some that will not. And

I would offer -- maybe this is too late to do this -- that that sort of overlapping classification is a better one than simply this diagnostic, prognostic, pharmacogenomic, because they're overlapping. Sorry.

DR. MC NEIL: No. No. That wasn't a conversation stopper. Catherine. Oh, I'm sorry, Steve, and then Catherine.

DR. PEARSON: I was just going to say what I like about this question is sometimes what's important is what's on the page, and sometimes it's what's not on the page. And what's not on the page here is an option that information is important by itself, that somehow we're starting with the idea that we need to see some change in physician-directed patient management.

So just kind of perhaps to restate the obvious. But, you know, greater precision and just information is not on this list is what we would consider to be sufficient to infer whether it improves patient centered health outcomes. And I actually think that's a very important statement for the MEDCAC

to clarify.

DR. MC NEIL: So say that again in different words.

DR. PEARSON: Well, again there have been lots of debates about what is the value of genetic tests. And there's an argument that you will hear that just knowing will improve patients' health outcomes.

DR. MC NEIL: Okay.

DR. PEARSON: There's a kind of something that will kind of lead to changes in behavior or other things by patients. You have to do a lot of assuming.

But we're really kind of starting with the assumption that we're looking for some tangible evidence that really changes the pathway of care.

DR. MC NEIL: That's what these are.

DR. PEARSON: And we're trying to figure out how far down that pathway to go in different situations.

So greater precision -- one of my favorite examples isn't even from genetics, obviously. But if

we had an oxygen saturation meter that could tell us not that it's 92 percent 02 Sap, but 92.5, that's greater precision, but it's not going to -- it gives us more information, but it's not going to change patient management.

So reclassification in my mind kind of even starts -- again we've already talked about it a lot. So it comes up as critically important in my mind for clinical validity to even establish whether there are possibilities for changing physician management, et cetera.

And I do think that in general, when it's a diagnostic situation, if we can have a very clear -- I mean if it's kind of established that there is a set pathway of care that's established for this diagnosis or this subgroup, then I don't think we necessarily have to see new evidence of patient change in behavior or management.

If we know that if this patient falls into this risk category, they are definitely always managed this way. And if we have a test, and we have good

evidence that the reclassification will shift more patients into that category, that, in my mind, would be pretty strong evidence and not require any further studies, if you will.

On the other hand, warfarin, to me, is an interesting example because there will be reclassification, but it could go into two directions. Right now, we kind of treat everybody in the middle with an algorithm. We may now create three pathways, still people in the middle, and some people will receive higher doses up front. And some will receive lower doses.

But they're the misclassification problem, the false positive and false negative issues. It may mean that we really do need randomized control trials to see how the change in patient management really plays out into patient outcomes.

So I don't think it's always going to be very straightforward to figure out what level of evidence you need. But I do think that this is the right list, and that starting with reclassification is

the absolute floor for clinical validity. Then moving into these three others, I think is the right way for us to go.

DR. MC NEIL: Catherine, and then Jim

DR. ENG: I'm a little bit more optimistic in terms of the changes in physician-directed patient management, especially physicians managing geriatric patients, who are very complex. And I think that all three, diagnostic, prognostic, and therapeutic, anything that would help the physicians manage and that would reduce the burden of further testing, I think that's something that I would add to reason.

I think when we are managing frail patients who have multiple co-morbidities, the last thing we want to do is continue to test in, you know, individual conditions. And if we had one test that would help us, the warfarin story is getting there.

However, I'm not so optimistic that I think it's a four or five. I think it's a three at this point because I don't think that the evidence is there. And if the evidence is there, it's not

presented or communicated in a convincing enough way to busy clinicians.

And I think that in addition to covering a test, I think that the societies, the different societies, have a role in helping to educate and having physicians believe this.

DR. MC NEIL: Let's see, Jim, are you going to comment?

DR. PUKLIN: Yeah. I just wanted to beat a dead horse a little bit longer.

Regardless of the genetic testing sequences, whether it be with warfarin or whether it be with various types of cancers, gene sequencing allows certain cancers to be better treated by certain drugs because they're more responsive and so forth.

I think the changes in physician-directed patient management is really going to be directed by where the tests are used and where the patients are cared for. So, for example, the trend in major metropolitan areas is for most people who have serious cancers to be treated at a cancer center funded by the

NIH. There are some of these around the country. And a lot of the intensive care is given there. And that's state of the art.

And if all the patients are sent to places like that, they'll get first rate care with implementation of all the latest diagnostic testing, including genetic testing. And it will reduce the mortality, the morbidity, and the adverse events, which is c in this question.

But if it's warfarin testing, to bring the INR into the proper level, I can tell you from working in central metropolitan areas and in the communities, that patients away from academic medical centers just aren't being monitored correctly, presently on coumadin and warfarin. And they're not going to get any better because the private practitioners, the general practitioners out there, probably aren't going to deploy it.

So the issue about physician-directed patient management really depends upon where the test is going to be utilized the most. And I think it

applies to whatever test. It's a highly variable situation.

DR. MC NEIL: Cliff, and then Marion.

DR. GOODMAN: I want to agree and strengthen the point. A, b, and c are basically - this question is asking you, how far to the right on the analytic framework do you need to go in order to be satisfied that evidence linking the test to a or to b or to c suffice to tell you that you're going to improve patient centered outcomes.

This is clearly going to reflect the particular circumstances of the situation. So I wouldn't be confident making any general statement about whether I'd always be satisfied with a or b or c. It's going to depend very, very much on the patient population, on the late indications, on the medicines that are going to be taken by the patient.

You could do very, very well linking something to changes in physician-directed patient management. But if you know the physician is going to say take drug X, and patients can't stand taking drug

X because of the side effects or the pill is too big or they can't otherwise comply, then clearly that's not good enough evidence.

