CENTERS FOR MEDICARE AND MEDICAID SERVICES

Medicare Evidence Development & Coverage

Advisory Committee

March 24, 2015

Centers for Medicare and Medicaid Services

7500 Security Boulevard

Baltimore, Maryland

Panelists

Acting Chairperson
I. Craig Henderson, MD

Voting Members
Harry Burke, MD. PhD
Josef E. Fischer, MD
Mark D. Grant, MD, PhD
Beverly A. Guadagnolo, MD, MPH
Mitchell R. Kamrava, MD
Marcel Salive, MD, MPH
Maren T. Scheuner, MD, MPH
Diana Zuckerman, PhD

CMS Liaison
James Rollins, MD

Industry Representative
Lakshman Ramamurthy, PhD

Guest Panel Member
Barry M. Berger, MD, FCAP

Invited Guest Speakers
Barbara Conley, MD
Jeffrey S. Ross, MD

Executive Secretary
Maria Ellis

TABLE OF CONTENTS

PANEL PROCEEDINGS

(The meeting was called to order at 8:14 a.m., Tuesday, March 24, 2015.)

MS. ELLIS: Good morning and welcome, acting committee 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 selected molecular pathology tests for the estimation of prognosis in common cancers, adenocarcinoma of the colon and rectum, breast cancer, invasive duct and lobular cancers, non-small cell lung cancers.

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 interests. 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 disclose if you or any member of your immediate family owns stock or has another formal financial interest in any company, including an Internet or e-commerce organization, that develops, manufactures, distributes and/or markets consulting, evidence reviews or analysis or other services related to selected molecular pathology tests for estimation of prognosis in common cancers, adenocarcinoma of the colon and rectum, breast cancer, invasive duct and lobular cancers, non-small cell lung cancers. This includes direct financial investment, 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 limit. We have numerous

presenters to hear from today and a very tight agenda, and therefore, cannot allow extra time. The light will begin flashing -- I'm sorry -- 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 Dr. Harry Burke, Dr. Josef Fischer, Dr. Mark Grant, Dr. Beverly Guadagnolo, Dr. Mitchell Kamrava, Dr. Marcel Salive, Dr. Maren Scheuner, and Dr. Diana Zuckerman. A quorum is present and no one has been recused because of conflicts of interest. The entire panel, including nonvoting members, will participate in the voting. The voting results will be made available on our website following the meeting. I ask that all panel members please speak directly into the mic, and

you may have to move the mic since we have to share.

The meeting is being webcast via CMS Live in addition to being recorded for the transcription. By your attendance, you are giving consent to the use and distribution of your name, likeliness and voice during the

meeting. You are also giving consent to the use and distribution of any personal identifiable information that you or others may disclose about you during today's meeting. Please do not disclose personal health information.

In the spirit of the Federal Advisory Committee Act and the Government in the Sunshine Act, we ask that the advisory committee members take care that their conversation about the topic at hand take place in the open forum of the meeting. We are aware that members of the audience, including the media, are anxious to speak with the panel about those proceedings. However, CMS and the committee will refrain from discussing the details of this meeting with the media until

its conclusion. Also, the committee is reminded to please refrain from discussing the meeting topic during breaks or lunch.

If you require a taxicab, there are telephone numbers to local cab companies at the desk outside of the auditorium.

Please remember to discard your trash in the trash cans located outside of the room.

And lastly, all CMS guests attending today's MEDCAC meeting are only permitted in the following areas of CMS single site: The main lobby, the auditorium, the lower level lobby and the cafeteria. Any persons found in any area other than those mentioned will be asked to leave the conference and will not be allowed back on CMS property again.

And now I would like to turn the meeting over to Dr. James Rollins.

DR. ROLLINS: Good morning. My name is Jim Rollins, I'm the director of the Division of Items and Devices in the Coverage and Analysis Group.

Originally this MEDCAC was scheduled for the spring of 2014. A technology

assessment was performed that reflected the medical literature at that time but because of matters beyond our control, it was postponed. In today's MEDCAC the technology assessment has been updated to reflect the state-of-the-art activities related to this topic.

And now I will turn things over to Dr. Henderson.

DR. HENDERSON: So, I'm Craig

Henderson, from University of California San Francisco, where I'm a professor of medicine and a medical oncologist. I think the first item should be to introduce other members of the panel, let them introduce themselves, and we will go from left to right here.

DR. BURKE: I'm Dr. Harry Burke, an associate professor of medicine at the Uniformed Services University of Health Sciences. I am representing myself, my views are not those of the university, Department of Defense nor the U.S. Government, and I have no conflicts of interest.

DR. HENDERSON: Okay, I just want to remind each person, and you were very good, I

have no conflicts of interest, and each one of you should mention specifically your conflicts of interest as you introduce yourself.

DR. FISCHER: I have no conflicts. I am the McDermott professor of surgery at Harvard Medical School.

DR. GRANT: I'm Mark Grant, I'm the director of the Technology Assessment Center for Clinical Effectiveness at Blue Cross Blue Shield Association. I'm here representing

myself and have no financial conflict of interest to report.

DR. GUADAGNOLO: I'm Ashley Guadagnolo, I am an associate professor in the division of radiation oncology at the University of Texas M.D. Anderson Cancer Center. I represent myself and I have no conflicts of interest.

DR. KAMRAVA: I am Mitch Kamrava, a radiation oncologist at UCLA. I represent myself, with no conflicts of interest.

DR. SALIVE: Good morning, I'm Marcel Salive, from the National Institute on Aging at NIH, representing myself, and I have no

conflicts of interest.

DR. SCHEUNER: Hi, I'm Maren Scheuner, I am chief of medical genetics at the VA Los Angeles. I'm also a professor in the department of Medicine at the David Geffen School at UCLA, and I'm representing myself. I have no conflicts of interest.

DR. ZUCKERMAN: Dr. Diana Zuckerman. I'm the president of the National Center for Health Research and I have stock in Johnson & Johnson, which I have just been told may be

related to some of these topics.

DR. RAMAMURTHY: Lakshman Ramamurthy, I'm a director at Avalere Health, which does consult with a number of companies. I am representing myself here and I have no conflicts of interest. Thank you.

DR. BERGER: Good morning, I'm Dr. Barry Berger, I'm the chief medical officer at Exact Sciences Corporation. I am here representing myself and I have no conflicts of interest for the matters at hand.

DR. HENDERSON: Thank you. So we will, before we have our first presentation I

just want to make a couple points to emphasize.

First of all, our discussion here is not directed towards coverage decisions in any way. Rather, we will focus on the questions that are specifically addressed to the panel, representing the questions that CMS has asked us to address because of our individual and collective expertise in these areas.

Secondly, I want to emphasize that we are going to focus on just three tumor types. In her opening remarks Maria emphasized those, and they are lung cancer, colorectal cancer and

breast cancer. And we will be focusing predominantly, or exclusively, on the markers that are listed here and their prognostic importance.

So with those opening remarks, I think we will proceed, then, to the first presentation, which will be by Cheryl Gilbreath on the CMS presentation and voting questions. Dr. Gilbreath.

DR. GILBREATH: Thank you and good morning, ladies and gentlemen. My name is Dr. Cheryl Gilbreath and I am the CMS analyst

for this MEDCAC regarding molecular pathology testing to estimate prognosis in cancer.

So, considering the evidence, today I will briefly discuss some background information regarding CMS's consideration of evidence thus far and will introduce the questions for this MEDCAC meeting.

CMS and genomic testing. Today's MEDCAC meeting is one of a series of five MEDCAC meetings over the last six years which have focused on various topics related to genomic testing. In 2009 we had genetic genomic testing as well as screening genetic

tests, in 2010 we did pharmacogenomic testing in cancer, 2013 we did genetic testing for cancer diagnosis, and today we are working on molecular diagnostic tests to estimate cancer prognosis, so prognostic versus diagnostic.

A prognostic test is a laboratory or imaging test to aid in the diagnosis or detection of disease in a beneficiary with signs or symptoms of an illness or injury. In general a prognostic test, in contrast to a diagnostic test, measures or detects one or

more markers called biomarkers that can affect prognosis in a beneficiary with a personal history of cancer.

For today's purposes, molecular pathology testing to estimate prognosis in cancer provides information about the likelihood of recurrence after treatment, death from cancer, or death due to another cause after the diagnosis of a cancer.

What factors affect prognosis? Several factors can affect prognosis and can be used to inform physicians' decisions and to advise patients about the likely courses of their diseases. For example, cancer-related

factors include the type, stage and location of the cancer, treatment-related factors include the body's response to prior therapies, and patient-related factors could include the patient's overall health, comorbidities, as well as the patient's personal values and wishes regarding cure versus palliative care.

Prognostics reviewed. Tests for genomic factors of cancer cells are being actively investigated as estimators of

prognosis. Some are designed to detect recurrence at an early point of time. Others offer views of cancer aggressiveness and may alter the physician's management. The molecular tests reviewed today estimate prognosis in cancer types frequently encountered in adult patients, colorectal, breast and non-small cell lung cancers. Genomic-based prognostic assessments of other cancer types are currently under investigation.

The outcomes of interest for CMS. Well, when CMS asks what effect do the results of these tests have in patient outcomes, we are especially interested in the outcomes shown here, overall survival, mortality, avoidance of

harm of anticancer treatments, quality of life and others.

So what is reasonable and necessary? Well, CMS assesses the evidence to determine whether it is adequate to conclude that the item or device, or specifically in this case the molecular test for cancer prognosis, leads to improvement of clinically meaningful outcomes in our beneficiaries. Today we will

be asking the MEDCAC panel to listen to the presentations of the evidence and advise CMS on how best to proceed.

So the MEDCAC questions, I will read for you to keep under consideration during the subsequent presentations: For the panel's reference, a crosswalk is supplied to help relate the MEDCAC voting questions and the TA key questions which will be discussed after my presentation.

Here is a voting question grid. The MEDCAC panelists will vote on each prognostic test per question. This is a sample of the voting grid to identify your confidence level or score for each question regarding each prognostic test. If the mean average panel

score for a certain prognostic test is above 2.5, then we will continue the voting questions for that particular prognostic test. If the mean score is below 2.5 then we will stop the voting for that particular test. The TA to MEDCAC crosswalk may be used as a reference during your voting.

Other acronyms used. Well, we all are

aware of FDA, CDC, and there are quite a few other acronyms which are going to be used throughout the meeting. For your reference we've also supplied a working glossary of several of these other terms that will also be used during the meeting. For example, test validity measures, today's voting questions will focus on analytic validity, which is a test's ability to measure the genetic trait of interest; clinical validity, the test's ability to identify or predict the disease or condition of interest; and clinical utility, which is the balance of benefits and harms when the test is used to influence patient management. This information is also available in more detail in the reference glossary that was provided.

On to the questions. MEDCAC question

1(a). For each prognostic test listed, how confident are you that existing evidence is sufficient to confirm the analytical validity of the molecular pathology test to estimate prognosis for Medicare beneficiaries with that cancer type?

Again, if the answer for question 1(a)

is at least in the intermediate range, a mean score of 2.5 or more, then we will move on to question 1(b). If not, we will save that particular test for the discussion question number 4.

Question 1(b). For each prognostic test listed, how confident are you that existing evidence is sufficient to confirm the clinical validity of the molecular pathology test to estimate prognosis in Medicare beneficiaries with that cancer type?

As mentioned before, if the mean score is 2.5 or above, we will move on to question 2 for that particular prognostic test.

Number 2. How confident are you that there is sufficient evidence to conclude that using the molecular pathology test to estimate prognosis affects health outcomes, including

benefits or harms, for Medicare beneficiaries with cancer whose anticancer treatment strategy is guided by the test's results? If the range is above 2.5, we move on to question 3. If not, we skip to question 4.

Voting question 3. How confident are

you that there is sufficient evidence to conclude that using the molecular pathology test to estimate prognosis has clinical utility, meaning that it improves health outcomes either due to increased benefits and/or reduced harms for Medicare beneficiaries with cancer whose anticancer treatment strategy is guided by this test's result? Again, you vote, and if the mean score is 2.5 and above, that's great, because we're going to move on to question number 4, the discussion question.

The discussion question: Please discuss whether the following factors change generalizability of evidence about molecular diagnostic tests estimating cancer prognosis. (A), regulatory status of the test, i.e., FDA, or approved/cleared by the FDA versus LDT; (b), performing laboratory type, i.e., academic medical center laboratories, independent

commercial laboratories or other; (c), demographic subgroups within the Medicare beneficiary population; and (d), cancer genomic characteristics.

This is the conclusion of my

presentation. I will invite Sreelatha Meleth to give the TA presentation. Thank you.

DR. MELETH: Good morning. My name is Sreelatha Meleth, I am representing the RTI-UNC EPC, evidence-based practice center, and we conducted the technology assessment for the, to look at the tests for the estimation of prognosis of common cancers.

A lot of people helped us, this is a fairly extensive review, and I just wanted to start off by acknowledging that, also acknowledging that the funding, was based on the AHRQ contract, and none of the investigators had any conflicts of interest.

For the actual technology assessment, I understand that we've changed the focus of the cancers a little bit, but the actual TA looked at four cancers and we looked at 11 tests, and the objective was to conduct this systematic review assessing the 11 molecular

pathology tests that might inform estimation for prognosis. The overarching question was whether there is direct evidence that the addition of the results of the tests would give

you more information about the prognosis in addition to the traditional prognostic markers that are traditionally used, and would improve clinical outcomes for adult patients.

So, the table there shows you the tests that we looked at. For breast we looked at MammaPrint and Oncotype Dx. For colorectal cancer we looked at BRAF, KRAS, MSI, MLH1 and Oncotype Dx. Then for the objective lung we looked at EGFR, KRAS and ALK translocation, and for bladder we also looked at UroVysion.

Just a brief background. It is estimated that there will be approximately 1.67 million new cases of cancer in 2014. It's the second leading cause of death in the United States. However, the death rates in general for cancers are declining. In the last five years death rates decreased 1.8 percent a year for men and 1.4 percent a year for women. The declines are largely due to the declines in death rates in the four major cancers. In the

last five years lung cancer death rates are down 34 percent for men, nine percent for women; 34 percent, 45 percent and 46 percent

respectively for breast, prostate and colorectal cancer.

And part of this improvement in survival can be attributed to advances in molecular pathology, which has resulted in better understanding of cancer subtypes and development of treatments based on the subtypes. For example, the identification of the human epidermal growth factor 2 receptor resulted in targeted therapies for breast cancer. Advances in molecular pathology have also helped identify tumor characteristics that help predict prognosis for a patient in addition to traditional markers such as stage and differentiation.

We would like to, before we go into the matter and results, clarify that the two, questions 2 and 3 in MEDCAC, the questions that were just presented, ask about the anticancer treatment therapy being guided by these tests. It's important to know that these genetic tests were used in two different contexts. In one

the tests are used in a specific context of a test and therapy combination where the test is

being used to predict response to that particular therapy, so it's a way of looking at the test with respect to a very specific treatment.

In the second context these tests are used to estimate a patient's prognosis and physicians use this information in a variety of ways, including informing choices from a variety of different treatment options.

CMS requested this report to evaluate the second context, not the first. Therefore, studies that evaluate specific test/therapy combinations were excluded from this review.

So, we'll just briefly go through the methods. We started off with key questions that were then refined through discussions within, between the EPC, AHRQ and CMS, developed an analytic framework, we searched the databases, we looked, did a systematic review of published evidence, summarized evidence qualitatively and quantitatively, with a meta-analysis when that was appropriate.

The methods used in the review were

based on the framework developed at the CDC,

the Evaluation of Genomic Applications in Practice and Prevention, EGAPP, by that working group which was established in 2005 to develop a systematic process for evidence-based assessment that is specifically directed towards, focused on genetic tests and other applications of genomic technology.

The methods developed and recommended by this working group share many elements with other existing processes, such as the USPSTF and the AHRQ Evidence-Based Practice Center. It also recognizes that the gold standard for direct evidence, randomized clinical trials, may not be available in the evidence base for these new tests, and outlines the process for building a chain of evidence.

And that chain of evidence is also called the ACCE model, or the A-C-C-E, which stands for the analytical validity, clinical validity, clinical utility and, where applicable, ethical, legal and social implications, so those are the sort of pillars that we use to build the evidence.

Analytic validity is the technical

performance of the test, so does the test actually measure what it's supposed to. The common ways in which we assess that are sensitivity and specificity.

Clinical validity is looking at the strength of the association between a genotype and a result of interest. The strength of this association determines the test's validity to diagnose a disorder, assess susceptibility or risk, or provide information on prognosis or variation in drug response.

Clinical utility looks at the evidence that the test results can actually change a patient's management decisions and improve net health outcomes down the road.

So, how did we apply the ACCE model? So based on the working group's recommendation we first developed an overarching question that sought to find direct evidence addressing our primary question. And then we created a set of sub-questions based on the ACCE model that would help build the chain of evidence that could help answer the overarching question indirectly.

So what was the overarching question?

The overarching question was, is there direct evidence that the addition of the specified molecular pathology tests used alone or in combination with traditional prognostic factors changes physician decision-making and improves outcomes for adult patients with colorectal cancer, breast, lung or bladder cancer, compared with the use of traditional factors to predict risk of recurrence for adults with these cancers? Ideally we hoped to find published evidence that directly answered this key question.

In the absence of direct evidence, in the event that there would be an absence of direct evidence for the key question one, we developed key questions to build a chain of evidence that would help answer KQ1, and the chain again was based on the ACCE model. So the analytic validity question was, does the existing evidence establish technical accuracy and reliability of these tests for detecting the relevant molecule analytes? The clinical validity question was, does existing evidence establish the prognostic accuracy of the test for predicting recurrence? And clinical

utility was, does existing evidence support clinical utility of the molecular pathology tests?

Clinical utility was further refined into the impact on physician decision-making and patient centered harm.

I have the red light flashing, is that actually right? Okay.

What is the evidence that prognostic information, the first question for clinical utility was, is there evidence that the prognostic information provided by these tests modifies physician decisions regarding use of adjuvant antineoplastic chemo and/or radiotherapy, enhanced diagnostic testing for recurrence, and/or surgery among adult patients with malignant tumors?

And the KQ4b was, what is the evidence that modified decisions leads to improved outcomes, including patient-centered outcomes such as improved quality of life, reduced disease recurrence, increased overall survival or disease-free survival, or reduced therapeutic side effects?

And KQ5 looked at the harms associated

with treatment decisions informed by the molecular pathology tests.

So based on that, we created an analytic framework where, we are looking at those, so that's how we expected to come to our conclusion, to assess our evidence and go throught the results. So we have the cancer, we have the tests, and then you're looking at whether the test is actually doing a good, you know, detecting what they wanted to detect, and then we're looking at whether it actually predicted risk of recurrence of prognosis, and then looked at treatment decisions, and then looked at health outcomes for the patients, and also whether patients were harmed.