You might have very good evidence linking the test to b. But there's no really good evidence showing that the intermediate health care outcome has anything to do with the patient centered outcome. That's going to be situation specific.

DR. MC NEIL: Marion, did you want to?

DR. DANIS: I have just a repeat of this, but just to say it's the setting. In the clinical setting, it's also how complex the patient's illness is and if patients are frail, elderly, having falls, not remembering to take their medicines, nutritionally not in good shape. Whether the INR is in good control is not just a function of what their coumadin dose is. So I think the likelihood it's going to have an effect is really going to be small.

DR. MC NEIL: So I'm trying to think of what we've just been saying. Does that mean it's not possible to answer these questions in the absence of a

specific example in a specific setting? Is that what we're just saying?

DR. GRANT: Sort of. But can I just say -- I want to add. I don't disagree. But I think it's important to lay out some principles of evidence that yes, everything is specific. But I also think -- I feel very strongly, that a -- I tend to agree with the one. There's uncertainty there. And there are probably cases where it suffices, but they are probably not very common. All right?

The same thing with the surrogate marker, yes. There are instances. But to say that in fact this is something we'll accept as evidence, I would say no. That's not a good standard to consider when you're designing a study, unless you have a compelling and convincing argument to demonstrate in fact that this proposed surrogate marker in fact correlates quite well with your outcomes.

Whereas, what we desire are outcomes, clinically relevant patient outcomes. And I think that it's important to be - I feel it's important to

be on record stating that, that these in fact do differ. They differ considerably. There may be exceptions. There always are.

DR. MC NEIL: I think I lost track here. Did somebody else want to -

DR. PERFETTO: Yeah. I wanted --

DR. MC NEIL: Oh, Eleanor. Sorry.

DR. PERFETTO: I was actually just going to make the comment that there was a question that was asked earlier that is relative to a. And it was, should we be considering this in the context of whether or not you actually see changes in practice or whether you should see changes in practice.

And I think that's something that we have to be very pragmatic about because we've seen a long history of things that should have changed practice. And it took many years for it to actually happen. So I think we just have to be practical about this.

DR. MC NEIL: Linda, were you going to say something?

DR. HOLTZMAN: Might I?

DR. MC NEIL: No? I thought Linda was going to say something. Okay. Neil, you're on.

DR. HOLTZMAN: Well, I agree that it's very difficult to make general statements in answer to this question without considering what the test is and what the disease is. I also think that this a,b,c classification is overlapping.

It's hard to believe for me that one would bode strongly that test meets changes in physician-directed patient management or allows for physician directed-patient management that doesn't at the same time have an effect on c, patient outcome. And so I'm a little confused here as I was about the inputs diagnostic prognostic pharmacogenomic. I'm also confused about the independence of each of these three outcome factors.

Now, just historically a little bit, I mentioned when I introduced myself that I chaired the first real task force that looked at the safety and effectiveness issues of genetic tests. And one of the things that we suggested there, in trying to decide

how much evidence was necessary, was to develop a hierarchy of the diseases or the situations and when the tests would be used. And that some would require stringent scrutiny.

And, for instance, one of the things that we talked about here would be common diseases, for a number of reasons. One is the effect of a relatively large part of the population. And secondly, from a genetic point of view, they're much more complex. When we don't have a single gene that is going to, with very rare exceptions, account for prognostic or therapeutic differences.

So it seems to me that if MEDCAC and CMS are really concerned about validating or deciding on decisions about reimbursement for genetic tests, then one has to take the different view of classification. And I would argue many of the things that are underlying what I'm saying, analytic, clinical validity and utility are extremely important and the bedrock of making any of these assessments.

But I think that one cannot look for

generalizations.

DR. MC NEIL: Okay.

DR. HOLTZMAN: Genetics itself is the study of variation. And I think that one has to set up some sort of hierarchy of the stringency of what the evaluation requires.

DR. MC NEIL: Mina, you had a comment?

DR. CHUNG: You know, what is desirable may not be what's necessary or what's actually achievable. And while it is good to - I agree with your comments, but I am stuck on the clinical validity versus clinical utility issue.

You know, whether we saw evidence on those studies that looked at warfarin, APOE, and MTHFR, there is only one study in the warfarin studies that applied to clinical utility. Zero in the others.

And we may wish to have all -- I just fear that although, yes, it would be great to have these studies, evidence that support changes in physician-directed management and health outcomes in a surrogate or in a harder endpoint, that that data just does not

exist.

DR. MC NEIL: Can I try a statement? I've heard a couple of things. One is that generalization is not possible, whether it's across diagnostic tests or across these three categories. At least I think I heard that. Or did I not hear that?

UNKNOWN FEMALE VOICE: I think you heard it.

DR. MC NEIL: Okay. Did I also hear that it's very difficult for us to answer these three questions? But if we were to make an answer, if we were to answer something, we would say that the most important end point was number c, and that we would want the most rigorous evidence for number c. And that we would accept less rigorous for number b, realizing that there is an interaction between b and c. And that when we think about question 4A, we get into the problem of should versus did.

And that's a little bit hard to answer that because then that gets to be very physician specific. All of them get to be physician and setting specific.

I don't want us to perseverate forever on

our going around the fact that we can't generalize, and we certainly -- well, I don't feel comfortable saying one, two, three, four, five for each one of these. But I feel like we're stuck. So I need a suggestion. Linda?

DR. BERGTHOLD: Well, I don't know if this will help. But I've been just struggling with this "is sufficient" the whole time we've been talking about it. And I finally now think that we're talking about "would be sufficient" not "is currently sufficient."

DR. MC NEIL: Oh, okay.

DR. BERGTHOLD: Because the question is not is the evidence -- is there rigorous evidence to infer now about these things. Isn't what you're asking us --

DR. MC NEIL: Would be sufficient.

DR. BERGTHOLD: -- "would" -- I don't know. I mean, clarify it for me.

DR. PHURROUGH: I think the interest -- this is fascinating each time we have these MEDCACs, and we

spend lots of time trying to get these questions just right. And Steve and Barbara spend time with us trying to get these questions right, and we know what we want. And then we have this conversation here, and it's, what was it that we wanted?

If you make the assumption that there is good evidence, however you define that, that there's good evidence that a genetic test, a diagnostic genetic test, changes physician-directed patient management, is that sufficient for coverage? Is that the kind of evidence that we ought to use to make a coverage decision? Assuming that there is good evidence.

DR. BERGTHOLD: What else would there be, bad evidence?

DR. PHURROUGH: Sure.

UNKNOWN MALE VOICE: A lot of bad evidence.

DR. PHURROUGH: Most of the evidence is bad.

DR. BERGTHOLD: Of course, if there's methodologically rigorous evidence --

DR. PHURROUGH: Even if it's

methodologically good evidence, but the outcome is only physician-directed patient management changes --

DR. BERGTHOLD: Oh.

DR. PHURROUGH: If that's the only evidence you have is physician-directed patient management, should you use only physician-directed patient management in making your coverage decision?

DR. BERGTHOLD: I didn't see that.

DR. PUKLIN: That would be sufficient if you want to get to the patient. You can't get to the patients without the physicians. So if that's a categorical paradigm, then you have the criteria for a, and that automatically gets you to c.

DR. MC NEIL: No. It doesn't.

DR. PHURROUGH: No, no, no.

DR. MC NEIL: No. It doesn't.

DR. PHURROUGH: You may never know. We have a good example right now. We are just about to finish a coverage decision on petscan, nothing to do with genetic test. Petscanning in cancer patients.

And we've been looking at PET scanning on

cancer patients for my lifetime. And the level of evidence has been non-existent, except in rare instances. So our last decision, we said, all right, we want you to collect more evidence. And we allowed that to be done in a registry of physicians reporting how the results of the petscan when added to the other diagnostic tests available changed their management of the patient.

So we have this data. It's all the data we have. We don't have anything that says, well patients were better, patients live longer, I gave him less chemotherapy, I gave him more chemotherapy, I operated on him more often, I operated on him less often, they were hospitalized more, they were hospitalized less, they lived longer, they had longer disease-free survival. We don't have any of that data. But we do know that physicians change their mind about what they would do.

DR. BERGTHOLD: Or said they changed their minds.

DR. PHURROUGH: Well, let's make the

assumption they were honest. And physicians always are. So that's not an uncommon presentation to us. That's the only evidence we have. Physicians are going to do something different because of the test that's been provided. Is that sufficient for us to decide, or should we use only that level of evidence to make decisions around whether we should pay something or not?

DR. MC NEIL: All right. If that's the question, let me just do a straw poll here because this is shedding a little bit more light on where I think we want to go.

So you're saying, Steve, if the only evidence you had was good evidence on a test, a genetic test, and it changed physician-directed management -- and we'll smush together diagnosis, prognosis, and pharmacogenomic therapy for the moment -- would that be good enough for this panel to think that CMS should cover the test?

So how many people think on a one to five scale, one being terrible, five being good, how many

people think that one is the appropriate answer to this? That is, this is not sufficient.

Oh, you've got cards. I don't have cards.

UNKNOWN MALE VOICE: Do you want us to hold the cards?

DR. MC NEIL: Yeah. Why not? Okay, good.

(Whereupon, the panel voted.)

DR. MC NEIL: I can't vote, but I see one, one, one, one. I dare anybody to put up a two. Oh, five.

MS. ELLIS: On the left hand side of your folder in the back, there is a score sheet for you to circle your answer and put your name at the bottom, and I will collect them.

DR. SCHEUNER: Barbara, I have a question.

DR. MC NEIL: Maren, yes?

DR. SCHEUNER: So when we were on the phone on February 6th or whenever that was, and I asked the question, if the test is going to end a diagnostic dilemma, in other words I'm not quite sure what this patient has. They have a bunch of complaints, issues,

and I do this genetic test and it gives me an answer, a true diagnostic test. Is that just --

DR. MC NEIL: Well, that's going to change the management. You're not going to do more tests. Right?

DR. SCHEUNER: Right. That's what I was going to say. In terms of c, indirectly you might avoid more procedures, et cetera. So that would be --

DR. MC NEIL: Right.

DR. SCHEUNER: Okay. I just wanted to make sure where that would fit.

DR. MC NEIL: No?

DR. PEARSON: Maybe we need a little bit more conversation. That's a very good question. But my assumption is that they really are asking us whether we would use that level of evidence without having further evidence of the linkage to patient outcomes. That there's an assumption.

Now, I was trying to make the point that sometimes you have a really clear, hard assumption that seems blazingly obvious. But sometimes it's not

so clear that a diagnosis will lead to changes in management. That's the hard part.

DR. SCHEUNER: I'm now thinking in the realm of the rare Mendelian disorders, where you are trying to understand what is going on with this patient.

DR. MC NEIL: Maybe for the sake of this question, we should - I don't mean to be too blase here, but maybe for the very rare situation, we should have a separate discussion.

DR. SCHEUNER: Well, that would be helpful to make that distinction.

DR. MC NEIL: Would that be helpful?

DR. SCHEUNER: I was trying to prod us along that whole --

DR. MC NEIL: Right. Why don't we try and do that because I think it is a very specific situation. And we can ask about that separately.

DR. PHURROUGH: And as the discussion has gone all afternoon, there are always exceptions. And so with whatever recommendations you're going to give to us, we will always take those recommendations and

apply them in a specific circumstance. And that circumstance may change how we view that particular -- the results of a particular body of evidence based upon those circumstances.

So I think we are asking a general question. In general, for the majority of the genetic tests, what would you find?