We searched PubMed, the Cochrane Library and EMBASE for English-language studies published through November 2013. We also searched the reference list of pertinent review articles and studies that met our inclusion criteria. We searched as well for unpublished studies relevant to this review, for which we used relevant websites, clinicaltrials.gov, the FDA website, Health Services Research Projects

25 in Progress, and the European Union Clinical

1 Trials Register. We also requested information
2 from the College of American Pathologists and
3 from relevant companies, asking for data that
4 they believe should be considered for the
5 review.
6 A very brief look at the population
7 and intervention, so the population includes
8 adult patients with one of the cancer types of
9 interest that evaluated an eligible test. The
10 comparators for KQ1, 4 and 5 included studies
11 that compare at least one of the tests plus
12 standard prognostic factors with the standard
13 prognostic factors alone to determine whether
14 the test adds independent prognostic value or
15 benefit, or introduces additional harms.
16 We did not include studies focused on
17 patients with advanced or metastatic cancer or
18 studies focused on predicting response to
19 treatments. And just in case there's a
20 question, because the task of the technology
21 assessment was focused on risk of recurrence,
22 so it didn't seem relevant to look at advanced
23 metastatic cancers, and we already talked about

response to treatments.

The comparators for KQ2, which is the

analytic validity question, we included studies of test performance, including intra/inter-lab reproducibility for the included tests. For KQ3, which is clinical validity, we included studies comparing patients with different test results, example, those with a mutation versus those who are wild-type, to establish prognostic value, with a multivariate analysis to adjust for known factors. We required that the results were either adjusted for known factor using a model, or were specifically addressed in other ways which could be either the inclusion/exclusion criteria for the study or something like stratification.

So, all studies, for everything we had independent dual review to assess for eligibility. If there were conflicts, they were resolved by discussion. We used structured data extraction forms. One team member abstracted the data and a second reviewed the data for accuracy.

To assess the risk of bias we followed

the Methods Guide For Medical Test Reviews and the AHRQ Methods Guide for Effectiveness and Comparative Effectiveness Reviews and the RTI's

own question bank.

For analytic validity we used relevant questions from QUADAS-2 to assess the potential for bias due to flaws in the sample selection, testing protocol, reference standards, verification procedures, interpretation and analysis. For clinical validity and utility we assessed the potential for selection bias, confounding, performance bias, attrition bias, and detection bias.

Two independent reviewers assessed each study, assessed as low, medium, high or unclear, and conflicts between reviewers was resolved by discussion, and a consensus was reached.

The strength of evidence, overall evidence for our particular questions was graded as high, moderate, low or insufficient. We used the guide established for the EPC program and, which incorporates four key domains, that is the risk of bias which we

assessed individually for each study; that includes study design and aggregate quality, consistency, directness, and precision. Once again, two reviewers assessed each domain for

each key outcome and determined an overall grade based on domain ratings, and differences were resolved by a consensus discussion, or, if that was not possible, by consulting with a third investigator.

For the data synthesis we had enough evidence to do an actual meta-analysis only for KQ3, and so we estimated a summary hazard ratio for outcomes of any given test-cancer pair with three or more independent adjusted hazard ratios. We tested the null hypothesis of homogeneity of effect sizes across the studies for each of the outcomes. If effect sizes were nonhomogeneous then we used random effect models to create a summary effect size. If they were homogeneous then we used a fixed effects model to create that.

I'm not sure we need to go through this. We started off with 5,445 records from the three databases, of which 1,884 were

duplicates, and so we removed -- let me go through that. We removed 2,702 duplicates, then 3,850 were assessed, and from that abstracted, and then we reviewed 1,828 full text articles, out of which we included 112,

and the reasons for exclusion are here.

So we're getting into the results now. So for analytic validity we had fairly limited data on analytic validity in the published literature. Therefore we, in order to supplement that, we approached CAP to give us proficiency test results for five tests. CAP focuses on inter-laboratory reproducibility. Based on CAP evidence, BRAF, EGFR, KRAS, MSI and UroVysion are reported to have between 90 to 95 percent inter-lab reproducibility. Oncotype Dx is reported to have high inter-lab reproducibility by Genomic Health.

We will go through the clinical validity for each cancer. So for breast cancer we are looking at MammaPrint where we took that as, the poor prognosis versus good prognosis. Evidence from multiple studies supports association between test result and prognosis

for risk of recurrence and cancer specific survival. There was only a single out of those studies for overall survival.

Just to let you know, I should have probably mentioned this when we were looking at risk of bias. Any study that had an unclear or

high risk of bias was excluded from the summary of the results.

So you can see the hazard ratio there, the summary hazard ratio, and the Ns for the different studies and the number of studies that went into our assessment.

For Oncotype Dx where we were comparing high risk to low risk evidence, there was an association between test results and prognosis, again, for risk of recurrence and cancer-specific survival, and there was just one single study again for overall survival. And once again, you have the number of studies that we looked at, and that was the total N for all of those studies together.

The clinical validity for lung cancer, we looked at EGFR testing, mutation versus wild-type, and KRAS mutation testing mutation

versus wild-type. Six studies looked at the prognostic value of EGFR for risk of recurrence, with a total N of 1,870, and overall survival 1,820, and the summarized evidence suggests no prognostic value.

Again, this is not, just to remind you, this is not looking at impact of treatment

or the impact of EGFR on response to treatment.

Some evidence that KRAS testing had prognostic value, so the results of that were that we had studies that looked at risk of recurrence and overall survival. For risk of recurrence we had enough studies to do a meta-analysis, for overall survival we just had two studies, so we just listed the hazard ratios there for the two studies.

The clinical validity for colorectal cancer looking at BRAF mutation testing, wild-type versus mutation, evidence suggested added prognostic value in colorectal cancer for cancer-specific survival and overall survival. It was not significant for risk of recurrence.

Again, KRAS mutation testing, wild-type versus mutation, evidence suggested

no added prognostic value for KRAS mutation testing in colorectal cancer for either risk of recurrence or overall survival. There was added prognostic value for cancer-specific survival.

And for MSI testing where we were looking at microsatellite high was stable, and evidence suggested prognostic value in

colorectal cancer for all three of the outcomes of interest, risk of recurrence, cancer-specific survival and overall survival.

Oncotype Dx, there was just one study with 690 patients, and there was no published evidence that met our criteria for other outcomes.

UroVysion, we did look at bladder cancer, but there is a caveat which the company has been very vocal about letting us know, that UroVysion was not, it was designed as a diagnostic test, it was not designed to assess prognosis. Despite that, there is some, very limited evidence that it may be useful in predicting risk of recurrence. There were no studies for cancer-specific survival or overall

survival.

For patient outcomes there were no published studies that assessed the impact of the test on long-term outcomes for the patients, example, the impact on risk of recurrence or survival after the test is done, whether the actual doing of the test impacts recurrence or survival. Even in cases where the tests seemed to provide added value in

determining prognosis, that is, even though there was evidence of clinical validity, we found no evidence that using the test was related to improved outcomes for patients.

There was moderate evidence about treatment decisions for one of the tests, that is Oncotype Dx. We found that there is evidence that it leads to changes in decision-making. And although the decision changes were observed in both directions for individual patients, studies consistently showed an overall shift to less intensive treatment recommendations as a result of using the test, with fewer recommendations for chemo, and therefore potentially less harm, exposure

to chemo and the harms. But studies did not actually follow patients to actually report on harms, or to assess the overall balance of clinical benefits and harms.

There were no studies that focused specifically on the Medicare population or assessed the prognostic value of the tests stratified for the Medicare population. However, almost all studies included patients from the Medicare population, and we found no

evidence to suggest that clinical validity would differ for this population.

So, the limitations. Many of the included studies had methodological limitations, including some risk of bias. For example, most of them were observational studies assessing associations between test results and outcomes, and are susceptible to potential confounding. There were no studies specific to the Medicare population. Many of the included tests are currently used to predict responses to specific treatments, that was not evaluated in this report. Determining whether the tests have clinical utility for

predicting therapeutic response is beyond the scope of this review.

So in summary, the weight of published literature to date has focused on the clinical validity of the tests of interest. Relatively little emphasis on how these tests can be incorporated into the overall care of patients in terms of changing decisions or the effect of those changed decisions on downstream patient-centered outcomes.

Oncotype Dx was the exception here,

with a relatively large number of studies showing an impact on treatment decisions resulting in fewer recommendations for chemotherapy, but there is still insufficient evidence on downstream outcomes.

So, the conclusions for clinical validity. Good evidence supporting added prognostic value beyond traditional prognostic factors for the following tests for risk of recurrence, cancer-specific survival, and/or overall survival, so those were MammaPrint and Oncotype Dx for breast; KRAS mutation testing for lung; and BRAF, KRAS and MSI for colorectal

cancer.

Clinical utility. Oncotype Dx for breast leads to changes in treatment decisions, and we had no studies that directly assessed the impact of test use for any of the included tests on downstream health outcomes to establish clinical utility.

And I don't think the questions are now, I think the questions will be later. Thank you so much for your attention.

DR. CONLEY: Good morning. My name is Barb Conley, I'm, my day job is the associate

director of the cancer diagnosis program at the NCI, but I have no disclosures, and my opinions of course don't represent those of the NCI or DHHS necessarily.

So, I was given the task to sort of review this huge number of tests in half an hour and discuss the clinical validity and clinical utility, as well as the analytical validity of all of them, and I will do my best, but fortunately the previous speaker did lay it out quite nicely. The point of view that I'm going to look at is more that of the medical

oncologist probably, in these tests.

So I start out with colorectal cancer and you've heard the number of analytes that are under discussion today, so 40 percent of patients with colorectal cancer will have KRAS mutations, but 14 percent will also have other RAS mutations. Clinically, you know, it's not that associated with prognosis, but the way you would use it clinically is in the metastatic sense, metastatic condition, these mutations can predict nonresponse to EGFR monoclonal antibodies, and there's an ever increasing number of RAS and other mutations that seem to

predict nonresponse. It doesn't quite work the other way, that if you have normal genes, that you will respond.

So there is a, an FDA-approved test for these mutations in KRAS to use in identification in patients with metastatic colon cancer for treatment with cetuximab or panitumumab, and there are other tests around as well as tests done in various laboratories.

The clinical validity and clinical utility were laid out basically in a

prospective way in three trials that had already been completed, but then the patients were analyzed for RAS mutations and their response or nonresponse was then gauged, and these trials are listed here. There was the CRYSTAL trial using FOLFIRI with or without Cetuximab, the PRIME trial using any chemo with or without panitumumab, and now there's OPUS using FOLFOX4 with or without cetuximab. So in all of these trials, there were no benefit to patients who had a mutated KRAS.

But there are some questions remaining clinically. We do know that some RAS mutations other than KRAS codon 2 seem to indicate that

their patients will have no benefit to EGFR receptor inhibitors, particularly monoclonal antibodies, but, you know, we don't know what percentage of alleles that have these mutations, for example, is it five percent of the KRAS mutated enough, or do we need 50 percent or more. There's some preanalytic variables to pay attention to, percent of tumor. If the test is done by what some people think is the gold standard, Sanger, you're not

going to get below the 20 percent prevalence there. And then there's the question of LDTs, which was addressed in the previous presentation, and the College of American Pathologists has addressed that one.

So for BRAF, five to 15 percent of patients with colon cancer have BRAF V600E mutations. There are other BRAF mutations, of course, and this seems to be a very strong negative prognostic factor and it's often associated with MMR deficient somatic tumors, not particularly the MMR deficient dermoid tumors. We don't really know what to do with this. There's been some trials addressing can we, do we have drugs that can address this BRAF

confusion, but they don't work as well as they do in melanoma for colon cancer. So there's no particular platform or assay recommended by the NCCN for this.

There is an assay kit that's approved for melanoma V600E or K mutations, not other BRAF mutations and not colorectal cancer. The M.D. Anderson looked at the prognostic capability of BRAF mutations to predict in

colon cancer patients, and noted the shorter progression-free survival in patients who had BRAF mutations. NCALGB study 80903 also saw something in that direction.

I'm going to spend a little more time on MSI because, it stands for microsatellite instability, and it turns out that in some particular genes in the coding region, there are nucleotide repeats, there are small numbers of repeats, and these represent an area where you can have mismatched repair and normally the body will fix that mismatched repair so that you don't get mutation. But if you have a defect in the mismatch repair capabilities, then you are prone to get further mutations and that is thought to lead to cancers, as well as

potentially lead to changes in the behavior of a cancer.

So, there is a syndrome called Lynch syndrome, and it is tied to mismatch repair problems, and the reason why you want a diagnosis is because, of course, it's a hereditary syndrome, it's passed down in families and ideally, if you can monitor these

patients and catch the cancers early, they can be cured.

We do know that on discovery of this syndrome, or at least the mismatch repair which is either genetic or semantic, that patients who have this mismatch repair deficit are unlikely to benefit from this and rely more on adjuvant treatment for colorectal cancer. It is not quite certain whether they would benefit from the more modern adjuvant treatment or not, which includes oxaliplatin, and of course we're hoping that it will actually pan out in adjuvant treatment of colorectal cancer, but it's possible that patients who have a mismatch repair defect might be more responsive to this treatment, and there's new treatments being developed for this mismatch repair deficiency

state. Prognosis-wise, though, mismatch repair defects seem to have a better prognosis in stage II or III detectable colon cancers.

So, they do have certain pathology features that you can note. There's tumor infiltrating lymphocytes, the watchword these days for possible response to checkpoint

inhibitors. They have a mucinous signet ring appearance. They could have Crohn's-like lymphocytic reaction and in fact mismatch repair defects have been noted in Crohn's disease, and they seem to have medullary growth.

There are certain DNA repair genes that are the subject of these studies and they are listed up in here. The hereditary form exists in three to five percent of the colorectal cancer patients but the sporadic form also occurs in ten to 15 percent of cases.

They tend to occur in women of older age on the right side of the colon, and potentially with inflammatory conditions, and they're associated with a certain phenotype called the CpG island methylator phenotype, which tends to produce some of these mutations.

In the '90s and revised in the early 2000s, the Bethesda Guidelines for picking out patients clinically that might have a hereditary syndrome, a Lynch syndrome, and that is early age of onset of colon cancer, less than 50; if they've had synchronous or

metachronous Lynch syndrome tumors regardless of age, and I will go over that in a second; if they have the histology that would be consistent with Lynch syndrome when their age is less than 60; if they have more than one first degree relative with a Lynch syndrome cancer, one of which is diagnosed at less than age 50, or two or more first or second degree relatives with Lynch syndrome cancers regardless of the age.

So this is not a perfect set of criteria, these Bethesda criteria. Up to 50 percent of Lynch syndrome patients don't actually meet the Bethesda criteria. 90 percent of them, though, are MSH high and they lack expression of at least one mismatch repair protein by immunohistochemistry.

So germline mutations of the mismatch repair genes are detected in at least 50, or

more than 50 percent of patients that actually do meet the Bethesda criteria, and the key here is that the lifetime risk of colon cancer in these patients is 80 percent, but a surveillance and removal of premalignant

lesions lowers the risk. Also, endometrial cancer is a little more tricky there. So looking at screening of patients with tumors, of screening tumors of patients who do meet the Bethesda criteria have been shown to be cost effective.

So, what is that screening? So, if you think about the mismatch repair genes there are two mismatch repair genes, MLH1 and MSH2, both of whom require partners to be able to work, and these are the most commonly affected in the genomic or hereditary situations. And so early colonoscopy, very early colonoscopy is recommended, and more frequent colonoscopy is recommended in this situation.

There are also endometrial and ovarian cancers that can be more frequent in this syndrome, and basically here we are left with enhanced attention to the symptoms of such diseases, and in considering surgical removal

of the uterus and ovaries on completion of child-bearing.

This is a partial list, probably, of the Lynch syndrome associated cancers, and you

can see the GI tract is represented in there, also the genitourinary tract and sebaceous gland adenomas and keratoacanthomas, as well as the brain.

So, there's some Lynch-like cancers that we didn't quite find the reason for, they're MMR deficient but they don't have hypermethylation of MLH1 promoter or any MMR mutation that we can find.

In the somatic pathways, as we noted before, they tend to have BRAF mutations in the vast majority of these patients.

So if the patient has, how do you test them? If the patient has a known familial history with MMR tumors, sequencing would be helpful. If you have (inaudible) deletions, though, that can cause a little bit of a false negative there.

If they don't have known familial history then you can use IHC or MSI, PCR, but there's about a ten percent false negative and

the sensitivity, as you see there, is reasonably good.

You can also test the BRAF mutation

and if you have a BRAF mutation, then it's likely that your tumor is not hereditary but it's a somatic tumor. About seven percent of colorectal cancer patients who have MMR actually do have Lynch syndrome.

So, the testing of MSI for a PCR tumor or normal, there are various markers that are suggested to be tested, none of them are perfect, none of them absolutely have to have normal tissue, and some of them don't. If you have MSI-H, which is really the only category that you need to distinguish from other, then there's instability in two or more of these markers that's not specific for Lynch syndrome, and it might underestimate that, and various ones of these markers have various sensitivities, as noted there.

So for MLH1 and MSH2, sensitivity is about 90 percent, but it's a little bit less for MSH6 and PMS2. The specificity is about 90 percent, however.

Promoter methylation is another way,

because promoter methylation would be presumably a way of getting your somatic MSI

behavior. The A region methylation is seen in 16 percent of the stable patients, and the C region methylation is common in MSH high. You can also do bisulfite conversion followed by real time PCR for this methylation.

IHC is a widely available alternative and it actually looks at the function, or actually it looks at the presence of the protein, it doesn't really look at the function of the protein, so it's possible that you might have a mutation affective protein function but not the energetic domain, in which case that would result in a false positive.

The other thing to note is that both MSH2 and MLH1 have obligate dimers and therefore, if they are mutated or methylated, the dimers will also not be seen.

There's a ten percent false negative in IHC testing.

So there's been some suggestion that there should be reflux testing for Lynch syndrome, and test everybody who has an early diagnosed colorectal cancer. EGAPP has found

this to be cost effective and it is endorsed by

NCCN, but I don't think it's being done out there in the clinics. Or you can test everybody who is less than 70 years old and they meet the Bethesda guidelines, so that sensitivity is 95 percent and specificity is a little more than 95 percent, and that is also endorsed by the NCCN.

So, we're going to talk a little bit about Oncotype Dx colon. This is an assessment of seven cancer-related genes that correlate with recurrence and they represent genes associated with activated stroma, cell cycle and early response or genotypic stress.

There's a 12-gene recurrence score that was derived in stage II and III patients who participated in NSABP adjuvant studies, so this was a cooperative group and is seen as a more energetic cooperative group now.

The analytic validation of this test was published in 2010. The intended use was to assist in a decision for adjuvant therapy, particularly in stage II where we know that some cancers can be aggressive and most of them are not probably, but also stage III where

there might be some overtreatment of cancers that tend to be not so aggressive.

So published in 2014 was a study looking at both MMR and Oncotype Dx, and of 221 patients, 141 had T3 MMR proficient tumors and of these, looking at the Oncotype Dx, 71 percent turned out to be low risk, five percent were high risk, and 25 percent were MMR deficient, so you would tend probably, although not definitively, not to treat the MMR deficient patients, and ideally you'd want to make the decision not to do an adjuvant treatment for the low risk patients.