DR. SCHEUNER: But let me ask you, what do you think the majority of genetic testing is comprised of? I mean what do you think is happening out there? Do you think it's - I mean, actually there's a lot of genetic testing for rare Mendelian disorders that each one of them are rare, but collectively there's a lot. So --

DR. MC NEIL: Let's divide the question in two because I think we're going to not have an answer to your question, Maren. So let's answer the question, removing the exceptionally. I mean what you are saying is cumulatively a lot of rare tests add up to a lot of tests.

And I think what we have been talking about

is tests that are more common, a few tests that are more common and may add up to the same number. But it's a few tests on a smaller group of patients versus lots of rare diseases.

So let's just talk about the non-rare situation, and then we'll come back. Because I think we're going to confuse this group, this panel too much. Deborah?

DR. SHATIN: I have a question. This won't be used for coverage decisions?

DR. MC NEIL: No. We're not talking about coverage. These are criteria.

DR. SHATIN: Right. But what will it be used for then?

DR. MC NEIL: Let's answer the questions first, and then have Steve answer it later, unless you want to answer it now.

DR. PHURROUGH: We're asking you for recommendations on what kind of evidence we should use in making coverage decisions. That's your recommendation. And so should we use evidence that

only shows changes in physician management as sufficient for making a coverage decision?

DR. MC NEIL: All right. Let's answer that. So that's the question.

UNKNOWN FEMALE VOICE: That's different from if it's the only one. I mean --

DR. PEARSON: Think of them as steps on a ladder. It's a hierarchy. And we're talking about as we go up the hierarchy towards what's considered to be higher quality data usually, according to the Frybeck schema, et cetera. Does your confidence in whether it will be sufficient -- if that's as high up as they get and no further, how sufficient would that be to make a decision?

DR. MC NEIL: Steve said it exactly right, Catherine. If that's all you've got, how do you vote?

DR. DANIS: Let's ask Steve about that. So Steve, you're saying that you might imagine that physicians are making management decisions where they don't expect changes in outcomes?

DR. PEARSON: No, not that they don't expect

or that we wouldn't expect, necessarily. But that we do not have published literature to support that.

DR. DANIS: Okay.

DR. MC NEIL: That's what I think Steve was talking about with the pet study. That would be an example of that. There were data on changes in physician-directed management, but that's as far as the data chain went for those particular data sets. Now, they may go farther in other data sets. But for that particular data set, that was all it was.

So are we clear on this? So if the only information you have is changes in physician-directed patient management, voting from one to five; one you don't think much of it; five you think it's an important outcome that should be considered highly when CMS then goes to evaluate genetic tests for coverage for the common genetic diseases. And we'll come back to Mendelian later.

DR. PHURROUGH: Hold your numbers up, and keep them up until Maria has a chance to look at them.

(Whereupon, the panel voted.)

DR. MC NEIL: Good? Okay. Now, we've moved up in Steve's word, one step on the ladder. And we have a change in an intermediate outcome, like lab tests, presumably a warfarin dose -- I mean a coumadin dose.

DR. ENG: On top of a?

DR. MC NEIL: Yeah. We're voting.

MS. ELLIS: Good.

DR. MC NEIL: Okay?

DR. PHURROUGH: Please mark your papers also, we have to officially record your thinking.

DR. MC NEIL: And how about direct patient centered outcomes? Mortality, morbidity, functional status, adverse events, blah, blah, blah.

(Whereupon, the panel voted.)

DR. MC NEIL: Now, I'm hoping as a check for the understanding of these questions, there's nobody whose numbers don't go sequentially up. That would not be good. Nobody's going one, three, one, or something like that. That would not be good. Right?

So now let's go to Maren's question. And

I'm not quite sure what you do with this, Maria. You're interested in knowing for a rare disease in which the genetic test stops the work-up --

DR. SCHEUNER: It gives you an answer.

DR. MC NEIL: I'm sorry?

DR. SCHEUNER: Yeah. It gives the answer.

DR. MC NEIL: It gives the answer. Stops the work-up.

DR. SCHEUNER: You've got a patient with all kinds of complaints, and you actually figure out why.

DR. PHURROUGH: Isn't that an outcome?

DR. GRANT: That's an outcome. I would say the answer really has value in and of the way it translates into help outcomes. In this case, avoiding, if you can say you avoided further work-up, and that work-up had negative consequences, that's an outcome. You avoid negative consequences. That's how I would see it.

DR. MC NEIL: So you'd have a footnote for question 4C. And say this, in particular, is the only relevant or may be the most relevant -- well, how

would you word that? I mean I think the conversation is right.

DR. PHURROUGH: I think that's under C at the top of page two, that's another for example.

DR. MC NEIL: Okay. Oh, good idea. Excellent idea. And Neil, did you have a comment on this?

DR. HOLTZMAN: Well, on this matter of rare genetic diseases. I mean we are being restricted to the Medicare population.

DR. MC NEIL: Well, not yet. We'll get to the Medicare later.

DR. HOLTZMAN: Because the number where that would be applicable to a Medicare population over 65 is negligible.

DR. MC NEIL: Right. Okay, are we done with question four? Maren, did you want to contest that?

DR. PEARSON: Can I just make one quick comment?

DR. MC NEIL: Sure.

DR. PEARSON: And actually it's just going

to echo what Mark said earlier, which I like to think echoed what I said earlier. But it was you had framed this as we can't really generalize. And I would say I feel like what we were saying is that although it's difficult to generalize, there are clear principles of better evidence and thresholds of evidence that we think should be met in order to have sufficient evidence. I hope I'm not overstating that for anybody on the panel. Let me know if I am.

But it's not to leave the feeling that we feel that it's just a soup. And I think our votes kind of show it that way. But that there are principles that you can use, in a sense, to say that there is good evidence that should be required for the use of these tests.

DR. MC NEIL: Okay.

DR. PEARSON: Mark, am I accurately --

DR. GOODMAN: Can I add to that?

DR. MC NEIL: Yeah. Absolutely.

DR. GOODMAN: What we've done is we've shown a relative preference, but not an absolute preference.

In any given circumstance, we always prefer to have c over b, b over a. But there's certainly going to be some circumstances if a given genetic test comes here for a national coverage determination, it may very well be that rigorous evidence for b would more than suffice because b is very strongly and definitively linked to a patient centered outcome. And in that case, b would be fine.

DR. GRANT: No hematocrits for ESAs.

DR. MC NEIL: I'm sorry?

DR. GRANT: No hematocrits for ESAs.

DR. MC NEIL: No hematocrits for what?

DR. GRANT: No hematocrit targets for ESAs.

DR. MC NEIL: Oh, okay. Okay. Got it. Okay. So, yes, Eleanor?

DR. PERFETTO: Just to emphasize Cliff's point, I think it goes back when you were summarizing earlier. And you were summarizing what the points were.

The first thing you said was that each of these things were going to have to be looked at

individually because of all of this variability. So it's kind of our vote laid on that statement, I think covers what we're trying to convey.

DR. MC NEIL: Okay. Good. All right. Now, let's get to ethics and have a little bit of a discussion before we vote on this. Actually, we're not going to vote. We're just going to discuss. So here we are.

DR. PUKLIN: Do you want to limit this to non-Mendelian diseases, or is it an open forum?

DR. MC NEIL: It's open. Do you think it shouldn't be? No. It's open.

DR. BERGTHOLD: Can you give us an example, Steve, of what you mean by an ethical issue that would alter the rigor? So we don't waste time on this one.

DR. PHURROUGH: Are there privacy issues or the concerns about things that would come out with a broad genetic test that affects large amounts of the population? In a clinical trial that would say we are going to do this clinical trial differently because of our concerns about those issues.

DR. JACQUES: Barbara, if I could? One of the things that we had mentioned a little bit earlier in the call is for example would blinding or double blinding of a trial be particularly different, knowing that the genetic information might impact people other than the subject themselves. And that was just an example.

DR. MC NEIL: Yes. Linda?

DR. BERGTHOLD: I think the answer to that is yes. What I do think we're going to see more and more -- and we're certainly seeing it in the popular press now -- is a tremendous amount of distrust of how data is used. And some of that came out in the stimulus bill of the discussion of comparative effectiveness research. And an unnamed radio host blew it all out of proportion. And so I think privacy issues are really super important to lay out.

And I can imagine a clinical trial just being destroyed because somehow sort of the privacy of the data or how the data is going to be used doesn't get clearly explained. So I think we can't

underestimate that.

DR. PHURROUGH: But does that issue change the kinds of trials that we would accept, versus we need to do the trials differently to ensure that those ethical issues are managed appropriately?

DR. BERGTHOLD: I think it could. I can't predict, but I think it could. I think it could prevent you from doing the most rigorous trials that you would like to do. And I don't know what to suggest.

DR. PHURROUGH: And if it does, then would that alter the answers to the previous question?

DR. BERGTHOLD: Yeah. Sure.

DR. PHURROUGH: Okay.

DR. MC NEIL: Marion?

DR. DANIS: I think that when you think about the ethics of research, there's not only the ethics of human subject's protection, there's the question of generating scientific data that is valuable to society to make therapeutic decisions and decisions about whether we're going to spend money for

it.

I don't think that we should undermine the rigor of the science and the demands we have for scientific rigor. I think we might have to do some special things, certificates of confidentiality, et cetera, alerting people that if they're going to be tested in the course of research to test for new genetic markers, that they should be fully aware.

The consent process has to be careful. But I don't think that we would want to say that we would accept less good genetic science because of this issue.

DR. MC NEIL: Teresa?

MS. SCHROEDER: I know the consent process has become more and more rigorous. And even for retrospective studies, all patients have to be consented. So I think the ethical issue in protecting the patient, they walk into a doctor's office and they automatically get consented for any medical institution or -- I lost my train of -- academia. You walk in, these patients are consented.

They're immediately consented for a patient going into a registry. They're consented before they go to surgery. They're consented that their data may be used for peer review journals. They're always consented before they go in.

And more and more, the consent process is getting more and more strict. Every time I go into IRB, I have to change consent form, so that those patients are protected, that their HIPPA rights are protected. So I know we go back and forth about with what data is out there and what's available, and keeping the patient protected. They are protected, I think.

I haven't seen -- maybe I'm a little bit naive, but I haven't seen a case where a patient's rights have been trampled to go forward with research. The goal is to protect that patient. We're not giving out any personal identifiers. So they are protected. So I don't understand what the --

Maybe I'm missing what the ethical issue may be. As long as these patients know what their data is

being used for, know it's being collected, know it's going into genetic, know it's going into a registry or wherever it's going, and then they're being told this is what's happening. Most patients want to be part of this. They want to know.

DR. MC NEIL: I think maybe what I heard Linda say is this is the time where we want extra security around those data. Consent aside, the issue is absolute security and privacy, I think is what you're saying. So let's see, Jim, and then Neil.

DR. PUKLIN: I was wondering if anybody here is familiar with the legal aspects that protect patients' records with regard to sensitive genetic findings that actually negatively influence their life expectancy and medical problems they're going to be facing? So that's one question.

The second thing is that the answer to your question is there are enormous ethical issues surrounding all of this. And as we just heard, the patients are signing all sorts of consents allowing the de-identification of materials so that all sorts

of tissues can be studied from a genetic perspective.

But in the clinical realm, in an age group, the gene I think that's associated with breast cancers, is it the BRAC 1 gene?

DR. MC NEIL: BRCA.

DR. PUKLIN: And so I have known several people who have that gene in their family, And the parents are in their 20s. And in order to find out whether they were positive, so as to not adversely prejudice their health insurance and their employability going forward, they actually went to Europe. They had the gene testing done, and they paid cash for it, so as to have the results hidden from their medical records. So I mean the ethical implications, the social implications are enormous in this endeavor in that age group.