So they looked at how do clinicians receive this information and what do they do with it. 33 percent of the 45 percent who had changes in treatment plan decreased the intensity of their adjuvant treatment and 11 percent increased it. We don't know what that did, though, we don't know how patients fared with that. Chemo recommendations, however, decreased from 52 to 30 percent.

So, how I view this right now is prognostic, it's really not predictive, because we don't know really what treatment it would or

would not predict for giving. Interestingly, if you look at all stage II cancers, there's a little hint that adjuvant treatment might help, but the higher risk patients using the Oncotype Dx would have the same relative tiny benefit, but probably a higher absolute benefit, and you would use it in conjunction with the T stage already indicating some problems, and the MMR status, and the rest of the clinical information on the patient.

I'm going to switch over to the non-small cell lung cancer, EGFR, ALK and KRAS. EGFR and ALK is a Category 1 recommendation to be performed in all patients with metastatic lung cancer, as recommended by the NCCN. They recommend multiplex sequencing or FISH. It's clear that patients who have activated mutations of EGFR which, we have most evidence of deletion 19 or L858R, will be sensitive to the first generation EGFR inhibitors and likely the second and third generation EGFR inhibitors as well.

However, the T790M mutation is a resistance mutation and that predicts that the patients would not be sensitive to these EGFR

inhibitors except for the third generation ones, and T790M is the most common resistance mutation happening in patients who then progress after having a response to EGFR inhibitors with lung cancers.

Interestingly, T790M can be a germline mutation as well and this tends to be a lung cancer family, I'm not sure if it relates to other cancers at this point, but it's an area of active investigation.

ALK translocations are another area where, you know, ten years ago we just lumped all non-small cell lung cancers together, it didn't really matter. Now we have adenocarcinomas and squamous carcinomas, and all of these mutations I'm talking about are mostly in adenocarcinomas, very rarely in squamous carcinomas. But the ALK fusion or the ALK translocation is highly responsive to an ALK inhibitor and it does tend to present metastatic -- most ALK and EGFR tend to present in patients who are nonsmokers or less smokers, but they are not exclusively that.

So these tests, basically, I think are used mostly in a predictive sense. In the

prognostic sense, generally these markers mark cancers that are maybe not as aggressive as your garden variety non-small cell metastatic lung cancer. We don't really use them very much in a situation of resectable lung cancer.

But there's a new trial with the NCI and the National Clinical Trials Network called ALCHEMIST. This will take all comers with resectable lung cancer, they will get standard adjuvant therapy versus, if they have an EGFR activated mutation or an ALK mutation, will get also adjuvant treatment with the relevant drug for that situation. And as part of that, we may be able to find out a little bit more about the prognostic abilities of these in patients who have resectable tumors, as well as the prevalence.

So there are two approved companion diagnostics for EGFR, the therascreen and the cobas, so the analytical validity is taken care of, as well as the clinical validity, right there for metastatic disease.

And for ALK, co-approved with the anti-ALK drug crizotinib was the FDA-approved

VYSIS break-apart FISH assay for the selection

of patients, previously treated patients with non-small cell lung cancer for crizotinib treatment.

In KRAS I don't think, while there may be a little bit of prognostic evidence for KRAS, I don't think clinically it's used very much because we don't have much to do with KRAS as far as treating it. However, one thing that might be useful is that it is mutually exclusive with EGFR, so if you have a metastatic patient and you do have a KRAS mutation tested on them and it is a mutation, mutated KRAS, they probably don't have any EGFR activated mutations.

Let's turn a little bit to the breast cancer, invasive breast cancer right now, and MammaPrint and Oncotype Dx.

MammaPrint is an Agilent gene expression array using 70 genes. The intended use is to predict recurrence risks at five years in early stage breast cancer that has been treated with surgery and other treatments. It has been reviewed by EGAPP, it also has FDA

clearance for both fresh tissue as well as for paraffin-embedded tissue for women less than 61

years of age with stage I or II lymph node negative breast cancer.

The clinical utility of this is one of those requiring a randomized trial, and MINDACT is currently cooking, so we don't know that evidence yet but hopefully we will.

Oncotype Dx, a similar situation. As was stated previously, the analytic validation was published. The intended use is to predict the ten-year recurrence risk in early stage estrogen receptor positive breast cancer patients after surgery, and initially lymph node negative breast cancer patients.

Now recall, or know now that over the last, you know, several decades, more and more patients were being treated with adjuvant chemotherapy for breast cancer and it got all the way to the early stage disease, the lymph node negative, even ER positive breast cancer patients, and yet data were coming out that we weren't quite sure the chemotherapy was doing a whole lot for ER positive patients, maybe the

key treatment was the endocrine manipulation. So this was clinically validated in two retrospective prospective trials, and also

they're looking at it in ER positive patients, and we are currently waiting for TailoRx which is the lymph node negative, and Rxponder which is a lymph node positive, for the studies to complete.

EGAPP did not find evidence of clinical utility for either test, but again, we're waiting for results, and then that will be it for me.

DR. ROSS: Good morning. Thank you very much for the opportunity to address the committee this morning. I'm Jeff Ross, I am a molecular pathologist and I practice at the medical school and medical center in upstate New York in Albany.

I draw your attention to the disclosures. Unlike many of the other speakers, I have a wide range of disclosures to share with you, and in particular I ask you to focus on my relationship with Foundation Medicine, a company that I'm a co-strategic

founder of, the full-time medical director, and a shareholder. Foundation Medicine evolved in the era of predictive testing for precision medicine and placing cancer patients on

targeted therapies custom-designed for them.

Molecular pathology has evolved significantly in the last 40 years. We began with immunohistochemistry guiding breast cancer, in particular estrogen and progesterone receptor testing, and prognostic factors emerged both for solid tumors and for hematologic malignancies during the 1980s. The switch to messenger RNA-based expression profiling was particularly heralded by the two molecular RNA profiling tests, the Oncotype Rx and the MammaPrint, as prognostic tests but also with a therapy guidance component.

We expected messenger RNA expression profiling to really become the signature assay for cancer classification and therapy, but unfortunately that did not occur, and surprisingly in the late 1990s and then constantly now in this century, DNA sequencing became the major technique for getting patients

on therapies that are designed for them and based on matching the available new library of targeted therapies to the genomic alterations that these different cancers have.

The HER2 story is, I feel, an

excellent way to tell the background. This is an almost 40-year story of the discovery of the HER2 gene by Robert Weinberg in the early 1980s, to a day here in the Maryland area that was somewhat dark and dreary in December of 1998, a day in which my career changed, and I think the careers of everyone who practices oncology changed. It was December of '98 when one side of the street in Gaithersburg, the ODAC of the FDA approved the anti-HER2 antibody therapeutic trastuzumab, or Herceptin, and across the street in the in vitro diagnostics branch a test that was incorporated into the label of trastuzumab, the Dako HercepTest, an IHC test, was designed only for selecting patients as eligible for treatment with this drug in the metastatic setting.

As you follow this long arrow, appropriately in pink, but you will see no

longer having to stay pink, is the idea that multiple drugs have been developed, both more antibody therapeutics and the introduction of kinase inhibitors, all the drugs targeting this HER2-driven disease.

What we've also seen is how a target

can change, how a disease indication, breast cancer, could expand in 2008 now to gastroesophageal cancer, also showing amplification of the HER2 gene and responding to anti-HER2 targeted therapy.

And then most recently, we begin to see emerge the fact that the slide-based test, the immunohistochemistry and FISH that we used traditionally to detect HER2-driven cancers, are not sufficient to detect all of them, because many other tumor types are driven by mutations and the sequence of HER2, not by increase in the copy number or the protein. So we happened to see the fact that we needed to have tests that could evaluate all the classes of genomic alteration, and stop limiting them to where the tumor started, to find the drivers that would get patients on individualized

19 therapy.
20 The traditional tests, especially
21 immunohistochemistry, FISH and the so-called
22 Hot Spot DNA sequencing tests presumed that the
23 tester knows what alteration is likely to be
24 there and then goes and just tries to confirm.
25 But cancer is a complex disease and each tumor

1 type can contain any number of different
2 alterations, and if you don't cast a wide net
3 and use a test that will establish a complete
4 study of all of the alterations, you run the
5 risk that either you don't detect the
6 alteration you need to, or the one that you
7 want to detect you can't detect because you've
8 used up the sample doing each test
9 individually, and now have no more formalin
10 fixed paraffin-embedded material, and the
11 patient may not be well enough to undergo
12 another biopsy.
13 The Hot Spot panels of DNA sequencing
14 also presume to know what alterations and which
15 portions of the genes may be altered and then
16 they may only be looking for a substitution
17 when the alteration is something else like a

short insertion, a deletion, or even a gene fusion.

But our goal is to have disease outcomes, like shown in this slide and several others I'll show you. Here a patient with BRAF V600E mutated metastatic melanoma, who was one of the first patients to receive the targeted anti-melanoma BRAF drug vemurafenib, had

obtained this kind of advanced response to bone scan in just a two-week period.

But we have more examples of this type of thing happening in different ways. This is a breast cancer patient with widespread metastatic disease whose tumor was negative by FISH or IHC for HER2 gene amplification, but instead had a sequence alteration in HER2, actually two of them, one in the kinase domain and one in the extracellular domain, and was put on anti-HER2 targeted therapy in the absence of a positive, quote, test, because the gene DNA sequence showed that this HER2 gene was driving the tumor by hitting base substitution mutation, and achieved a dramatic response.

Here's another example of a patient with widespread HER2-driven disease who tested FISH negative and IHC negative, who had an extra cycle of domain HER2 mutation, placed on a combination of lapatinib and capecitabine, and had a dramatic response as the scan shows here. Now we're learning that what we thought was a target for a single disease and was a single alteration type now was a target for

many subtypes of solid tumors and also can be more than just an amplification, it can also be a base level substitution. We have to test for all the alterations and we have to do them in a sensitive way.

Here's an example of using the FISH technology in lung cancer and putting the patient on the wrong therapy. This individual presented with a brain metastasis from untreatable metastatic lung cancer. The FISH test, which you see in the upper left corner, shows the red dot and the green dot very close to each other; in other words, they're not broken apart in the FISH test and the EML4 ALK fusion was reported as negative. The patient

continued to progress on frontline standard of care chemotherapy and then the sample was sent for a sensitive hybrid capture base next generation sequencing assay, and indeed an EML4 ALK fusion was found. Indeed, the FISH test only identifies about 70 percent of the ALK-driven lung cancers and 30 percent of patients who have ALK-driven lung cancers do not go on crizotinib, the anti-ALK drug, because the test is not sensitive for all the

alterations that can be ALK drivers.

This is one of the issues that always comes into play when a gold standard test in the label of a drug after approval turns out not to be the gold standard. This patient, by the way, is four years now on crizotinib, and his fatal disease, fulminant disease is now converted to a chronic disease because the targeted therapy continues to inhibit the driver alteration and the patient lives on with the malignancy.

These types of technologies of looking for all the alterations can also discover new driver alterations such as the RET fusion

alteration in lung cancer, which rapidly led to the use of RET inhibitors approved for thyroid cancer, now placed into lung cancer patients, a different alteration, not an activating mutation in RET like in the thyroid cancers, instead a fusion of RET with another intron driving the disease, still responding to the drugs.

So, we've heard a lot of it this morning, I don't want to repeat the issues of analytic validation and clinical validation and

clinical utility. I'd just like to make a couple of personal comments here about doing this kind of testing, and I test hundreds of patients each day, I do gene sequencing analysis and report this now to more than 60,000 patients in the last three years, and false negatives in my own practice are far more important than false positives.

When tests are done right, false positives should be controlled and excluded, but not using a sensitive enough test, given the impurity of the samples we receive, we don't get pure tumor populations, we don't grow

these tumor cells in culture first and then sequence them, we get them mixed with all sorts of nonmalignant cells, inflammatory cells, stromal cells, epithelial cells, and benign epithelial cells mixed in. We have to use tests that can still detect the alteration even when the actual target has been diluted down to a very small fraction of the total DNA abnormalities in the sequence. So in this sense analytic validation, in my view, is extremely focused on eliminating false negatives.

In order to do that, again, you have to have all the test sites being done at the same time, you have to be able to test for the base substitutions, the short insertions and deletions, the copy number changes, both gains and losses, and especially the new gene fusions which are becoming very celebrated targets for the large emerging wave of targeted therapies being looked at in these clinical trials.

Analytic validation for this is very challenging because there is no built-in internal control. Human cells do not carry

cancer-associated genomic mutations and alterations that can be used to make them sensitive enough not to miss the target alteration. You don't have EGFR mutations or Exon 19 deletions, or KRAS mutations that are associated with normal cells, so since we only can be certain that we are sensitive enough, we have to create a surrogate control system. We need to run cell lines diluted down to very low mutant allele fractions along with the patients, so that we're able to make certain our assay is working each time we do it for each and every patient we're testing.

Enter the complex disease. As we know, every patient's cancer is unique, its complex genomics are the entire sequence of the human genome. But curiously, despite the ability for cancers to have a wide array of alterations in any particular tumor type, for any one individual patient the actual number of clinically relevant alterations is relatively small. Some patients have hypermutation, phenotypes, they have DNA instability, or they have been exposed to very serious

cancer-causing environmental stresses like excessive UV light or heavy smoking, we may see more alterations. But especially in young people, the number of alterations can be very small, but in the Medicare population the number of mutations may be greater because of the age of the patients and their exposure to mutagens and toxins.

So the number of critically relevant alterations in a single patient is relatively low, but it varies among thousands of so-called passenger alterations that we have to detect. And now that we can detect them, as we heard earlier, a disease that was pretty much small

cell undifferentiated versus non-small cell undifferentiated lung cancer today now looks like a widely, widely rainbow-like pie chart of different small percentage subtypes. And this is a common disease, so although a subtype may be small it's still involving thousands of Americans each year, and now we think of lung cancer as a disease that only by doing careful genomic analysis we can separate each patient into their own individual category.

And for example, an alteration like MET, which we link to amplification only, will emerge just in the next few weeks with multiple publications fighting the Exon 14 mutation as an exquisitely sensitive target to MET inhibitors, and this is something that's changing monthly and daily, and that's why the testing has to match its capability to put the patient on a therapy that's matched to the alteration.

I found this chart and I had attributed, because I thought it summarized clinical validation quite well. I think if you look in the lower right corner, that's what we're hoping for, a test that not only can be

done in a reliable and reproducible way, but it's also hitting the target which will predict clinical utility. As you can see, we have tests that are not only not valid for being reproducible each day, we have tests that are reproducible but they're actually off target. We can never hit a hundred percent for positive predictive value but we certainly want to try whenever possible for that.

The negative predictive value is more important because most if not all of the therapy selection tests are based on negative predictors. They're not telling us who is going to respond to the targeted therapy, they're telling us who is so unlikely to respond to it that we should consider a different treatment option.

The clinical utility also, this study that's widely cited talks about impact in health outcomes, strategies, the probabilistic nature of the test itself, and how it compares to other tests that are attempting to do the same thing. In daily practice we look for the most sensitive test with the least risk for false negatives that has the best chance to get

the patient to get the kind of responses to targeted therapy that I showed to you earlier.

In my own practice, which began with prognostic testing development -- and I developed and patented the first FISH test for HER2 testing in the early 1990s. It rapidly gave way when the targeted therapy for HER2 and the approval of trastuzumab made that test much

more important to patients for getting them on their own personalized therapy, not predicting the likelihood of whether the disease would relapse or not after primary treatment.

Another example of how this plays out is taking patients who are tested by one parameter, not placed on target-specific therapy, and then retesting them with a more sensitive test and seeing whether the new sensitive text not only will detect alterations that were missed initially, but whether those patients would then, now known to be matched to targeted therapy, that they will respond to the treatment. This was a recently published study which showed just that, where in a small subset of more than 30 patients who tested negatively for lung cancer were found to actually have

targeted therapy by a more sensitive test than the original screening test. Eight of the 11 of them have already shown dramatic responses to therapy that they would not have gotten, and they would have been placed on generic chemotherapy had not a reflex been done to test them on a more sensitive test.

In a disease like colorectal cancer, for example, we see this long tail or distribution chart of genomic frequency against patients with relapsed metastatic disease, that there are genes that are widely overexpressed repetitively that we have no targeted therapy for. And on the left of this chart you have alterations like APC, KRAS and PT53, where we don't have a targeted therapy for them. But what we're interested in is matching all of the other patients who also have alterations that do match the targeted therapies, and one of which actually is BRAF. Despite the initial report that BRAF, V600E in colon cancer is just a very serious negative predictive factor, we are going to see second and third generation BRAF inhibitors that are more potent and when combined with the right additional

chemotherapy, we can target BRAF in colon cancer after all.

But many others, the little ones, the ones that are one percent, two percent, one percent, two percent, all represent hundreds if not thousands of patients, because the disease

is so common. This gives us a chance to get each patient matched to a therapy, rather than think of them as just a large group that goes on and will respond to a full series, when indeed it relapses in the liver.

So, for the questions, I gave my own responses. It all depends.

For the question one about the analytic utility, some tests are prognostic only, some are both prognostic and predictive. I'm always going to favor the predictive test because today it isn't so much whether your cancer has recurred, it has recurred, it's how can we get you on an individualized therapy. And then the issue of no built-in control, we need to run the cell lines and parallels so we don't miss anything, and thus prevent ourselves from getting false negative results.

For the second question, I think it

also depends when we talk about assessing health outcomes. Predictive tests that find alterations that lead to effective therapies obviously is all our goals. But the greatest outcome benefit is to be looking for all the

alterations, not consuming the sample with individualized one-by-one-by-one tests and then denying the patient the therapy that could so greatly help them.

So, I thought it might be worth going over a certain story line that also reflects how drugs are being developed for cancer now. In the pre-targeted therapy era it was very difficult to recruit patients to clinical trials, the response rates were so low, the benefit in terms of overall and disease-free survival were minimal for lots of different chemotherapy regimens and cancer types. But now we have Richard Pazdur, the medical director of the ODAC of the FDA, giving an interview in which he says conventional therapy might get response rates of only 10 to 20 percent but he repeats, newer drugs are being developed that have response rates of 50 to 60 percent. And he says, would it make sense to

do a randomized trial in this setting? And if there was one, he asks, what patient would want to go on a randomized trial when the treatment arm would have a 50 to 60 percent response

rate, when you're having a big effect is kind of jaw dropping. He says there are response rates we haven't seen before in cancer.

So what this means is we aren't going to be having multi-arm trials anymore, we're going to continue to see single-arm trials where we have a test to define a patient's eligibility, we have a drug that has a high response rate, and thus early in development, maybe Phase I where we just prove safety, and then show a 60 to 70 percent response rate in less than a hundred patients, will allow approval of the agent, with follow-on clinical trials that do add a comparator arm. And when we've done that recently, especially in lung cancer, we've seen those follow-on trials today absolutely validate the early approval of the drug and show dramatic overall response rates, survival rates, and low P values when that data ultimately becomes available.

So on question three about the

prognosis and clinical utility, no question that diseases like breast and lung cancer were improving outcomes by doing this approach.

Colon cancer is lagging behind. We do have the KRAS test as a predictive test in order to deny a patient opportunity for the anti-EGRF antibodies, but here we should see that evolution.

Here we heard about the KRAS, the testing on codon two, but that's no longer in the NCCN guidelines because now we have to test the entire coding sequence of KRAS because other mutations initially not known to be negative predictive factors when you use anti-EGFR antibodies, we now know that even the NRAS gene has to be tested, so it's a change in clinical practice that happened, best within an overarching test testing for all of these things right in the beginning.

But we're going to see changes, we don't have a targeted therapy for RAS in colon but we have targeted therapies for a whole host of small subsets of colorectal cancer patients just sending in manuscripts on ALK fusions, the tests we do in lung cancer responding to ALK

inhibitors when those fusions are seen in colon cancer, and it's a very small subset of

patients. And we've seen MET amplifications respond, KIT mutations respond. Slowly we'll pick at colon cancer like we've already shown we can do with lung cancer, and maybe help that one catch up in terms of getting a significant population on the targeted therapies.

For leukemia, lymphomas, and the tumors like the gastrointestinal stromal tumors, and obviously melanoma, we're already showing the dramatic benefits by first sequencing, getting the target, and then matching it to the patient.

So in these last two slides, here are my recommendations as an outside, I'll call it an opinion.

For colorectal cancer, BRAF testing I give a green light because I think it's going to be a therapy target and a predictive test, not because I want to just know that colon cancer is more likely to relapse in the liver than one that's BRAF wild-type.

KRAS has to be considered a high value test because it's guiding the use of two

important anti-EGRF antibody therapeutics on

the market. For the prognostic use of testing to predict that a patient will relapse with colon cancer, because of my own personal assessment of the adjuvant therapy data, I don't see the value as highly because I can't see that the decision to treat is having such a dramatic impact on patient outcome.

Moving to breast cancer, I think that the tests that guide withholding of chemotherapy, the MammaPrint and the Oncotype certainly have value because this is allowing patients to avoid the toxicity of chemotherapy, but we always have to keep in mind that the value of those tests is driven by a patient willing to take the antiestrogen receptor targeted therapy, tamoxifen, for five complete years, and that is not so easy to achieve in a lot of women who have the side effects that lead them to decide to stop therapy.

Finally on the lung cancer, there's no question that ALK testing and EGFR testing are standards of care, but the tests that are actually approved are only really sensitive enough when the sample is heavily enriched with

malignant cells. For ALK, the FISH test, I've already mentioned in my opinion that it's only 70 to 75 percent sensitive and that's going to leave a large percentage of patients getting on chemotherapy that should be on crizotinib because the FISH test missed their ALK fusion. And for EGFR, the samples have less than 50 percent tumor cell purity. When the benign cells contain the tumor cells, the sensitivity of both EGFR sequencing tests in the labels of the two approved drugs lose their sensitivity and run the risk of causing a false negative result in a patient who doesn't go on EGFR targeted therapy.

In my own practice I've seen, you know, more than a half dozen in the last six months, patients retested for EGFR who are called negative by the standard of care test, who were found positive by the more sensitive next generation sequencing test, who went on vemurafenib, and all got dramatic benefit from the therapy. So we must not let patients miss their targeted therapy because the test that is used is not sensitive enough in their sample to detect the alteration.

And finally for KRAS on lung cancer, I did want to say that it's not a hundred percent that KRAS mutation precludes a targeted therapy opportunity. We have seen in one to two percent of KRAS-mutated lung cancers an EGFR, an ALK fusion, and several other targeted therapies. Almost a hundred percent, but really only in the high 90s.

DR. HENDERSON: You're running out of time.

DR. ROSS: This is my last slide.

DR. HENDERSON: Okay.

DR. ROSS: The last slide just says that this is what I would like to see this committee be discussing, which is all of the new emerging targeted therapies that are rapidly coming forward and giving patients a chance to get on to individualized treatment for cancer. Thank you.

DR. HENDERSON: Thank you for our speakers, we appreciate it. So now we're going to have a five-minute break, so it is, I believe we will start again at 10:03.

(Recess.)

MS. ELLIS: Good morning again,

everyone. At this time we are about to start the scheduled public comments portion of the meeting. I just wanted to make a statement to let everyone know that Dr. Sam Caughron, he is unable to attend today's meeting due to circumstances beyond our control, but you guys do -- I'm sorry -- the presentation was submitted and it has been posted to our coverage website for a week, the panel members, they did receive Dr. Caughron's presentation about three weeks ago, so we do have his presentation, his presentation is made available for the record, it just will not be presented at today's meeting. Should the panel have any questions in regards to Dr. Caughron's presentation, Dr. Jan Nowak, who worked with him on the presentation, will be available for comments during the comments to presenters portion of the meeting. Okay? You may begin.

DR. NOWAK: Good morning. I am Jan Nowak, from NorthShore University HealthSystem in Evanston, Illinois. I direct the molecular diagnostic laboratory there, so I actually perform many of these tests that we're talking about, I take the results to our cancer tumor

boards where the information is discussed with the oncologists there and is used to make treatment decisions, so that's what I do.

So today, I don't know if this is an important disclosure, but I'm salaried by NorthShore University HealthSystem, occasionally I help other institutions out, and sometimes they pay me.

So, I am here as a representative of the College of American Pathologists, and you've already heard a little bit about what the college does. So, all clinical laboratories in the U.S. must be certified by CLIA. CAP is deemed by CLIA to accredit and inspect clinical laboratories. In order to assist laboratories in satisfying CLIA and CAP standards, CAP also provides a robust and widespread proficiency testing program encompassing hundreds if not thousands of clinically relevant analytes. Most of the 7,600 laboratories in the U.S. -- are in the U.S., and these include laboratories in academic medical centers, community teaching hospitals, community hospitals, small private

25 laboratories, and even large reference

1 laboratories.
2 So, you have before you three issues
3 for discussion regarding laboratory testing
4 that we think we can assist you with.
5 Specifically, does FDA approval status make a
6 difference in testing for the analytics in
7 question; is there a difference in where the
8 testing is performed; and finally, are these
9 tests analytically valid?
10 Now this question, as I think other
11 people have already stated, is not clearly
12 framed. Analytical validity usually refers to
13 the ability of a laboratory method to detect
14 the specific biochemical analyte or its
15 variants. The ability of the test result to
16 inform about the presence or absence of a
17 disease state is appropriately called clinical
18 validity. My focus will be on the former.
19 Others, including Dr. Caughron's presentation,
20 focuses on the latter.
21 So, CAP provides proficiency testing
22 for all of these analytes and for ALK as well
23 through its cytogenetic surveys. So as you can

see, these programs have been in place for a number of years, some as long as ten years, and

the number of participating laboratories has grown. Now, not all laboratories who do this testing participate in this program, but I think most of the laboratories in large academic centers and reference labs, most laboratories who do this testing are represented here.

So, some summary statements on some of these surveys, but typically a survey distributes a number of blinded specimens for evaluation at least twice a year. These challenges are often designed to address specific issues of clinical relevance such as detection limit, confounding analytical factors, and are often supplemented by interpretive challenges. The surveys also collect information about methodologies used in order to discern any patterns of inadequacy among participating laboratories. This data is regularly reviewed, summarized and returned to the participating laboratories for their review and use.

Now it's important to recognize the consequence of a PT failure. An isolated instance of PT failure results in instructions

to the laboratory to review all aspects of the test procedure to understand the source of that failure, and to take appropriate corrective action and verify that the test is performing adequately. A second failure within a year results in an order to cease testing for that analyte and the failure is reported to CMS. The laboratory must then demonstrate good performance on at least two subsequent PT challenges before it is allowed to resume clinical testing. So this is a big deal if you don't pass this proficiency test.

So here's some information on the BRAF survey that you can read.

So for the laboratories, this is part of their ongoing quality assurance program, but the information is also useful in designing and implementing improvements in their specific tests and protocols. So as these tests evolve, as the proficiency testing goes forward, we can see that the performance of laboratories

actually improves as we go along, and that's the whole point of quality improvement. The long-term performance in all of these surveys has been truly consistently good.

Some information on the EGFR survey that was initiated in 2010, there's 213 laboratories now participating in this. Again, the performance has been very good.

The mismatch repair and the MSI survey has been in place for ten years now and, again, the performance is very good.

Now CAP attempts to do summaries after a number of years of the performance on these surveys, and they do publish this information, so that kind of summary was made on this survey just a couple years ago.

So in this slide you can see the most recent performance for laboratories reporting for the anticipated result for all challenges, so this is the end of 2014 for which the results have been reviewed, and you can see that for all of these surveys the performance has been really outstanding.

Now, some of the information that

these surveys collect. So this slide summarizes the wide variety of methodologies and test platforms laboratories use for these tests, so KRAS, BRAF, EGFR and MSI across the top. The highlighted category includes but is

not exclusive for an FDA-approved test, so you can see that for most of these tests, the majority of tests are not performed using an FDA-approved version, these are all laboratory-developed tests and laboratories use a wide variety of methodologies to get the answer. Yet again, you know, I emphasize that the performance by all of these laboratories has been uniformly outstanding.

So with the accumulated years of CAP proficiency testing for these analytes, CAP has not perceived any difference in performance that distinguishes FDA-approved tests from laboratory-developed tests, not for any of these analytes.

I'll also point out that all of these tests are only available as LDTs until the vendor thinks it profitable to develop a test and submit it for FDA approval. The MSI survey

has been around for ten years, on the far right, and there's still no FDA-approved assay for that, so these things don't start with someone bringing forward an FDA-approved test, these assays start because there's clinical need, and laboratories develop this test and

they take responsibility for doing it well.

So in answer to the questions before you, does regulatory status of the test, whether it's FDA-approved or laboratory-developed, make a difference, well, the CAP PT survey data shows that regulatory status is not a determinant of performance for the analytes in question. Test performance has been uniformly good.

Does the type of laboratory make a difference, whether it be an academic laboratory, a reference laboratory or a community hospital laboratory, and again, the CAP proficiency test survey data has not discerned any difference in test performance for these analytes based on the type of performing laboratory. Test performance has been uniformly good and consistent.

And then for each prognostic test listed, how confident are you that there's sufficient evidence to say that these things are analytically valid? The CAP survey data, in conjunction with the other requirements of CAP laboratory accreditation take into account both analytical validity and clinical validity,

and those assure us that the analytical validity of the testing for these tests is appropriate for clinical use.

Now, I just want to add that among the standards for CAP accreditation, the laboratory, the medical director of the laboratory is responsible for assuring that the test is clinically valid and that it's used appropriately clinically, so that's part and parcel of what this accreditation program does. 200-plus laboratories across the country would not be doing these tests if they did not have good clinical validity and clinical utility, and we wouldn't be offering the proficiency tests for these things if they did not. Thank you.

DR. HENDERSON: Okay. Next. Try to

keep your remarks to five minutes, please.

MR. VAN DER BAAN: Good morning. My name is Bas van der Baan, I am VP of clinical affairs at Agendia, I am an employee of Agendia.

So most guidelines in the U.S., like NCCN and ASCO, recommend to consider chemotherapy in early stage breast cancer with

tumors larger than one centimeter. As a result, the majority of early stage breast cancer patients receive just-in-case adjuvant chemotherapy. These guidelines tend to ignore the fact that for a decade there have been additive improvements in survival due to screening, education and better therapies.

The largest meta-analysis for benefit of chemotherapy in Stage I and II breast cancers demonstrates only a four to six percent absolute benefit of chemotherapy. Yet due to reimbursement incentives for chemotherapy and inadequate prognostic tools, more than 60 percent of women with endocrinic cancerous tumors continue to receive chemotherapy. There is a significant need for additional prognostic

information to more accurately decide on the risks and benefits of chemotherapy in early stage breast cancer.

MammaPrint was developed in 2002 and validated over the past 13 years in thousands of patients. The central clinical question which MammaPrint was designed to answer is to find the largest group of low risk patients who could safely forgo chemotherapy without

compromise to outcome. By interrogating the full genome, we selected the 70 most prognostic genes to predict outcome in breast cancer, and this provides substantial additional independent information beyond traditional clinical factors like ERP, HER2, size and grade.

Analytical and clinical validity has been extensively established in numerous FDA clearances of MammaPrint, published in peer reviewed journals. This is a summary of the analytical and clinical validity performance that we submitted to the FDA, and what a surprise to see that in the overview, FDA data is not used. I think we can safely conclude

with the six FDA clearances that we have, that clinical and analytical validity has been robustly established for MammaPrint.

So, we've also established clinical and analytical validity in three analytes. It's the only commercial assay that can be used both fresh, fresh frozen, and formalin-fixed paraffin-embedded tissue. MammaPrint's recognized clinical validation was the basis of its selection as a qualifying biomarker for

patient enrollment into the unique adaptive trial design of iSPY 2, looking at innovative drugs. Investigators at 24 of the nation's most prestigious NCI-designated cancer centers rely on MammaPrint to accurately identify low risk patients who can avoid chemotherapy, while high risk patients are enrolled in this innovative trial.

MammaPrint has been validated to outperform clinical risk assessments predicting outcome of early stage breast cancer at five, ten and 25 years. In the 30 percent of cases where MammaPrint results and clinical risk assessments are discordant, MammaPrint has been

demonstrated in prospective outcome trials with five-year followup to be more accurate in predicting disease outcome. MammaPrint is included in many global guidelines for prognosis, like Argos, ASCO and NCCN, and last week was again confirmed at the Sangalli Convention meeting to be accurate in risk stratification for breast. MammaPrint's low risk patients can safely forgo chemotherapy and MammaPrint high risk is an indicator for chemotherapy.

MammaPrint identifies between 60 and 65 percent of all early stage breast cancers to be low risk and those can safely be spared chemotherapy, as I will show in our prospective outcome data.

This is our RASTER study, started in 2004. RASTER examined the health outcomes of patients five years following treatment decisions that incorporated the MammaPrint result along with other clinical pathological factors in a real life clinical setting. When physicians and patients followed the advice of the MammaPrint low risk result despite

important clinical factors that indicate the patient had a high clinical risk, they had a 97 percent five-year distant recurrence-free survival. Patients who did not receive any adjuvant systemic therapy had a hundred percent distant recurrence-free survival. So there is outcome data available on those tests.

Earlier also, the MINDACT trial, the prospective randomized trial was mentioned. MammaPrint when used according to its intended use identifies the largest percentage of early stage breast cancer patients who can safely

forgo chemo of any commercially available multigene signature. MammaPrint lowers patients' established clinical utility by showing the same outcome with less harm, reducing unnecessary toxicity and long-term effects of chemo, while at the same time being cost effective by reducing chemotherapy costs.

The MINDACT trial has enrolled almost 6,700 patients and is a prospective randomized trial that will yield level 1.A evidence. The trial closed recruitment in 2011 so we expect the outcome data soon. Patients with

discordant risk between clinical risk assessment and MammaPrint were randomized and treated either according to clinical risk or to genomic risk. There was a significant reduction in chemotherapy in the MammaPrint arm of the trial, 1,550 patients were clinically high risk and MammaPrint low risk in this trial.

In the interest of time I'm going to skip the discussion slides, but briefly mention that MammaPrint --

DR. HENDERSON: Your time is up, so if you could wrap up?

MR. VAN DER BAAN: Sorry?

DR. HENDERSON: Your time is up.

MR. VAN DER BAAN: Okay. So concluding, analytical and clinical validity as well as clinical utility has been established for MammaPrint, and thanks for the time.

DR. HENDERSON: Introduce yourself and the conflict of interests, please.

DR. BILKOVSKI: Good morning, everyone. My name is Dr. Robert Bilkovski and I head up medical affairs for Abbott Molecular.

My conflict of interest is that I have a financial benefit with Abbott Molecular, I'm a full-time employee and receive compensation for that.

I will have a rather truncated or limited line of view for Abbott Molecular System's products with regard to the view of the analysis that was applied for this MEDCAC assessment with regard to the difference between predictive and prognostic. I have one slide but there will be a fair amount of voiceover and I would be surprised if that was going to take more than five minutes.

As has been just elaborated earlier

today with Dr. Ross and Dr. Conley, there is a difference between a test or an adjunct that is predictive versus prognostic. There is an overlap between the two; however, in our situation with regard the ALK, it has been utilized and attained FDA clearance as a containing diagnostic, for its ability to be able to be predictive to identify patients that have a specific type of non-small cell carcinoma that has a favorable response to

crizotinib.

So as part and parcel of the FDA's containing diagnostics mantra, a containing diagnostic links a diagnostic to a therapeutic for the purposes of one of three things: whether or not it selects a patient that is more amenable to therapeutic benefit, it deselects a patient that is less favorable to receive therapeutic benefit or it selects them away from an adverse event, and then thirdly, it can be helpful to monitor a patient in terms of response to therapy.

The break-apart FISH assay that we have secured FDA clearance for is a companion diagnostic with crizotinib that has been linked

to a favorable outcome response with the use of the Pfizer molecule crizotinib. With regard to the assessment that is delineated below, I wanted to highlight a couple of issues that we came away with when we looked at this.

First of all with regard to the clinical benefit, clearly from our standpoint there is a clinical benefit if we can be able to predict a patient population that has a more

favorable benefit to a therapeutic intervention, that being crizotinib. That was not included in terms of the assessment that was performed for the MEDCAC review.

With regard to efficacy in terms of patient care, we also have been able to demonstrate that in terms of clinical benefit. However, the ability to predict a patient that has a favorable response in and of itself has a prognostic benefit. However, the prognostic question is not the scope of what ALK was intended to do, the intended use was to be predictive of the presence or absence of a particular mutation that predicts a favorable response to crizotinib.

And then last, we covered clinical

benefits where the NCCN has established guidelines that recommend the utilization of ALK to select patients for crizotinib therapy, once again favoring the clinical benefit of the use of this test.

So in conclusion, I'm just trying to share the same feelings and thoughts that were shared by Dr. Ross and Dr. Conley earlier, that

the ALK assay has a clinical benefit in terms of being able to be predictive, but it is not within the intended use or in terms of the package insert or the labeling with regards to what we have to operate under the FDA that we are making a prognostic claim, so that is not the lens of which our assay was intended.

Based on that, I thank you for your attention, and I give you a minute back.

DR. HENDERSON: Thank you. I appreciate your timeliness.

So, that ends the scheduled public comments. I don't think we have any others, and we have no requests for open public comments, so we will go on now to the questions to the presenters. So if the presenters could take their place up here in the front, that

will be helpful.

So, one person asks a question to a presenter. Once that presenter comes up to the microphone then other members of the panel should, if possible, address that presenter at that time, so they don't have to keep popping up and down quite so frequently. And, let's

see, I think that's about the only major limitation here, so we're focusing predominantly on questions to speakers related to their presentations, particularly questions that will be helpful to us in making our decisions this afternoon in terms of how we're going to proceed with the voting. So who would like to initiate the first question here? Maren.