In the Medicare age group, of the type that we were referring to earlier, such as a gene sequence with regard to treating a cancer where a drug is more sensitive if it has certain gene sequences that make them more susceptible to treatment or in warfarin

therapy, I think the ethical issues aren't anywhere near as great because these patients already have established disease. And the treatment options that are available to them may be enhanced by knowing what the gene sequencing issues are that are permissive of better patient care.

DR. MC NEIL: Neil? Thank you.

DR. HOLTZMAN: A few separate issues. The first is in evaluating evidence and collecting the evidence that those studies that should be accepted should all be approved by institutional review boards. Now, among academe and governmentally funded institutions, that's not a problem.

Among clinical laboratories or for-profit laboratories, that may be more of a problem, in terms of how one constitutes an institutional review board and whether it just becomes a yes group or is looking at it more independently. So I think that is an issue that should be looked at as one evaluates the studies.

The second issue is a very tough one, and that's on archived specimens. Because if you have

archived specimens that have identifiers, the use of that specimen has not been approved by the person who originally gave a specimen. But the question arises as to whether you could go back to that individual and get corroborative information depending on, say, if you were looking at the genotypes.

DR. MC NEIL: That's a good point.

DR. HOLTZMAN: There have been a number of groups that have recommended where you draw the line in something like that. One being that you attempt to find that person before you use the specimen.

The third issue that was just raised about BRCA 1 testing -- and this is germ line testing -- as to what responsibility the person who wants to have the test has. And this doesn't come up so much in the Medicare population. What responsibility the person who wants the test has to notifying relatives that information about his or her genotype may be given as a result of the test result on the proband of the index case.

DR. MC NEIL: So I've heard several things

so far. And I want to make sure that there aren't others. But one was that we have to -- while everybody has consented as Teresa said -- are in general, except for maybe the private labs, which presumably could go these for-profit IRBs.

But putting that aside, we have to worry a little bit extra about privacy, which was Linda's point. So super attention to privacy.

The issue that Jim raised was the issue of underwriting. That people could potentially have their insurance status changed as a result of genetic information. While you said it wouldn't apply to the Medicare population as much as to somebody 20 years old, there are lots of people under 65 who could be hit by that one, should that be a problem.

The other one was what was the responsibility to relatives if an individual was found to have a positive test. And I suppose that responsibility could be either on the part of the person having the test or on the part of the physician who gave the test and what was his or her obligations.

And the fourth item was what to do about archived specimens. And I guess that gets to be tricky. If the patient's dead, it's not an issue. But if they are not dead, then you've got to go track them down. And if you can track them down, what kinds of biases are introduced by that. And I have no idea. But those are the ones that are on the table so far. Yes, Linda?

DR. BERGTHOLD: Well, I really want to hammer this point just a little more.

I think Medicare actually has a responsibility to go beyond this sort of normal explanation of privacy protection. The public knows nothing about IRBs, could care less. This is a bubble in here. And people don't understand it. And I bet they wouldn't even trust it if they did. So IRBs mean nothing to the public.

HIPAA, you know, when they go into the doctor's office and sign that form, they're not reading it. I did a lot of HIPAA consulting. I know that.

So what I do know is that the issue of privacy as genetic testing becomes more prevalent, it's going to be huge as a public affairs and a communication issue. And I just think that it would be prudent for Medicare to really develop a stronger communication process around the ethical issues of privacy.

DR. PHURROUGH: I'm going to do what I shouldn't do. The person in this chair is supposed to provide an explanation and some facilitation of discussion. But I'll go on record with an opinion.

DR. MC NEIL: Go for it.

DR. PHURROUGH: Similar to the issues that Marion brought up to begin with. While there are important issues around what could be done with genetic information, I think first of all that should in no way change how trials are done. Trials should be rigorous, regardless of the types of evidence being collected. If you're going to run a trial, then you got to do the trial appropriately. And you shouldn't accept lower standards because you're concerned that

data may not be handled appropriately.

Secondly, I think it's a bit inappropriate for us to consider that we're going to let some data have less of a control from a privacy point of view. And other data needs more control. If it's private data, it's private data. And it ought to be managed as private data. And if we're not doing that well in clinical trials, shame on us. Let's fix it. Regardless of whether it's genetic data or any other kind of data, it ought to be managed appropriately.

I think the biggest issue here is less that issue of how we manage private data, but the issue of what trials we're actually doing. Are we actually creating the potential for having greater exposure because we don't design trials that answer the questions that need to be answered and, therefore, we have to move on to another trial that may or may not answer the question that we need to get answered.

And I think the real ethical issue, perhaps more so in genetic issues than in others, is that we ought to be designing the trials up front that are

going to answer the questions that need to get answered. And it's unethical to not design the trial in that manner. And if we are holding those principles up, then I think the ethics of this are less of an issue.

DR. MC NEIL: Let's see. I had Catherine, Steve, and Cliff. Is that right? Steve?

DR. PEARSON: I'll just quickly say that you have agreement from one MEDCAC member.

DR. MC NEIL: All right.

DR. PEARSON: We want to send a signal, especially to the clinical researchers in the manufacturing community, that this shouldn't be an excuse for not producing the kinds of evidence that we need to make good judgements.

DR. MC NEIL: Thank you, Steve. Cliff?

DR. GOODMAN: All the issues we've talked about had to do with management of data, not the method used to generate the data. So I concur with Steve. If we make a distinction here, somebody's going to walk into this room some day with a new test

and say, look I know my evidence isn't very good, but after all there are a lot of ethical issues surrounding this, so give me a break.

Or someone in this panel might say, you know, you need to do coverage with evidence development there, which may require setting up a registry. And they're going to say, well, you know, there are ethical issues here. I don't want to do that.