DR. SCHEUNER: Dr. Meleth, am I pronouncing your name correctly?

DR. MELETH: Yes.

DR. SCHEUNER: I have a couple questions for you. I just -- this should be hopefully straightforward. I notice that the tech assessment, the date there was May of 2014, and I understand there was a delay in gathering us together, and the literature was

reviewed through November 2013?

DR. MELETH: Yes.

DR. SCHEUNER: Have you looked for any updates in literature since?

DR. MELETH: No, ma'am, the contract ended in February or March of 2014, so that's

-- it was extended and we got comments back until the contract ended, and that was the end of the review.

DR. SCHEUNER: Okay. So I guess I'm just going to ask my panelists, you know, I'm not familiar with literature since that point, but I think that might be something that would be a big question in my mind, are we really looking at the most current evidence?

The other was, I don't think I saw, and maybe you can direct me in the tech assessment, regarding ALK and MLH1 promoter.

DR. MELETH: We didn't find any literature that would meet our inclusion and exclusion criteria to include studies for those two tests.

DR. SCHEUNER: Okay. And then I was wondering if you had considered looking at tests in combination, like MSI and BRAF.

DR. MELETH: So, if studies met our inclusion criteria, we didn't look specifically at whether combinations of tests were, there was a difference in the prognostic validity in terms of recurrence, but there were several

studies that had, you know, that were looking at several biomarkers in a patient population.

I have -- I'm testing my memory here. I don't think there were many that had models that included multiple markers. They generally tended to, although they were testing on the same patients, they tend to be looking at individual prognostic analytic events, so I don't think a lot of them were looking at combinations.

DR. SCHEUNER: Okay. And then I guess my last question is with regard to the chain of evidence and thinking about this issue of prognostic, is it a requirement in your understanding of thinking about policy development, that we need to see evidence that there's prognostic or clinical validity, evidence of prognostic value before one might expect that there's any clinical utility?

DR. MELETH: I'm not sure that I

understand your question fully. So, the chain of evidence is trying to build something when there isn't enough evidence to answer the overarching question, so that, do you need, I

think your question is, do you need clinical validity before you can assess clinical utility? Is that your question?

DR. SCHEUNER: Yes.

DR. MELETH: I would say yes, because if you're not, if the test is not telling you whether it's predictive of what you're interested in, then you can't go on to say whether testing that then helps the patient outcome or not, so yes.

DR. SCHEUNER: But because we didn't look at clinical validity for the outcome of treatment response to therapy, we can't really address that clinical utility.

DR. MELETH: You can't, not from our technical assessment. We specifically excluded those types of studies, yes, and this is not just our decision, it was a decision that we came to with CMS and AHRQ.

DR. SCHEUNER: Thank you.

DR. HENDERSON: Dr. Burke.

DR. BURKE: So nice to meet you again. I have to disclose, we've met in the past. So, in the request from CMS, you talked about

sensitivity, specificity, positive and negative value for the clinical validity. You provided hazard ratios.

DR. MELETH: Yes.

DR. BURKE: We'll talk later on about the meaning of hazard ratios, but on your summary, your executive summary, you had a parentheses, discrimination and calibration, which is a measure of predictive accuracy, and what CMS asked for were measures of predictive accuracy. Did you provide in your tech report anywhere any measures of predictive accuracy?

DR. MELETH: No, and part of the reason is that there isn't a whole lot of literature there. So, we are only assessing published literature, we're not designing the studies, we are limited to what is available to us to assess, and what was available to assess was just studies that looked at hazard ratios.

DR. BURKE: Thank you.

DR. RAMAMURTHY: Hi, Dr. Meleth, while you're still standing up there, in terms of

your TA, most of the tests that you reviewed for analytical validation were FDA-cleared

and/or approved tests. You do say, as the public commenting to the peer report, asks that you include the CAP-approved criteria as well to evaluate clinical, I'm sorry, analytic validity for test performance. Did you actually compare the FDA clearance approval standards on analytic validity to CAP's standards for determining analytic validity?

DR. MELETH: No.

DR. RAMAMURTHY: So it's not clear they're really comparable, then, but in the sense that 70 percent of the testing by the other presenter that's done are mostly, essentially lab-developed tests, should not constitute FDA clearance approval.

DR. MELETH: Okay.

DR. RAMAMURTHY: I have one question for Dr. Jeff Ross; can I do that?

DR. HENDERSON: Why don't we finish with Dr. Meleth.

DR. RAMAMURTHY: Okay, thank you.

DR. HENDERSON: Dr. Grant.

DR. GRANT: First, just a comment on

your reply about the accuracy issue, and I just

concur that those data aren't specifically reported hazard ratios, but actually invariably CAP was provided with a little bit of work, actually a lot of work.

A couple questions about the TA, though, that you did, and I'm not sure they're critical, but nevertheless they struck me. One was simply excluding any study that was deemed to have high risk of bias or unclear. You know, typical practice would be to draw a little, a wider net, and perform sensitivity analyses as opposed to excluding them a priori. Could you comment on that?

DR. MELETH: Yes. We had, we did think of doing that, and I don't know -- so, I just want to say that it ended up being out of the scope of the report.

DR. GRANT: And the other question relates to reporting bias and its companion, you know, publication bias. So there was one instance where a nonsignificant result, of course they weren't pooled, but did you otherwise assess potential for publication bias?

DR. MELETH: I have to say no. We did look for unpublished studies, we asked all the, you know, the companies that were performing the tests to provide data that they thought we should be reviewing, and we also looked at publications that we got through the public review. We actually got, had to review almost a hundred articles from things that we got through peer review and public comment, and none of those actually changed what, the overall results, and they were included when they were reviewed.

DR. GRANT: Part of the reason it struck me was that in the strength of evidence ratings, specifically, reporting bias is included in that, and I just, as you sat there I just didn't see that, so thank you.

DR. MELETH: Okay.

DR. HENDERSON: Dr. Scheuner.

DR. SCHEUNER: I'm sorry, I have one more question.

How did you approach Lynch syndrome with respect to the MSI testing?

DR. MELETH: We didn't include -- we looked at only sporadic colorectal cancer, we

didn't include Lynch syndrome.

DR. SCHEUNER: Why was that?

DR. MELETH: Because, I don't -- this is a long time ago, but I want to say that the idea was that Lynch syndrome and the predictors of heterogeneic diseases were not appropriate for this technical assessment, we were looking at tests that -- we just felt it was a different question that we would be answering if we looked at the predictors of recurrence, and so we didn't want to mix things up with sporadic diseases.

DR. HENDERSON: Okay. Dr. Zuckerman.

DR. ZUCKERMAN: I have a question for Dr. Meleth and Dr. Conley. There was an inconsistency in, I guess, Dr. Meleth, in your summary it showed that for the EGFR test, it seemed to have no predictive value, or no prognostic value, and Dr. Conley, you specifically said it had predictive and prognostic value, and I was not sure if that's because yours is a more up-to-date literature review or if there's other differences between the two.

DR. MELETH: Just what we found in the

literature that met our inclusion and exclusion criteria, that was not looking at the impact of the EGFR testing on therapy, we found no evidence of prognostic, added prognostic benefit of doing the test with the traditional markers that we used. That's what I can say about what we did, and I'm sure Dr. Conley will.

DR. HENDERSON: Any more questions? I would like to ask you one question.

DR. MELETH: Sure.

DR. HENDERSON: So, I'm trying to get a general point but I'm going to take one particular slide, it's slide number 28, and I'll remind you that this is a slide on KRAS mutation testing where you're comparing wild-type versus mutation.

DR. MELETH: Uh-huh.

DR. HENDERSON: And you had three hazard ratios, first of all for risk of recurrence and your hazard ration was 1.02, and of course not significant. Then you had cancer-specific survival and your hazard ratio of 1.30, which is probably even marginal. And then overall survival, your hazard ratio of

1.22, which was not significant or just maybe just slightly short of significant.

The question I'm asking here and trying to get at is, with the studies of overall survival in this particular case, were you looking at, did all those patients have treatment? In other words, you're comparing wild-type versus mutation when you're looking at treated patients, or did you even have any untreated patients where you could draw some conclusion about the relationship between KRAS and survival outcomes?

DR. MELETH: I don't remember, so if you need an answer to that, I will have to -- I don't want to give you an answer that I don't -- I would like to go back and look at the studies. I want to say that there was probably a mixture of the two, but I'm not -- we didn't have a subset of studies that looked just at treated patients and KRAS versus untreated patients and KRAS, but to give you a definitive answer I would have to go back and look at the pubs again.

DR. HENDERSON: The reason I'm pushing

this, and maybe we'll get back to this on

discussion later, this is kind of a thorny problem when we get to this issue of predictive and prognostic factors. And if we go back, like some of the things that are well established that are kind of models for them, if we look at HER2, and that was given, I think Dr. Ross gave a very nice summary of that, and I think he did sort of a model for most of us in terms of looking at genomics now, but you know, we had an abundant database of HER2 and outcomes, both survival and recurrence outcomes, before we had any anti-HER2 therapy, which is kind of nice.

And in actuality, we can do that for colon therapy too so you can see, look in the database of patients who never got any hormone therapy where we have HER or ER data is relatively small, but it does exist. And I would think that, so therefore you could say that both of these are prognostic as well as predictive, and you could look at them independently. And it sounds like you did not make an attempt to do that, even look at that

issue with any of these markers. Is that a fair statement?

DR. MELETH: I'd actually honestly have to go back and look at it. I don't remember right now.

DR. HENDERSON: But is it fair for us to assume that the information you provided for us is going to be mixed, in other words, treated and untreated, there's no attempt to sort out this question the way we have with ER and with HER2?

DR. MELETH: Yeah, so what we did do is when we were looking at the quality of the studies, we did have a specific question about whether there was information about treatment in the study, and whether they controlled for treatment while they were looking at the prognostic value of the marker, okay? So yes, we did look at the impact of treatment on the -- so we did look at patients in the cases where there was treatment and there was information in the publication about the treatment, we evaluated whether the models were controlling for treatment or not, but it will

be fair to say that it was a mixture, I think.

DR. HENDERSON: Okay. So we can assume that for all of your hazard ratios, by

and large it's a mixture, presumably it's more of a mixture for treatment with overall survival than it is with risk of recurrence, and oftentimes these patients are treated only after they've recurred; is that fair?

DR. MELETH: Yes.

DR. HENDERSON: Any other questions of Dr. Meleth? Thank you.

Now, next question, who had a question, was it Dr. Zuckerman or Dr. Ramamurthy? Go ahead.

DR. RAMAMURTHY: Mine was for Dr. Ross actually.

DR. HENDERSON: Okay, so Dr. Ross will be next.

MS. ELLIS: I'm sorry, could you please state your name for the record?

DR. ROSS: Jeffrey Ross.

DR. RAMAMURTHY: Dr. Ross, that was a very nice presentation, thank you very much.

You also noticed us all rapidly

writing notes and copying your answers. You also have successfully solved our issue of predictive and prognostic by just conflating the two and solving that problem for us. But

by saying essentially that knowing the biomarker status can give information about the outcome here, your supposition is that the right therapy is reaching the patient and the therapy does what it should do as in give a positive outcome, correct? Which isn't as applicable, number one.

And when you also say five and four for some of those things, the questions are for various biomarkers, but the biomarkers have so far tested by various tests. For example, KRAS, as an earlier presenter pointed out, there are just a variety of technologies and often regular laboratories that offer the test, number one. Number two, also some of these genes like KRAS, we know there are seven mutations that are most validated. Some laboratories in the past, for sure, I don't know currently if they're updated on this, have offered a lot more than those seven mutations.

So when you suggest in your humble opinion those answers of fives and fours, can you give a little more clarity on how you were so confident on those answers?

DR. ROSS: Sure, I would be delighted

to, thank you for asking the question. I'd like to respond also to Dr. Henderson's question. When we think about this issue of doing a retrospective published literature search of a prognostic factor and then look at all of the outcomes of those reports, one thing we often don't do is actually evaluate the technique that was done. I did this in 1998 on 3,000 patients with HER2, it was still at that time a debate, was it a prognostic factor or not, prior to the development of trastuzumab, and I thought since they were almost a hundred percent except for our one FISH study, immunohistochemistry-based, I did a filter of just saying gosh, we agreed that more than 10 percent of breast cancer is HER2 positive, and gosh, I'll give a high end of 40 percent, which I knew at that time already was way above what the real rate was, and just exclude the studies

that had less than 10 percent positive or greater than 40 percent positive, which eliminated about one-third of all of the published reports.

And then the prognostic segments of the test went dramatically into an absolutely

predictive bad outcome, and then we said what could be wrong with those tests that didn't seem to get the right result, and I found out things like package inserts for the antibody used was specifically for frozen section material only, and this was 670 paraffin-embedded patients who were studied. Generally the authors don't look at the package insert of the actual test that was done, they look at the results of the test. These are the kinds of variables that can affect when you're trying to evaluate retrospectively prognostic tests.

So why did I say fours and fives? So for example, for tests where I said four rather than five as a prognostic test is because it was really being used like MammaPrint and Oncotype, much more than just a prognostic

test, people are making therapy decisions based on it, and the therapy decisions aren't always easy to assess because a lot of patients with low recurrence risks still want to take chemotherapy because they're so frightened of the disease, and that one or two or three percent will benefit enough for them that they

will ignore the fact, they're happy in the low risk group but they're still going to take treatment. That confounds the utility.

And then the other side of it is, with the utility is, for those two breast cancer tests, you have to stay on tamoxifen for five years for the test's actual predictive value to be valid, and a lot of patients aren't able to do that, so that's why I dropped them from five to four, but I think they're still very useful in the daily management of newly diagnosed breast cancer.

DR. HENDERSON: Dr. Zuckerman, did you have a question of Dr. Ross?

DR. ZUCKERMAN: No, mine was for Dr. Conley.

DR. HENDERSON: So we'll come back to

it. Anybody else have questions for Dr. Ross? Yes.

DR. SALIVE: So, on the flip side of the last question, could you justify why you gave low grades to the other ones? I think, you know, I would just like to hear your rationale there, I don't think we heard it in your talk.

DR. ROSS: So, my rationale for prognostic tests in colorectal cancer that might drive a decision to use untargeted adjuvant therapy, A, because I think adjuvant therapy shouldn't be untargeted, I mean the choice to use FOLFOX or FOLFIRI shouldn't be based on what the infusion nurses in the practice like to give, which is often what drives it more than some test that says that a platinum-based regimen would be more efficacious than an irinotecan-based regimen, that's not done, so that's one reason.

The second reason is I haven't seen in the literature dramatic impact of using either of those adjuvant regimens nontargeted on patients who are judged high risk with Stage II

disease, and that knowing the patient is more likely to relapse or not doesn't drive a better outcome for the patient because the adjuvant therapy has reduced that risk in a way that I have been impressed with, that's partly it.

The second reason for why I, for example, thought that BRAF as a prognostic test, I wouldn't have given it a high rating at all, but as a predictive test of response

therapy, I think it has potential, even though our first study with vemurafenib in colon cancer was a negative study, I see in the new emerging data that we are going to be able to target BRAF in the colon, so I want that test available for patients so they will know whether they are candidates for anti-BRAF therapy.

So those are my thinkings for those. And then obviously for lung cancer and RAS testing as a, quote, gatekeeper, where you would look for the RAS mutation first, and then only test the adenocarcinoma for a never smoker squamous cell, which the NCCN guidelines consider the same as an adenocarcinoma. You

wouldn't test them for any other alterations because you got a KRAS mutation, thus they're known targets, and I've had enough experience now to know that's not true. It's certainly rare that you would get EGFR or ALK in a KRAS mutated, but it's not never, it's just very very low, it's two or even closer to one percent chance, but there are so many other markers that we didn't talk about today that are KRAS-mutated lung cancer that could get a

patient on a targeted therapy, so I wouldn't want to use KRAS as a prognostic test and stop other treatment, other testing, because in my feeling that would deny a patient a chance to find a therapy that might work better than, you know, platinum-taxol in the front line, or methotrexate on the second line.

DR. HENDERSON: Okay. Yes, for Dr. Ross?

DR. BERGER: There's a little more granularity on the last question here. So given the multiplicity of techniques, and let's take KRAS for an example, okay, looking at it despite its problems, and despite the CAP's

proficiency testing program, is it fair to judge it prognostically in that regard, and assuming validity going forward to clinical validity, or do we have to subsegment the mechanisms of analysis in order to get a true answer on prognostic value?

DR. ROSS: In my opinion it is more likely than not that the KRAS mutation story as a prognostic factor will be similarly colored by the different sensitivity of the assays that have been used across the months and years it's

been done, similar to the way the poorly done immunohistochemistry clouded the obvious fact that HER2 overexpression was an adverse prognostic factor in breast cancer, which now we accept. It would take a long time to filter and look at whether it was pure Sanger or real specific EGFR with pyro-sequencing and with mass spec, or with next generation sequencing. But the issue is, if all the samples were pure and there were only tumor cells in the DNA that was extracted, all of the assays look pretty good and get it right, but as soon as you start dropping it to 90, 80, 70, 60, 50, different

assays lose their sensitivity for mutations, call it wild-type and then contaminate the prognostic issue.

DR. HENDERSON: Dr. Fischer and then Dr. Zuckerman. Dr. Fischer.

DR. FISCHER: I'd just like to ask a question of all the people who've actually dealt with large populations of people with various cancers. I didn't hear anything about nutritional status, which as a clinician has a profound effect on outcome. Did anybody look at the patient from the standpoint of where

they were nutritionally before start of treatment?

DR. ROSS: Certainly not in my comments.

DR. FISCHER: Do you have any data that was reported?

DR. HENDERSON: Dr. Ross, you don't need to sit down yet. We'll come back to you.

DR. MELETH: No, there was no study that was looking at nutritional status as a variable, or as a predictor.

DR. FISCHER: Unfortunately, that's

one of the prime outcome measures of patients with cancer, and I would hope that in the future when patients are treated, that that particular aspect of where they are clinically should be taken into account.

DR. HENDERSON: Okay. Dr. Zuckerman, do you have a question for Dr. Ross?

DR. ZUCKERMAN: Yes, I do, thank you. I guess this is a good followup to that one. Your graphics were very impressive, but with the N of one for each patient and whether it's nutritional status or other things, there's so many variables that affect outcome, and so I

guess, you know, I'm a little concerned because you talked about Phase I trials being used as approval status potentially for FDA, and your examples are sample sizes of one, so I wondered if you could speak at all to the issue of, you've shown these small number of patients, but obviously this is not what's happening with all patients, or perhaps not even most patients.

DR. ROSS: I think the philosophic focus of my presentation was that cancer is an

N of one disease and that patients should not be grouped, they shouldn't be grouped based on where their cancer started, they shouldn't be grouped on where it spread, it should be grouped based on what is the general driver alterations that are making it grow and spread, and are they targetable each patient one at a time, rather than the breast cancer group, whether it's HER2 positive or HER2 negative, or whether it's a brain tumor patient that's EGFR-driven or IDH1-2-driven, or NF2-driven, that the therapy should be decided by the actual genomic drivers of the disease, not the clinical status, the nutritional status or any

of those, because the new targeted therapies don't have the toxicity that current toxic drugs do. Patients with poor nutritional status, when they're being candidates for targeted therapy, tend to get the treatment and not have the nutritional status worsen because of the side effects of the drugs.

So that's where I'm thinking, every patient could be on that chart near the left side with an alteration very commonly seen in

that disease, possibly not particles like KRAS or P53, or they could be at the far other end, there's only a few people like them who have ever had an activating KIT mutation which progressed to colorectal cancer and was then denied therapy because the only indication for treating this alteration is if they have the diagnosis of this tumor type, when actually they have their own tumor type. We match the therapy to the genomic drivers, not to the clinical status, not to the immunologic status, not to my immunohistochemistry profile but based on the genomic drivers, which is, I think, much closer to the real value of these tests.

DR. ZUCKERMAN: But just to follow up, I guess it's a little confusing to me that -- I mean, you're talking about an N of one and I understand the concept, but that's not what CMS does, and also, that's not what these tests do. These tests are looking at responses that are not just for one patient but they are grouping patients.

DR. ROSS: I think certainly the

prognostic tests are trying to do that, but in my view of the predictive tests they're only considering one patient one at a time. Every individual patient's genomic signature is theirs and theirs alone, they don't want to go to the shoe store and get the only nine-and-a-half size that the shoe store gives them, they want to have their foot measured first and maybe get one that fits better.

DR. HENDERSON: Thank you, Dr. Ross. Anybody else? So I have two questions.

The first one is sort of incidental, maybe in a little more depth but it comes from the last discussion. And so you know, you clearly have a lot of knowledge about HER2. Why has HER2 been so slow in being evaluated as

a target for other tumors after breast cancer? There's almost a ten, at least a good solid ten-year lapse, and yet if you go back and you look at the early breast cancer data, as you yourself have indicated, there's quite a bit of controversy about the percentage of patients and so on. It wasn't that this was an obvious target to go after, in fact I agonized over

this quite a bit, whether to get involved in the HER2 receptor studies because it seemed like so few patients would benefit. So why was there a lag in testing this, for example in gastroesophageal or other cancers? There's a lot of other cancers out there where the data in terms of frequency and so on for HER2 is not much different from breast. Why hasn't this been done?

DR. ROSS: Well, I think there's a lot of -- that's a wonderful question and I'm delighted that you asked, because I think it's an absolute tragedy, that patients with gallbladder cancer are not tested for HER2 and put on anti-HER2 targeted therapy, because they will respond every bit as well as the gastric and the esophageal, it's not anywhere near as

mixed there.

DR. HENDERSON: Why hasn't that happened?

DR. ROSS: I think a couple of things, Craig. I think the testing was still not perfect when trastuzumab was released, we didn't know about the impact of this. Right

after the drug was released there was a sudden pullback of use because one national renowned laboratory said that they were getting more than a 50 percent positive rate with the knockoff Herceptin test, the unlabeled test.

Then of course we went out and looked and tried to figure out what was happening and they were using microwave oven antigen retrieval when the package insert for the knockoff Herceptin test said that the polychrome antibody was far too sensitive for microwave, you need to use a water bath when you can control the temperature better. When that lab then used the water bath and followed the instructions of the FDA-approved kit, then they got 28 percent or 25 percent positives, and then therapy began again. So the test held it back.

And I think that also held back, excuse me, the applications of the tumor type. We came so close in bladder cancer to getting a registration trial started but then they pulled back. There are many tumor types that would benefit from being able to be treated with

anti-HER2 targeted therapies that are not on label now. If we switched and said we don't care if it's colon, we don't care if it's bladder, we don't care if it's breast, if it's HER2-riven it gets HER2 targeted therapy.

It will cross tumor types. We proved it with gastroesophageal, it can go to bladder, it can go to a whole host of gynecologic malignancies, and I think this is, my own feeling of what we should be doing, rather than relying on the traditional way we develop drugs, and that's why I think we will see targeted therapies get approved more and more in a single arm, no comparator group, and even Phase I some day will be enough.

DR. HENDERSON: Okay. So I want to follow that with a more directed, well, actually now another one, so two directed questions.

Your last comment had to do with the broad question which we're going to be talking about later this afternoon, about FDA regulations versus CLIA tests, and let me clarify a bit more. As you were talking I

thought, well, you know, part of this issue might be related to the fact that if we are going to test, for example trastuzumab in a, you know, in order to get a clearance, an FDA clearance, and do it in each one of the tumor types independently, I think if you look at the amount of work that was involved even in the breast cancer, it was enormous. And further, the breast cancer studies were seriously underpowered and we probably got the positive result we did more by luck than by strategy.

DR. ROSS: Especially when we treated two-plus patients the same as if they were three-plus, when we know three-quarters of those are HER2 negative.

DR. HENDERSON: Well, we did that in Phase II, but anyway, that may have been an element. So that comes, you can make a comment, but I think this afternoon when we're talking about CLIA versus FDA, this is kind of

a footnote of some importance and that is, does the FDA focus and insist on a focus on each of these separately? If so, that makes it much more cumbersome, to me it would be a much more

cumbersome process, developmental process, so you may want to think about that. Do you want to make any further comments on that?

DR. ROSS: No.

DR. HENDERSON: So, I have another question which, before the discussion a minute ago, I thought up actually while you were talking. So is the KRAS assay used for colon and lung essentially different?

DR. ROSS: Certainly the KRAS in colon is FDA-regulated in the label for Erbitux and panitumumab, in other words, a specific protocol on how to do the test is in the label. That doesn't mean that laboratories don't do comparison studies with their own in-house tests and LDTs that show comparability with the label-approved test and then go and do their own in-house tests. But there isn't, for lung there isn't an FDA approved in label because there's no targeted therapy for KRAS and there's no restriction of use of the therapy

based on KRAS like there is in colon, so they're different in that way. There's FDA approval for colon but that doesn't mean every

test being done is the FDA protocol, but there is no FDA-approved test for lung for KRAS.

DR. HENDERSON: The reason I'm confused about this is because for, just looking at analytical validity, question 1(a) which we have to address this afternoon, you've given KRAS for colon a five and KRAS for non-small cell a two, and so that would suggest that the analytical validity, the ability to do the test and get the same result repeatedly is different for lung and for colon.

DR. ROSS: Oh, I'm sorry if that's the implication that you take from that. What I was trying to say there is that in colon it's a predictive test to guide the selection of personalized therapy. Erbitux and panitumumab are not indicated if KRAS is mutated and it's not just the codon two but the entire coding sequence of KRAS or NRAS, they all have to essentially make the use of those drugs not FDA-approved. But in lung cancer the KRAS is often being done as a screening test to

eliminate this need or a decision to do EGFR, to do RET, and then a whole host of others, and

there's a squamous side that can guide therapy like MET, FGR and others, and I don't feel that KRAS can be allowed to be a gatekeeper for who should get further testing on lung cancer because we will miss way too many opportunities to treat patients with targeted therapy.

DR. HENDERSON: But it seems to me that your answer addresses question 1(b), 2 and 3, but question 1(a) would be the same answer based on what you said thus far, and maybe I'm misunderstanding, it would be the same for lung or colon.

DR. ROSS: No, you're right. I was jumping to a composite of all the questions in terms of did I think this test had clinical utility whether or not it was clinically valid, so it's my summary opinion.

DR. HENDERSON: So you would say the questions on analytical validity --

DR. ROSS: They should be the same.

DR. HENDERSON: They should be the same for the two tumor tests.

DR. ROSS: Definitely, again regarding

the fact that the colorectal is FDA-regulated

whereas the lung is not.

DR. HENDERSON: Okay. So, any other questions? Yes.

DR. KAMRAVA: I just have a question. I think a lot of the summary things like in the slide are reflecting a desire to have a predictive marker, and I think some of the comments from the other speakers is that we developed this as a predictive marker, it wasn't meant to be prognostic, but our task up here is really to evaluate the prognostic significance of these things and I'm just wondering, do you think that's even a fair thing to do independently, especially for some of the markers like ALK and EGFR, because that's not really what, you know, I think the consensus here is that they were actually designed to do.

DR. ROSS: I agree completely. I mean, no one would not do EGFR or ALK testing in a non or never smoker with a lung adenocarcinoma because it was or wasn't prognostic. That's not what the thinking is, these are tests done to decide whether a

patient should possibly go on a lifesaving or a life-altering therapy. So whether it predicts a net outcome or a good outcome is irrelevant. These patients already have high risk disease when the test is ordered. No one's ordering ALK or EGFR in a T1 lung cancer to predict whether it's going to relapse or not. It's being ordered on patients who present with Stage IV disease or in patients who originally were a lower stage but unfortunately relapsed after surgery or after front line chemotherapy, so it's a very difficult list of tests to comment on because it was a mixture of pure prognostic tests, partially prognostic, partially predictive, and totally predictive tests, so it challenged, I think, all of the speakers to try to make sense of that.

DR. HENDERSON: Okay. Any other questions for Dr. Ross? Yes.

DR. SALIVE: I think you placed a lot of emphasis on negative predictive value in your talk and the question I have is, you know, partly driven by the sensitivity/specificity and if you have -- and you also made, I think, points about false negatives. And so you have

this great concern I think in practice if you have a worry about false negatives, you can't weight that negative predictive value as highly. So how does that play out, really, for evaluating these tests?

DR. ROSS: So when you're talking about the predictive tests exclusively, the tests that have been developed, even if we start first with, let's say HER2 testing, traditional HER2 testing, the goal of the test isn't to identify who's going to benefit from any of the HER2 target therapies available. From some of the adjuvant studies, of course, we know we don't want to miss any woman with HER2-driven breast cancer and the adjuvant therapies are our one chance to cure that patient for life. If we don't get that patient on HER2 targeted therapy in the adjuvant setting an she relapses, and then we discover that she's HER2-driven, we can do a lot, she will live longer and do better as long as we try to keep it out of the central nervous system, but we can't cure her, but in the adjuvant setting we believe we can, the adjutant data believes that if we target her to

the adjuvant setting.

So the negative predictive value or the negative test result is so devastating, for a woman to be actually HER2 positive, the test missed it for any reason, and now she relapses and she's positive and we go retrospectively and look again, oh, she is HER2 positive, we should have probably used this, we caused enormous amount of harm to that patient by missing that HER2 positive breast cancer early in the disease.

That's not so true for all of the other markers, because the adjuvant targeting therapy is very limited right now, and most of it is really for Stage IV relapse, and there's an example where getting a negative result, a false negative, is so really really devastating. In lung cancer, though, we also see patients like the patient I showed that got false negative ALK FISH results. Thank goodness we were able to find that with a different platform and then get the patient on the proper therapy, apparently early enough, because this fellow is now living in his fourth year on crizotinib maintenance therapy.

So, a false negative is devastating when it's a target that we have a drug or multiple drugs available, because we deny that patient a chance to benefit. That's what I was driving at when I felt that's more important than any other, it's what we want to drive at, we want to have the lowest possible false negative rate by doing the most sensitive test that will work even in the impure mixed tumor and benign sample, which challenges all the tests we do.

On the other issue you were asking about, negative predictive value of the tests, in general all of these tests do the same thing. They can't tell you a patient will benefit from the therapy, but what they're saying is if they test negative, we don't want to treat them because the toxicity, the cost, the false hope is just so high.

We don't treat HER2 negative breast cancer right now with anti-HER2 targeted therapy. There is a trial that I haven't heard the result yet that is trying that, I think it is going to be a negative trial, we really need

the marker to be there in order to get the

benefit of the therapy, so we're doing negative predicting. We're not guaranteeing a positive test result will drive a probable positive outcome, what we're trying to do is not treat a patient who has a negative test for those reasons, avoid toxicity, avoid costs, avoid false hope.

DR. HENDERSON: We need to be moving here.

DR. BERGER: On that question, so to that end, do you have a specific recommendation for a test to be technically valid, whether its signal-to-noise ratio should be for distinguishing wild-type mutation?

DR. ROSS: Well, you know, it's difficult, as I mentioned, we're still not sure if this testing is done with no attempt to control it while it's happening, and when a test is a single gene, a hot spot alteration, even the on-label test in the drug label is done without some kind of what I call surrogate control, a cell line mixture of different concentrations of the malignant cultured cells

that have the alteration, a lab will report a negative, did not find EGFR in this lung cancer

specimen, did not find KRAS in this colorectal sample. That result is either a true negative or a false negative. We don't know because there's no built-in control.

Now if you run the cell lines along with it and you identify them at very low concentrations of EGFR mutations or KRAS mutations, you're just saying my system was in control, it detected the controlled cell lines like we expect it to, and the patient's DNA that was run right along with them, right through the same flow cell or through the same PCR reaction, tested negative, you're relatively confident that it's a true negative. If you don't run controls, your negative result is either a true or a false negative.

DR. HENDERSON: Okay, thank you, Dr. Ross. Don't go away, we may have more questions.

Dr. Rollins, you had a question or comment?

DR. ROLLINS: Yes, I have a comment

from CMS's perspective. Earlier during this discussion there was some discussion about Medicare policies and clinical validity and

clinical utility. It is true that clinical utility, in order for a diagnostic test to have clinical utility it's got to have analytic as well as clinical validity, so that's the basis for clinical utility, whether or not it reaches it or not, in terms of whether or not it's a diagnostic test or a predictive test or whether the results will apply.

In terms of treatment, since we're specifically looking at the prognostic value in terms of whether or not the disease will recur, considering the treatment perspective is sort of a moot point. So since we're still focusing on the prognostic value of the test, we're still looking at recurrence as opposed to treatment.

DR. HENDERSON: Okay, thank you. So, Dr. Nowak, you wanted to answer one, or make a comment on one of the questions addressed with Dr. Ross.

DR. NOWAK: Well, I have, I'm kind of

champing at the bit here because I have a lot of information that I think you would find useful. I can address your question and clarify what the issue is with FDA, if I can.

DR. HENDERSON: Sure, go ahead.

DR. NOWAK: FDA approves tests for specific indications, so there's a BRAF assay that has been approved for testing metastatic malignant melanoma. It has not been approved for use in colon cancer, it has not been approved for use in lung cancer, so any use of that test outside of metastatic malignant melanoma makes it a laboratory-developed test. There is no FDA-approved test for BRAF testing outside of that setting.

Same thing, KRAS testing in lung cancer, there is no FDA-approved test for that. The story with RAS testing in colon cancer in 2004, the targeted therapies against EGFR were first approved and the companion diagnostic that FDA approved at that time was immunohistochemistry testing, a specific immunohistochemistry test that tested for EGFR expression, the rationale being that if you

could detect EGFR in the cells, then there's a chance it would respond to the drug. We went for five years using that companion diagnostic and it didn't work.

In 2009 there were a couple of

retrospective studies that were presented to ASCO that said it's not expression of EGFR that's important, you need to look at what happens downstream and if you have a mutation in KRAS, it causes constitutive exacerbation of KRAS, and inhibiting an upstream molecule, EGFR, makes no difference. The day after the ASCO meeting all of our oncologists came home and said we need KRAS mutation testing, KRAS codons 12 and 13 mutation testing.

The laboratories developed these tests and from 2009 to 2014 there were only laboratory-developed tests that were available for KRAS testing. In two-thousand, I guess 2013 or 2014, finally an FDA-approved test for KRAS codons 12 and 13 became available.

So in the meantime, it's laboratory-developed tests that are providing this testing for cancer patients, and the test

that was approved in 2013 is specific for codons 12 and 13. In the subsequent year we've learned that it's not only mutations in codons 12 and 13, but mutations in 12, 13, 59, 61, 117 and 145, and those same mutations in NRAS are also predictive of response. So once again,

laboratories have developed their tests to answer this need and all oncologists are using this test now, and we're waiting for an FDA-approved test to appear that actually does that.

So that's the situation with both RAS testing and with BRAF testing. Using any of these tests in a different organ system, those are laboratory-developed tests, and it's the laboratories that take responsibility on their validity.

The comments on validity, absolutely laboratories are concerned about validity and controls, you need to know not only how good your test is, you need to know how bad it is. No test is perfect but you need to understand its shortcomings, what it means when you say it's positive, what it means when you say it's

negative, and I think you all know that.

The comments about co-testing with BRAF and MSI, I would be happy to address that for you. So in the context of predictive testing, as Dr. Ross points out, well, maybe BRAF isn't very important, but outside of that as a prognostic marker, BRAF is hugely

important. BRAF and MSI, after grade and stage in colon cancer, are the most important prognostic markers that you could look at, and you need to know this.

You know -- I know -- you know, Dr. Caughron has slides that specifically address this, I'd be happy to go through those for you. It's hugely important. If you have colon cancer, if anybody in your family has colon cancer, after you have the stage, the most important things to know are MSI status and BRAF status, because that tells you your risk for an inherited predisposition, Lynch syndrome, and if you're not MSI-H, then BRAF, a BRAF mismatch repair proficient tumor is the worst kind of colon cancer you can have. It's that simple. You need to know this.

You know, I would really like to go through some of Dr. Caughron's slides, I think it would clarify a lot of issues for you, but I'll leave that to you.

DR. HENDERSON: I think that was decided earlier.

DR. RAMAMURTHY: Can I ask a quick follow-up?

DR. HENDERSON: Yeah, sure.

DR. RAMAMURTHY: I completely appreciate that it takes time and resources for a test to be FDA-cleared or approved, and that is valuable time when a patient cannot lose access to care that they need critically. That point's been made over and over, and I completely sympathize with that point. The issue at hand is when the tests work and patients are saved, great. But when the tests don't work, there is no adverse event reporting system, is there?

For example, I was at ASCO last year and I befriended a doctor from Princeton, an oncologist, and she wrote me an e-mail since I used to work at the Agency before, that seven

of her patients were wrongly statused for HER2, and so they had to resend the samples off to a different laboratory because she thought her physical examination of the patient's cancer did not comport with the results of the test.

So I think that's the wanted question. I know you were chomping at the bit, I could tell, but that is the point, it's about knowing when it doesn't work. When it works, yes,

lives are saved.

DR. NOWAK: So, I made a point of mentioning that proficiency testing is one part of a quality improvement program. The other parts of a quality improvement program are running your controls every day hour by hour, watching whether your controls work, and you know hour by hour, day by day, week by week how good your test is. You do the proficiency testing so that you know how well your test compares to what they're doing at Mayo, what they're doing at Sloan Kettering, and in doing that comparison, everybody has to rise to a certain standard, and hopefully that standard continues to go higher and higher.

Adverse event reporting, you know, the adverse events that FDA requires, and I just, I talked about this at the FDA workshop in January, the adverse events that FDA requires are death and irreversible serious injury. At the end of the week or the end of the month when I sit down with my laboratory and go over our quality assurance monitors where I ask about how things are working, I don't ask how many people have died as a result of our

testing this month, how many people have suffered irreversible injury. Those are important but those are very coarse measures of how well these tests are performing. I know how well our tests are performing. Most laboratory directors, if they're conscientious, are monitoring their testing hour by hour, day by day, week by week, and month by month, and those are very granular measures of the quality of their testing.