So we have said nothing here that pushes up or down any of these three hierarchies at all, whether they be analytic validity, validity, clinical validity, or clinical utility. Nothing we've said so far pushes us up or down those hierarchies. All it does is caution us yet again whether for genetic testing or others, we have to manage people's data very carefully.

DR. MC NEIL: Okay. Yes, Jim?

DR. PUKLIN: So with regard to the amount of genetic testing that needs to be encouraged, where I sit, since I'm Chairman of the IRB at Wayne State

University, I see all the research protocols.

In the last several years, every cancer study that comes from any of the cancer cooperatives, regardless of what kind of therapy they're employing, even if it's not chemotherapy, it has specimens of the blood drawn and archived for an indefinite period of time.

When there's biopsies, all of the tissue biopsies are being archived for genetic studies. They can be studied and restudied. Not only that, but in the psychiatry department, patients with certain mental diseases are having their bloods archived with manic depression, depression, drug addiction. Those patients are being studied genetically. Their specimens are being archived.

It just runs the gamut that people are looking for all sorts of genetic markers in all sorts of diseases. So I think there is a wealth of information that's going to be coming out in the next few years, just from the general research that's been done on all these widespread diseases. I've just

given you several examples.

DR. MC NEIL: Marion, final comment on this topic, unless there are burning other new issues.

DR. DANIS: Well, I was just going to say I think that no matter what design you use for your study, if you're doing a genetic test, you're doing the genetic test. And it's going to be something you need to deal with as a matter of privacy. And you might argue that the more rigorous the study, the more likely you're going to be getting informed consent.

If you take retrospective data and don't have people's permission, you might actually have less consent. And so I don't see that there should be a worry that with more rigorous design. You have more worry about the ethics.

DR. MC NEIL: Good. All right. Are there any further comments on this? I shall move on to the last question, which I think we've touched on a little bit. At least Jim did, with the underwriting issue. But I want to make sure that we aren't missing anything.

There's one obvious one, which would be that if we're looking at prognostic information, and we're starting with a Medicare beneficiary, then the length of follow-up might be less than we would want. But that's life or death. Yes, Linda?

DR. BERGTHOLD: Okay. So I'm going to put Steve on the spot, since you're in the mood to be disposing, Steve. By age, do you mean like if you're 85, you're too old? Do you mean sort of different age ranges within Medicare?

Because frankly, I think age is getting to be somewhat irrelevant, particularly, I mean, it's really fragility or - what's the word I'm looking for here? It's really your health status. There are 80 year olds that are healthier than 65 year olds. And 65 year olds that are sicker than 90 year olds. So I didn't understand what you meant here by age.

DR. PHURROUGH: I think the question is more around is there an age range at which genetic testing may not provide any information because of the age of the patient. Recognizing that at some ages, some

people are chronologically not the same age as they are functionally.

So is there a benefit in a 75 year old getting a particular genetic test if their life expectancy is 10 years, if the genetic test doesn't offer a therapy that would improve the life span longer than that 10 year life span they have?

DR. BERGTHOLD: Well, then I think age, you can answer that. But I think health status is way more important than age in that respect.

DR. JACQUES: There is one other way to look at that question too, which was also intended when we wrote it. Is there anything about older DNA that has essentially sort of churned and churned and accumulated whatever random errors may or may not occur in our genome that would, in fact, affect the accuracy or the usefulness of genetic tests, if you're doing them in a 75 year old, rather than a two year old.

DR. MC NEIL: Neil?

DR. HOLTZMAN: To answer that question, I

must say I was a little surprised when I was invited to serve on this panel because I didn't really think there was much genetic testing that would be beneficial to Medicare beneficiaries. And I still feel that way.

I mean, If one looks at genetics of disease, if we go back to the relatively rare forms, they almost all appear early or earlier than age 65. For instance, take Alzheimer's or different types of dementia. The ones that are clearly genetic with high penetrance, perhaps single gene, almost always appear before age 65.

So that the utility of genetic testing for Alzheimer's in the population over 65 becomes much less. You're not going to find much. And that is true to a lesser extent among the relatively few common disorders from which multiple gene associations have been found, that they appear generally in younger populations.

So if I had to put my money on how to improve the health of the Medicare population, I

wouldn't put much of it in genetic testing.

DR. PHURROUGH: So it shouldn't be part of the welcome-to-Medicare exam?

DR. SCHEUNER: But what about for prognostic and pharmacogenetic testing?

DR. HOLTZMAN: Okay. So you're right. Thank you very much for making that, because I think I'm dealing here with germ line mutations. And that's an important distinction that hasn't been made. When one deals with somatic cell mutations and the mutations in cancer, the age situation does not arise at all. So I would make an important distinction between germ line and somatic.

DR. MC NEIL: Deborah?

DR. SCHEUNER: No, because you could have germ line inheritable traits that influence drug metabolism. It doesn't have to be acquired. Right?

DR. HOLTZMAN: That's a good point.

DR. SCHEUNER: Yeah. I think --

DR. HOLTZMAN: I think in pharmacogenetic, I think that would be true. And, in fact, I guess in

one of the papers that we looked at, there was a more pronounced effect of one of the CYP2C9's in the older population than in the younger population.

DR. SCHEUNER: Yeah. So I think --

DR. MC NEIL: Deborah please, and then Maren.

DR. SHATIN: I beg to differ on this. And, in fact, I was kind of perturbed by the wording of this question because it's just talking about the challenges. And I would say that there may be some opportunities for the elderly population with genetic testing. And specifically I'm thinking of pharmacogenomics.

DR. MC NEIL: Well, that's what Maren's point is, I think.

DR. SHATIN: Yeah. If you could know that you'd get a severe adverse reaction particularly with polypharmacy with the elderly, that's critical. So I just would like to raise that that's the flip side to look at.