The other thing I should say about uniformity of testing is that the professional societies have developed guidelines for how these tests should be done. So there is the

ASCO-CAP guidelines for the performance of HER2 testing, and while ten, 15 years ago the quality of testing may have been nonuniform, I would say that now it is pretty good and there's data to support that. There are guidelines in place for the molecular testing for tyrosine-kinase inhibitors in lung cancers and those are sponsored by, so it's CAP, AMP, IASLC, the International Association for Study of Lung Cancer. Those guidelines are now being revisited and they are going to be renewed, so

there will be recommendations made for how testing should be done, how tests should be validated, and how good they need to be. And also, they will touch on in which patients, which are the correct patients that need to be tested and which specimens need to be tested and how should they be handled.

There has been an ongoing effort that involves ASCO, CAP, AMP and ASCP to develop practice guidelines for laboratories for testing for molecular markers in colorectal cancer. I'm part of that effort and I can tell you that we've been going through the

literature, these are to develop evidence-based guidelines, and we've recently done a refresh on our literature because we expect to put out draft recommendations probably by the end of this month.

One of the cochairs of that effort is Dr. Stanley Hamilton, who's sitting in the back of the room. You know, ask him, ask him whether that effort is showing that there's evidence or not for the testing of these things. I mean, he has that data, and I know that some of that data was included in

Dr. Caughron's presentation, and that's why I was so anxious to share that with you.

DR. HENDERSON: Okay. Dr. Burke.

DR. BURKE: Thank you for a nice presentation, I liked it a lot, and I'm from Evanston, so I have to disclose that. So, a couple questions. How many laboratories are doing BRAF testing in the United States today?

DR. NOWAK: From the CAP proficiency tests?

DR. BURKE: No, just total.

DR. NOWAK: It's at least that many.

DR. BURKE: Right, but how many total in the United States are doing BRAF testing?

DR. NOWAK: I can only estimate that it's probably something less than 300.

DR. BURKE: So if the BRAF testing was found to be useful, lots more laboratories would probably be doing it. 50 percent more do it than are subscribers right now in your system, your system is not a random sample of laboratories. So the question is, how do we know that once we move beyond the 204 subscribers which are the, you know, the high performers, that other laboratories would be

able to do the job? Have you done a random sample of other laboratories who do BRAF testing?

DR. NOWAK: So, first, I don't believe that the number of laboratories would increase that much. If you -- the proficiency test for MSI testing, which requires a certain kind of technology, has basically been stable at just over a hundred laboratories. Testing for some of these markers is maybe a little bit easier but there's a limit to the number of

laboratories that actually can do this.

DR. BURKE: And why would there be such a limit?

DR. NOWAK: Because to do these kinds of tests, particularly as a laboratory-developed test, you need to have personnel, you need to have professionals who understand how to develop these tests, how to validate them and then how to perform them, and many smaller hospital laboratories simply don't have that expertise.

DR. BURKE: So you're thinking that CMS wouldn't be involved in more than 300 laboratories with this test going forward,

that's your prediction?

DR. NOWAK: Yes, if they're simply looking at the test. But if they're going to look at every indication for the test then you need to multiply that, because it's testing for melanoma, it's testing for lung cancer, it's testing for thyroid cancer, and on and on, and so if they're going to do it that way, then there is a huge number -- it makes every indication a different test.

So I've argued what they should really approach is to identify the companion analyte, what is it that you're testing for, identify a BRAF mutation. After all, it's a biochemical alteration in the DNA, this is a matter of analytical chemistry. How well can you detect it, how certain are you that you have it. There are other variables because you're dealing with biology. You get tumors that have different amounts of cellularity, acidosis, different amounts of fibrosis, that complicates the issue, but that's why you have pathologists looking at these things.

DR. BURKE: That was my next question, thank you.

So there has been some talk in the literature, now I just have something on non-small cell, about the sufficiency of biopsy material, the different quality, different amounts, and your analytic testing of the 204 subscribers wouldn't address the issue of quality of the specimen or amount of the specimen.

DR. NOWAK: Oh, absolutely they do. I

mentioned that among the CAP surveys there are supplemental questions, and one of the things that the survey has been focusing on lately is simply the pathologist's ability to accurately assess the proportion of tumor cells. So they've sent out a bunch of challenges, either microscope slides or actual pictures and asked people, in this picture what's the proportion of tumor cells? And their performance is not great, and there's various reasons for that because the statistics are just lousy if you're looking at a group of cells.

Nonetheless, they're trying to define that and make pathologists who do these things aware of the limitations of what it is they're doing, and in fact there is now a new survey

coming out of CAP that is specifically directed at that, it's an assessment of the proportion of tumor cells. The other part that comes into this is what kind of specimen, do you get a resection specimen where you have grams of tissue, or do you get a core biopsy, and even though you may have a core that has very nice tissue, there could be very few tumor cells in

it. So again, you need to understand and appreciate the sensitivity of your assay and relate it to the specimen and understand when you have sufficient specimen that you have confidence in the negative results and when you simply don't. You can still do the test but you simply won't have confidence in the result.

If it's positive, you may be in, but if it's negative, you don't know that it's a good negative.

DR. BURKE: So, do you have those numbers off the top of your head, about the performance with inadequate specimens or poor number of cells? You alluded to the performance not being really great, but do you have any sensitivity/specificity numbers for us?

DR. NOWAK: Overall, no, I can't cite you overall evidence but I can give you my own experience, and I think every laboratory can go through this the same way. If I test colon cancers or if I visit any laboratory that's testing colon cancers for KRAS mutations, I expect they're going to tell me that they

detect it in about 30 to 35 percent of cases, because that's what the literature says. If they tell me they're only detecting it in ten percent of cases, then I think there's an issue.

It's the same thing with HER2, as you heard, if you're detecting it in 50 percent of tumors if it's amplified, well, there's an issue, and if you're only detecting it in five, there's an issue. So there's an immediate benchmark there, that if you can detect, in your large specimens if you're detecting KRAS 30, 35 percent of the time, I'm reassured that you're at least achieving the benchmark that's reported in literature.

Then I can ask, well, if you're doing this on FNA core biopsies with liver mass where the tissue is much smaller, I'd ask the same

question, you know, is your assay detecting KRAS mutations in the appropriate frequency for what is reported in liver mass? So that's your first line assurance that you're doing well.

And then you can take it further. In my lab we work off of cytology smears

sometimes, we take groups of cells off of pap stain smears, and I worry a great deal about the sensitivity of our assay there. But still there, I'm achieving that same frequency of testing, so that assures me that I'm probably doing okay.

But then there are other issues when you get to the very very small samples, or if you get to a very very sensitive test, then you need to start worrying about tumor heterogeneity. It's great if you have an assay that can detect things down to .01 percent, and maybe that's what Dr. Ross's assays do, but that's a different animal, and we have not yet begun to understand or to address the importance of tumor heterogeneity. Sure, you may detect something at a very very low level and I don't doubt that it's there, but then we need to understand what that means clinically.

DR. BURKE: All right. So the results that you gave us basically have an asterisk of what you've told us since then, your slides have an asterisk on that sensitivity and specificity, the accuracy, based on what you've

just told us about inadequate specimens, heterogeneity of material, would degrade that accuracy.

DR. NOWAK: I mean, it depends on how you're looking at sensitivity. If you defined it in a very broad way, then it certainly would.

DR. BURKE: Okay.

DR. HENDERSON: Okay. So we've actually utilized the time allotted here, so you can sit down, Dr. Nowak.

But Dr. Zuckerman had a question that goes back, and I didn't want to ask you to wait, so I think it was for Dr. Conley?

DR. ZUCKERMAN: Yes, for Dr. Conley, and feel free to expand, because I feel like we've focused a lot on studies with an N of one, and I would like to hear whatever you have to say that might expand on studies with larger sample sizes, but to start with the question of

the inconsistency on the testing for EGFR.

DR. CONLEY: Yeah. So the reason that I said that EGFR mutations may be prognostic too is because, you know, when EGFR inhibitors

were first tested, they were tested indiscriminately, we didn't know EGFR activated mutations, and so we had patients treated with chemotherapy and patients treated with chemotherapy plus EGFR inhibitors, and on some of those trials we were able to go back and look at the tissue and see whether or not the patients had an EGFR-activated mutation. So only about ten percent of patients will have an EGFR-activated mutation, which is very low. However, when you look at how they do with standard chemotherapy, they do a little better than people who don't have it across the board.

But of course if you treat them with an EGFR inhibitor or a standard chemotherapy, they do much better with the EGFR inhibitor and they might do a little bit worse with the standard chemotherapy, so it is a little complex. There's a little bit of prognostic ability there and it seems that these tumors, although they present metastatic, may grow a

little bit slower than those that don't have this mutation. So as that data are coming out, we're sort of taking that into consideration,

and the T790M mutations, that data is very new so they wouldn't have gotten to it, I think.

DR. ZUCKERMAN: So it is newer data, and the samples are reasonably large or small, or what do they look like?

DR. CONLEY: Well, yeah, here we go, okay. So this is lung cancer, and if they present metastatic they're just going to be small samples. If they had resection previously, there may be larger tumor, but the issue that we have there is does that tumor reflect the tumor you have now, because sometimes the mutations that you can detect later after the tumor comes back, or certainly after treatment, might be different.

DR. HENDERSON: Okay. Just a second, Dr. Conley. Any other questions directed to Dr. Conley?

DR. ZUCKERMAN: Actually I have a followup, thanks. I would like for you to address this issue of, you know, we don't have much sample for the Medicare-aged populations,

and we know that there are differences for older people on a variety of levels, but partly

their exposures over time, so I was wondering whether you have any insight into the research that's been done looking at patients 65 and older, or if not, what the implications would be for using a subset of younger patients to try to understand what would happen to Medicare-aged patients.

DR. CONLEY: Yeah, that's a lot of questions in one. So, the incidence of the cancers we were discussing, with the possible exception of breast cancer sometimes is, you know, the incidence increases as you get older, and the median age is getting older and older of the diagnosis of the cancers. Yet, we do have a dearth of older patients in clinical trials. That median age tends to be stuck in the 60s, whereas the patients you may see in practice could be in their 70s or 80s, so it is a relevant question as we use the tests, the prognostic tests, the truly prognostic tests and the all-mix predictors, so I would sort of like to make a distinction between the all-mix predictors and what we call the multi-analyte

tests where you can do all the tests in like

one test like Foundation Medicine has done, and it is not exactly an all-mix predictor as much as a multi-analyte test. An all-mix predictor, though, is more like the Agendia test or the Oncotype Dx test. They take a bunch of genes and their relative composite result is the test, you know, it's not one gene by one gene.

So those are a little different and it is an area of, that we are trying to encourage research in, I don't know how much research there really is, as to, you know, what do these other things have to do with it, the nutritional status, the age of the patient, what their exposures are, I think those are ongoing research questions.

What we are talking about today, though, these really specific tests, to my knowledge it doesn't really matter about the age, with the exception of the MSI, which tends to occur in younger patients, and has ramifications beyond the patient who had the initial tumor.

DR. ROLLINS: Can I make a quick comment from CMS's perspective? It is true

that about 85 percent of all Medicare beneficiaries are 65 and older, but we still have a sizable number of persons who are younger than 65, so that information would still be of value to CMS.

DR. HENDERSON: Dr. Kamrava?

DR. KAMRAVA: Do you have any insight into the heterogeneity question again in terms of, we have been thinking about some of these markers as yes present, no present, but do we have any information about their prognostic value in terms of, you know, what percent of cells are ALK positive, does it matter if it's 60 percent versus 20 percent, because we've kind of just been treating it as yes-no.

DR. CONLEY: Yeah, right now we treat it as yes-no, but it's obvious that even if you put all the patients who have the particular eligibility marker in the same trial, they're not going to all have the same response, and I think right now we're starting to tease out why not. I mean, it started kind of the other way around where we had the exceptional responder kind of patients where you had patients on the trial, only one had a CR and they had a

particular mutation, let's say, and you looked for that mutation in all the other patients, a few of them had it, most of them had a little bit of decrease but none of them had that CR, right? So there are particular differences among patients even if they have a targetable eligibility criterion, and we're just starting to tease those out at the moment.

DR. HENDERSON: Okay. Dr. Scheuner.

DR. SCHEUNER: I hope my question is brief. I'm just wondering about the MSI again and I think we saw it's been around for ten years and if it's present it's a good prognostic factor, and those folks generally are not responsive to the 5FU type med, so, has there been no literature on physician decision-making, patient decision-making around that, do you know if anything is coming down the pike? I was surprised that there was nothing in our tech assessment with regards to that.

DR. CONLEY: You know, I will have to, again, invoke Stan Hamilton back here, an expert in this area as well, and has been for those ten years and more looking at this issue,

but I don't think we have a definitive answer yet on what to do with an MMR deficient tumor. We are starting to delve into the DNA repair mechanisms and their interactions with particular therapies, and to try and figure out if we can choose a little bit better those patients who will respond to it or not, because obviously if you didn't really benefit from irinotecan or oxaliplatin, nobody would raise their hand to get it, you know, if you knew that you weren't going to benefit from it. On the contrary, if you knew you were going to benefit from it, well, okay then, it's worth the risk of doing it, they're decent drugs.

The question is, why do you use this test and why we used it, we would use it, we started it in Stage II, right, like I was saying, and so what's the big dividing line between Stage II and Stage III, maybe there are better behaving Stage IIIs that should be grouped with Stage II instead of Stage III if you're going to be grouping, and we do know that adjuvant chemotherapies benefit patients who have Stage III colon cancer. It might be a little benefit, same thing like breast cancer,

but there is a benefit, and it won't benefit everybody, and again, you know, how do we actually look at that?

I have not seen any, to me, useful information to inform how to change what one would do other than, you know, you would consider not doing an adjuvant treatment in Stage II if the patient was mismatch repair deficient.

DR. HENDERSON: Okay. Any other questions for Dr. Conley? Okay. We're in overtime now, so let's keep it focused, please.

DR. BURKE: So, with MammaPrint, just a couple quick questions. So let me preface it by saying when I see ROC, AUC and C index, they all mean the same mathematically. And I should also say in 2004, you know, I published an ROC of .7 or higher for clinical validity, so I have to tell you that up front, which would translate to a sensitivity/specificity pair of about 140, and I like the signal detection question, that was very nice.

So first, let me see if I understand. Your test predicts low risk for recurrence; is

that right?

DR. VAN DER BAAN: Correct.

DR. BURKE: So it doesn't predict who will not respond to therapy; is that correct?

MR. VAN DER BAAN: Well, so in this exercise we were not supposed to talk about predictive therapy.

DR. BURKE: I understand, but I'm just trying to understand your test. So you don't predict who won't respond to chemotherapy?

MR. VAN DER BAAN: The risk of recurrence and low risk is so low that actually we can theoretically calculate that the benefit of --

DR. BURKE: No, I understand the analogy, but I just want to be sure about what your claim is.

DR. VAN DER BAAN: The claim is the FDA claim.

DR. BURKE: That you can find, your scientific claim, that you can find patients with low risk for recurrence.

DR. VAN DER BAAN: Correct.

DR. BURKE: And so it could be in fact

the case that not all low risk patients have
the same outcome, some have better outcomes

than others, and in fact there's nothing in
your test that would preclude somebody who
would be chemotherapy responsive from not being
in your population.
DR. VAN DER BAAN: As long as there is
recurrence in either group, and chemotherapy is
a broad spectrum agent, theoretically everybody
benefits from chemotherapy.
DR. BURKE: Okay. Now in your slides
you presented, I think, a five-year
sensitivity/specificity pair of .9 and .42, and
a ten-year of .84 and .42, which is under my
threshold of 140 combined, I should tell you,
and in an ROC for your test has been claimed as
.68, so I looked at -- I don't know if you're
familiar with this study, it's by Mark Ceroni
and it's a pretty well-known study, and they
did a validation, an independent validation of
MammaPrint, and they found an ROC of .59 with a
confidence interval of .55 to .62.
DR. VAN DER BAAN: Did they use our
test?

DR. BURKE: They took your genes and they reproduced -- now hold on.

DR. VAN DER BAAN: That's not the

same, I'm sorry.

DR. BURKE: Hold on. I understand what you're saying and we can talk about that, but they did an independent validation, okay? So it wasn't your company doing its own validation, it was these people who had no vested interest, and they came up with a .59, and yet these things are supposed to be replicable, right?

MR. VAN DER BAAN: We have a hundred percent interlaboratory, we have three labs, we have a hundred percent. Our (inaudible) validation, which would have been a better paper, was analyzed in Lausanne independently and also had a hundred percent, so -- sorry. It's really important that the test is the test on the same microarray platform using the same SOPs.

DR. BURKE: That was their goal, and by the way, their paper has been cited in a number of other publications, they haven't been

criticized for not doing it properly, so, you know, sometimes when people do their own testing, their results are a little better than when other people do their testing, and so this

suggests that rather than an accuracy of .68, predictive accuracy, you actually have an accuracy of .59 on the independent data set. And one of the reasons I set an ROC of .70 is because the variance on these things is usually about .04 on the ROC, .05, something like that, and so if you're pretty far below .7 you're going to dip into this range, which you did.

So the question I'm asking you, so you would say that this .59 on this independent validation data set, that that wouldn't be correct and that your own data would in fact be the correct data?

DR. VAN DER BAAN: Correct. And I think that what we've shown in the prospective outcome data, the 97 percent metastasis-free survival at five years is actually, even in the real life situation, it's even better than in the retrospective validation that we published.

DR. BURKE: Okay. Well --

DR. VAN DER BAAN: And now hear me out. I'd like to stress, to make one remark on the earlier comments on the elderly patients. So actually we went to the FDA for an independent postmenopausal validation and

clearance on our assay for exactly that reason. So we felt that's one potential biological change that might influence breast cancer, so we took a separate data set and took it to the FDA, so one of the six clearances is for postmenopausal specifically.

DR. HENDERSON: Thank you. So if there are no other really urgent questions, and we've, by the way, gone over what was allotted for this question part, so if you don't have anything that's really urgent, then we're going to break for lunch. So, we expect to be back in exactly 60 minutes, so that's four minutes of one.

(Luncheon recess.)

DR. HENDERSON: So, we are reassembled here, I think everyone is here, in fact I think every seat is filled now, so let's get started.

Now, this segment that we're moving

into now is an open discussion, and I'd like to encourage everybody to really stay focused now in our afternoon session, because obviously a lot, it's obviously a complicated topic, I think that came out in the questions this morning, but we should remind ourselves that

we're focusing specifically on the questions, the questions are focusing very specifically on the issue of prognosis, not predictive factor, not therapy, not the effects of therapy, but specifically on prognosis.

Now the first thing I want to raise as an option, if anybody would like to focus on any of the questions. We don't really have the option of saying I don't like this question so we're not going to vote on it, even though I know from casual remarks that some of you might say oh, that's what you would like to do. But if on any one of the four questions, anybody has specific changes in wording of a question that you think will make the discussion more meaningful or more doable, now is the time to make those suggestions and we'll discuss them.