DR. SCHEUNER: I guess I have to leave, but

--

DR. MC NEIL: Yeah. I know a lot of people have to leave. So we're going to wrap --

DR. SCHEUNER: I wanted to just make one more comment about this single gene stuff. And that maybe for example with fragile x, we've learned that there are phenotypes in the elderly that we didn't really know about until very recently, with tremor in males who are carriers of the pre-mutation. So I think that maybe the focus on the rare Mendelian stuff has been in the pediatrics. But maybe we still have some more yet to learn about the older population as well.

DR. MC NEIL: So the only challenge I've heard so far -- I mean I've heard comments around the edge, but the only challenge I've heard is, is there an issue about the health of the DNA. I've heard these other questions, but they don't exactly address this particular question, do they? Mina?

DR. CHUNG: Besides some of those very rare diseases that may have some late onset, there are some

other common diseases in which some susceptibility loci have been identified, coronary disease and atrial fib.

DR. MC NEIL: How does that relate to this question?

DR. CHUNG: So I think that those should not be excluded.

DR. MC NEIL: Oh, no, no, no.

DR. CHUNG: Those would actually be very appropriate in a Medicare age population.

DR. MC NEIL: They would not be excluded.

DR. ENG: One of the challenges I see is there might be more potential for greater harm in the Medicare population, particularly in terms of pharmacogenomics.

DR. MC NEIL: Okay. Or a benefit.

DR. ENG: Pardon?

DR. MC NEIL: Or a benefit. You could go either way.

DR. ENG: Yes. That's right. There is a chance. And so because of this, then I don't think

that the Medicare beneficiary should be excluded from this.

DR. MC NEIL: I don't think that was the intent. I don't think there was ever any intent. If there were, we wouldn't be sitting here. Right?

UNKNOWN MALE VOICE: You weren't going to cover these --

DR. MC NEIL: Right. Right. Yes, Steve?

DR. PEARSON: One comment. I don't know anything about whether old DNA means that the test should have lower analytic validity or not. It's an interesting question worth finding out about.

But the other thing just to point out is that insofar as the results of genetic tests are related as relative risks or associations for elderly patients who have much higher attributable risks to other causes or other co-morbidities, et cetera, it just has to be something that's always kept in mind that the information as it's portrayed may seem to offer a health benefit.

But in an elderly population, the

attributable risk is something that really needs to be looked at very carefully, in terms of what additive information genetic testing can provide.

DR. MC NEIL: All right. All of their DNA, to attributable risk and potentially greater harms. Yes, Linda?

DR. BERGTHOLD: One more thing is the consistency of the application, whatever it is that you decide to do in terms of covering testing. What was sort of disturbing today was to here how the local, the LCPs differ in Utah and -- you know, that's something that makes a Medicare beneficiary, like me, just go crazy.

DR. MC NEIL: Well, we talked a little bit about that. That relates to the issue of screening. And the BRCA, if used as a screening test, doesn't fall -- it was a little ambiguous. I think it's unfair to pull that one out.

DR. BERGTHOLD: But just the importance of--

DR. MC NEIL: Right.

DR. BERGTHOLD: -- NCDs to set a kind of

consistency across all the contractors.

DR. PHURROUGH: That would give the Medicare a standard line. That's the way Congress intended it. So if we intend to do something else, then Congress ought to tell us to do it some other way.

DR. MC NEIL: I know people are going to leave. So I want to make sure that --

DR. PHURROUGH: It's enough of a hint.

DR. MC NEIL: No. No. One already left right in the middle of --

DR. PHURROUGH: I'm sorry. I was --

DR. MC NEIL: Right in the middle of a famous sentence I was making.

Are there any other comments on these questions or general discussion points? Then I have one final thing I want to say. Yes?

DR. GOODMAN: This question wasn't asked, but I think it's going to be relevant to Medicare coverage decisions and the kind of evidence that you might expect.

And that is, some of the genetic testing

about which we've spoken today is going to enable different kinds of evidence generation and different kinds of trial designs.

And the ability of genetic testing to allow those different kinds of trial designs such as adaptive clinical trials or some of this basing stuff, where you can actually track and see how small cohorts of patients are worth -- whether some are succeeding with a certain therapy or not, test something about their genetic profile, and then help you redesign a trial based on that information is something that will be enabled by this. It might be even an option for calling for evidence requirements, including for coverage for evidence development, by the way.

DR. MC NEIL: Good point. Yes?

DR. DANIS: I'm just coming back to the coverage with evidence development. I really think you should put that on the schedule because it's something that could really facilitate the generation of evidence.

DR. MC NEIL: I think it's on the table.

Been on the table for a couple of years. All right.

I want to say one thing I learned today, actually this morning, and this is with a lot of sadness, that our friend Steve here is abandoning us. And some of you may not have heard that. But after leading this group for seven years?

UNKNOWN MALE VOICE: Over seven years.

DR. MC NEIL: Over seven years, and being one of our major partners in this MEDCAC group, he is going to AHRQ to work in comparative effectiveness.

So for me, personally, I think this is just an enormous gain for AHRQ and an enormous loss for CMS. But I did want to tell you that and to certainly express my thanks -- and I think I can speak for everybody here -- to Steve for all he's done for us. So thank you, and good luck.

(Applause.)

DR. MC NEIL: So we are adjourned. And I believe there is a bus out there for those of you who are going to BWI. You should go. You should not wait around and talk to your close friends.

(Whereupon, the meeting was adjourned.)

CERTIFICATE OF REPORTER

I, MAUREEN O'DONNELL, a Certified Verbatim Reporter, do hereby certify that I took the stenographic notes of the foregoing proceedings which I thereafter reduced to typewriting; that the foregoing is a true record of said proceeding; that I am neither counsel for, related to, nor employed by any of the parties to the

action in which this proceedings were held; and, further, that I am not a relative or employee of any attorney or counsel employed by the parties hereto, nor financially or otherwise interested in the outcome of the action.

_____

Maureen O'Donnell, CVR
Certified Verbatim Reporter