So we're looking now for not a

critique of the questions but any suggestions for specific word changes that you think are going to be appropriate in order to make them clear. Dr. Ramamurthy, did you have one?

DR. RAMAMURTHY: I think we kind of touched upon this before, but just to kind of lay, for the record to consider is, when you

talk about how confident are you that the existing evidence is sufficient to confirm the analytical validity, I'm not sure one can blanketly talk about tests in a blanket fashion because as we realize, there are lots of different tests using lots of different technologies, so that it's hard to say that. So, I mean, it would be great to qualify the question a little bit or something like that if you were to consider that.

DR. HENDERSON: Specifically which question are you talking about?

DR. RAMAMURTHY: Question 1(a).

DR. HENDERSON: Now we will be doing it for each one of the tests independently.

DR. RAMAMURTHY: Right, but each test, here is each biomarker; there are also several

tests that offer, for that biomarker, there are several mutations tested and given biomarkers and so on and so forth, so this is a very blanketly open question, that's all.

DR. HENDERSON: Okay. So, do you have any suggestions for specific word changes that you think would solve the problem?

UNIDENTIFIED PANELIST: Can you give

us a specific example like KRAS or BRAF, or the different subgroups, is that what you're saying?

DR. RAMAMURTHY: Right. I mean, except the kind of modification I'm thinking of would be a substantial change, so I'm not necessarily asking that we change the question now, but I wanted to place on record a qualifier that this question should be answered with the qualification that when you say BRAF, there are many many tests that offer BRAF biomarker analysis, they all have varying levels of analytical validation available. Therefore, when one takes analytical validity, it -- I mean, it can't be determined for a biomarker, it's determined for a biomarker done

by a specific procedure is what I mean to say. So you can maybe say based on procedure, if you want to modify that.

DR. HENDERSON: One way of dealing with what you're suggesting would be not change the question, but to take the comments that you have just made, and any others you might want to add to it, and putting that into the final remarks. Remember, as we go through this

discussion now, which is a little open ended but it is preparing for the vote, then we do the voting, and then each panel member is able to add additional commentary which does go into the record and becomes part of the commentary that Medicare will take into consideration in terms of developing policy.

DR. RAMAMURTHY: That sounds good.

DR. HENDERSON: Does that solve the problem for you?

DR. RAMAMURTHY: Yes, it does.

DR. BURKE: Do I understand this question, though, as an any versus all question? In other words, does this apply to any one test versus all the tests, is that your

question?

DR. RAMAMURTHY: Right, exactly.

DR. BURKE: Yes. So the question for BRAF, can we vote on any one test being highly analytically valid, versus if all the tests have to be analytically valid, that's the question.

DR. SALIVE: So, I think it's actually a well written question and I think it asks about the evidence, and so some of the evidence

we saw was from CAP that told us how the tests would vary, which is what you're talking about. And I do agree that it should be discussed in question 4 that this, you know, the specific comments on each test have to be put there, because that is a key issue for question 1.

I mean, you know, but here we are in this state of play as it was told to us by the speaker from CAP that, you know, we don't have a standardized test, so the evidence, you know, does pertain to that, and we can still have confidence on some level on this question based on what we were presented and what we know.

DR. FISCHER: I wonder whether you

really want to separate carcinoma of the colon and carcinoma of the rectum, because they're basically different diseases as far as when patients are treated what you have to look out for, whether or not there will be continued function of the rectum or whether they will have to have a colostomy, and I think that's an important difference, and you may not want to go into that detail.

DR. HENDERSON: Well, the question here is, I'm directing this to the panel and

trying to take your thinking one step further, although I said there are two questions here. Why don't we come back to this?

Let's finish this first one. And so, we have the possibility of using comments. I was just wondering, though, if you were to change the wording of the question whether it would help any, Dr. Ramamurthy, if we add something like this, for each prognostic test, and put in parentheses, assuming the most commonly used test, or the most commonly used assay. Would that help at all?

DR. RAMAMURTHY: That's difficult to

qualify, quantify and verify. I mean, it's really difficult to say which is the most commonly used test.

DR. HENDERSON: Okay. I mean, it's not about saying an FDA test because in many cases there isn't an FDA-approved test, as we've heard many times this morning, and so that's why I was substituting, because I presume that most of the tech assessment was based predominantly on one or two tests. Maybe you could say --

DR. RAMAMURTHY: All the tech

assessment that was done in the presentation we heard this morning, were from FDA-approved and cleared kits, and separately was added to the comment that the -- and also some laboratory-developed tests, and added to the tech assessment, that they met the criteria set by CAP. I'm satisfied with including the comment in the voting.

DR. HENDERSON: Dr. Scheuner, did you have something to add here?

DR. SCHEUNER: Not to this issue, but --

DR. HENDERSON: No, let's talk about this issue, I just want to resolve it.

DR. BURKE: I'm not resolved. I don't understand --

DR. HENDERSON: No, Dr. Zuckerman was next, and then we'll get back to you.

DR. ZUCKERMAN: This is a different question, and that is --

DR. HENDERSON: I want to focus on this question and get it resolved.

DR. ZUCKERMAN: I mean, I think it's a followup to that, I think it is, and that is the extent to which -- I mean, I understand

that we have an individual choice of how to use the information that was provided in public comment, but because it wasn't, I gather, vetted by CMS, I'm trying to get a sense of how we're supposed to use data that was presented during public comment but that wasn't officially part of, you know, the vetted material.

DR. HENDERSON: Well, as I understand it, and Dr. Rollins can correct me if I'm mistaken here, but as I understand it we vote

on these questions, but then there is time in the last section after the vote for commentary, and as I understand this, you record all this, it becomes part of the record, and therefore it becomes our official communication of this panel to CMS in terms of our reservations. So we've made comments, I think it has been made here, what Dr. Ramamurthy has suggested, and I'm trying to summarize it, is we have to make a qualifier about question 1(a), so we're going to vote on it with the wording as is, but the comments he made which have already been recorded will go into the record as qualifying our vote, and there are a number of you who

seem to agree with Dr. Ramamurthy, I don't know that we need to take a formal vote on it, but a number of you agreed with it, so that would become somewhat authoritative in terms of our collective recommendations to CMS.

DR. BURKE: What is this addendum? It's assuming that there is a single test, that's what the question assumes, so getting back to the point again, am I voting on the constellation of tests? In other words, am I

saying well, given all the tests for KRAS, I think that they're all valid, or are we picking -- because the question is for each prognostic test, okay? And so this is one, but there are more than one, so am I voting for the constellation of tests for KRAS?

DR. HENDERSON: Well, as I understood the discussion thus far we are saying you are voting for the constellation, but we're putting a footnote on that vote, and the footnote is we're voting on the constellation but we have to recognize that it may not apply appropriately to every test across the board, so we're providing a general comment about putting that qualifier on it. Okay?

DR. BURKE: Okay, I got it, thank you.

DR. HENDERSON: So now let's go on before we come back, Maren, to your question, let's finish on the colorectal issue that you raised. Now, does anybody have further comments on this issue? I must say practically, I'm not certain how we could look at colon and rectum separately. In other words, we might have to deal with this the same

way we did the last one without the panel, or most members of the panel having looked at the data separately for colon and rectum, even though this may be appropriate, is it not reasonable that we again make this a footnote, if you will, to the vote?

DR. ROLLINS: Yes. Another point is, even though there may be a difference between colon cancer and rectum cancer, the way we have our questions set up, as well as the response, we could not accommodate splitting the two, so as I said, because of the way that, in terms of quantifying the responses, I mean, if we had known this weeks ago we probably could have set them up separately, but we cannot at the present time split them up.

DR. HENDERSON: I suspect you couldn't even if you wanted to, because I think that you would lose so much statistical power, particularly looking at rectum, most of the data is probably colon, not rectum, to begin with, and so you might make a conclusion that it was colon and nothing for rectum and just say there's not sufficient data.

DR. BURKE: I take your point, that the data is probably mostly on colon, probably very little on rectum, so this probably doesn't even apply to rectum at all, so one idea is just to limit it to colon cancer.

DR. HENDERSON: Since we don't know that, though, wouldn't it be more appropriate to use this as a footnote qualifier, and say we recommend you go back and look at this more carefully, rather than limit it to the colon?

Okay, Maren, you had a question, or a comment?

DR. SCHEUNER: I guess it's more of a comment. It's just the way the question is phrased.

DR. HENDERSON: Now which question are you talking about?

DR. SCHEUNER: 1(a) and 1(b), just the first phrase, for each prognostic test. So, I think this morning we've heard that many people don't consider these tests as prognostic tests, so perhaps, and maybe this isn't a necessary adjustment, but the way item 2 reads, where it says how confident are you, blah, blah, blah,

that using a molecular pathology test to estimate prognosis, and then it goes on and on. I just prefer the wording of 2 and 3, because it's kind of saying maybe these aren't, the intended use is not a prognostic test, but for some reason we're considering it as a prognostic test, so, does anyone understand what I'm trying to say?

DR. RAMAMURTHY: Well, not -- well, I mean, at least one of the tests does have a clear prognostic claim, so you cannot say all of them don't have prognostic value.

DR. BERGER: So take the word prognostic out.

DR. SCHEUNER: That's what I was wondering.

DR. HENDERSON: I'm sorry?

DR. BERGER: Just take the word

prognostic out and leave it, for each test.

DR. HENDERSON: Oh, for each test, remove the word prognostic.

DR. ROLLINS: We think that's a reasonable request. If that helps to answer the question, that's what we'll do.

DR. HENDERSON: Are you comfortable with that? I agree exactly with what you're saying, and this may be a small fix or a partial fix, but do you think that would help?

DR. SCHEUNER: Perhaps, you know, in history. I don't know, because I just want to respect people's opinions that not all of these tests, their intended use is as a prognostic test.

DR. HENDERSON: Is there anyone on the panel who is uncomfortable with removing prognostic from --

DR. ROLLINS: We could say for each of the molecular pathology tests listed above.

DR. HENDERSON: Okay. That would apply to 1(a) and 1(b), for both of those we'll use molecular pathology tests.

DR. BURKE: You have prognosis twice in each of these and they're consistent, so you

have to be consistent.

DR. ROLLINS: We should be able to keep the second prognosis in the sentence because we're asking you about this particular molecular pathology test being able to estimate

prognosis, so the second prognosis should stay.

(Inaudible colloquy among panelists.)

DR. HENDERSON: We are acknowledging the point that Maren made, though, which is the tests are not indeed prognostic tests, these are --

DR. SCHEUNER: They may, they may not be.

DR. HENDERSON: In fact, probably only one of the three were approved, but I think there are, most were developed ostensibly as prognostic tests, although knowing the developers, they had other things in mind long term, but that's the way the studies were set up, but there are others, for example, ALK, we know was set up specifically for an opposite purpose, so that's why I am not calling them prognostic tests. So, anybody opposed to removing the word, changing the phrase? Yes.

DR. RAMAMURTHY: Just to be clear,

it's only the first word, the first prognostic?

DR. HENDERSON: We're changing prognostic test to molecular pathology test in both 1(a) and 1(b).

DR. RAMAMURTHY: That's right.

DR. HENDERSON: Okay. So, Dr. Berger, are you okay?

Now, I think we've covered the ones thus far. Does anybody want to raise a refinement in any of the other -- yes?

DR. KAMRAVA: I'm just a little confused between question 2 and question 3, because it seems like for question 3 it says does it have clinical utility, and then meaning that it improves health outcomes, but that seems like that could also fall into 2, in that it affects health outcomes, including a benefit or a harm, where the benefit could be that it improves their health outcome. I guess, can someone clarify what is really the difference between the two, because there seems to be an overlap.

DR. ROLLINS: Yes. The difference between 2 and 3, in 2 the question is how confident are you that these tests affect, and

affect can be a positive effect or a negative effect, whether or not it's positive or negative effect, it still affects. So from

that, then question 3 looks in the direction of a positive effect by demonstrating that it improves the population. Sort of like a two-tailed test, it can affect it on both sides, but the one-tail test you're moving in a certain direction, so question 3 is more indicative of a one-tail, whereas question 2 is more indicative of the two-tail test, the fact is that it affects, it can be positive or negative.

DR. BURKE: So 1(b) asked is it an accurate test, in other words, does it do what it says it's going to do, and then 2 asks well, if you use this accurate test, how, what would it affect, and number 3 is if it affects something, does it have improvement.

DR. ROLLINS: Right, that's a way of looking at it.

DR. BURKE: All right.

DR. RAMAMURTHY: In terms of 1(b) there is still one more qualifier that I would add to finish off 1(a) and 1(b) if you don't

mind.

DR. HENDERSON: Okay. So let's

finish, though, these questions on 2 and 3.

DR. RAMAMURTHY: Okay.

DR. HENDERSON: The question that Dr. Kamrava raised, which is the differences between them. So the way I think about it is, from a practical point of view, item number 2, if for example this would cause a physician to change a therapy, we don't know if that's better or worse but we do know there are a number of studies, particularly with Oncotype, showing that in fact people have changed the frequency with which they give adjuvant chemotherapy. I think that was mentioned with a number of other studies, it wasn't universally covered, but there were places where people said this has changed the pattern of therapies given. So the answer is if such data exists, that would be, the answer to 2 would be yes, it does.

But question number 3, it doesn't necessarily mean that they're better off getting less chemotherapy, so you would say we don't know whether that improves outcomes. Is

that fair? That's why the questions are

separate.

DR. GRANT: Can I just clarify that? The way I interpret it is that question 2 in the framework of a diagnostic or prognostic test is really, I think, asks about discrimination, maybe a little about calibration, sensitivity, specificity, those kinds of things, does it really have some specificity to screen people who do and don't have the outcome, but who would ultimately experience an outcome.

Now, question 3 asks, yes, it discriminates, but does it define the thresholds appropriately so that one can use them to make a decision? So in other words, you can say for example that you can define low risk of recurrence for breast cancer. Now, you can discriminate between a woman who's at 15 percent or less risk and it may do that quite well, but it doesn't do it sufficiently to change decision-making in such a way that the net benefit would in fact be improved.

DR. HENDERSON: Okay. Dr. Burke, do you want to comment on that, because you were

shaking your head here?

DR. BURKE: So I think that really 1(b) is your 2, 1(b) is the sensitivity/ specificity, you know, the discrimination, you know, how accurate, how clinically accurate is this test, okay? And then 2 says what's the utility of the test, in other words, if you use the test perhaps you could discriminate or calibrate the test, you know, does it actually affect anything in the world, and then the other one is, well, if they do use it, is there really an improvement?

DR. HENDERSON: But would you agree that the use of utility, I mean, we're getting so -- and you've done a good job in your e-mail that you sent around about talking about the definitions of these things, but doesn't 3 really apply to clinical utility in the classic sense as you defined it earlier, which is improving net health outcomes? If we say something has clinical utility, we're assuming that a formal evaluation has been done and that it improves net health outcomes. Isn't that part of the definition?

So number 2 doesn't necessarily imply

clinical utility.

DR. BURKE: Well, yeah, they split clinical utility into 2 and 3. Well, first I think they're just asking about use, do docs use this test in their practice to change their management; is that right?

DR. ROLLINS: For all practical purposes question 3 specifically looks at the clinical utility demonstrating that the management that resulted from such, I'm sorry, the management that took place as a result of the test resulted in some type of improvement from the patient's perspective, so 3 is going to look specifically at clinical utility.

2 looks at whether or not there is a change, a measurable change, it could be a positive change or it could be a negative change. That's why question 3 looks in the direction of improvement.

DR. BURKE: So what is 2 again? I mean, what kind of change are you looking for, are you looking for a change in management or are you looking for a change in outcomes?

DR. ROLLINS: It could be a harm or it could be a benefit. A harm would be a negative

outcome, a benefit might be a positive outcome. Then you go further from 2 to 3 and demonstrate the improved outcome from the patient's perspective as a result of the change in management that resulted from the results of the test.

DR. BURKE: So the physician's actual management of the patient is not relevant here, because these are actually outcomes in 2 and 3?

DR. HENDERSON: Well, there are -- in this as written here, 2 implies an outcome. In the discussion somewhere, I've forgotten where it was, maybe it was in the discussions this morning or somewhere in the discussion of question 2, changes in practice was considered one of the outcomes, but you don't specify that in the question.

DR. ROLLINS: No.

DR. HENDERSON: Do you think it would be helpful if we did?

DR. ROLLINS: I think that primarily our outcome would be from a patient's perspective, whether or not there was an improvement in condition. There may be a secondary outcome such as change in management,

but the primary outcome that we would be looking for is whether or not a patient's condition improves based on --

DR. HENDERSON: That's 3.

DR. ROLLINS: Yeah, that's 3.

DR. HENDERSON: But we're talking about question number 2, though, where at one point in our discussion, and I've forgotten which presenter or who it was, but somebody made the comment very specifically in presenting the questions, that changes in management, whether they were good or bad, whether we knew they were good or bad, was one piece of question number 2 but not question number 3.

DR. ROLLINS: Correct.

DR. HENDERSON: Would it be helpful to put specific wording in here, because the way it's written right now, the differences between question 2 and question 3 are a little confusing, and I'm going to say it was only because of bolding affects and improves that I finally came to grips with this. Having read over this set of questions four different times

25 in four different presentations it finally made

1 sense to me, but it did not make sense to me
2 the first time I read it, and I think other
3 people are having the same problem.
4 DR. ROLLINS: Often when we write an
5 NCD we often reference Fryback and Thornbury
6 and if you talk about the levels, level five
7 basically talks about an improvement from the
8 patient's perspective. Level four is going to
9 be a change in physician management based on a
10 particular test. Hopefully the change in
11 management is going to result in an improved
12 outcome, so that would be equivalent to
13 number 3, which is equivalent to number five on
14 a Fryback Thornbury criteria.
15 DR. BURKE: So what's 2?
16 DR. ROLLINS: Question number 2 is
17 whether or not the test results in some type of
18 action which the physician can change
19 management.
20 DR. HENDERSON: I want to suggest some
21 specific wording because I think that's the
22 easiest way to focus our discussion, starting
23 with you, Dr. Rollins, to see if you're

comfortable with this. Question 2 now, not question 3, so question 2 where you have the

parenthetical phrase, it says affects health outcomes, parentheses, including, and then I'm going to put in that parenthetical phrase, including change in management that might lead to benefits and harms.

DR. ROLLINS: That's reasonable.

DR. HENDERSON: So the phrase, change in management that might lead to, is acceptable to you?

DR. ROLLINS: Yes.

DR. HENDERSON: What about other members of the panel? Do you find that that would help clarify what we're talking about and distinguish what we're voting on in question 2 versus question 3?

DR. BURKE: But why would you put it in parentheses?

DR. HENDERSON: It's already in parentheses.

DR. BURKE: Yeah, but why don't you just say health outcomes and/or change in management?

DR. HENDERSON: I was just trying to get something that was as simple as possible.

DR. BURKE: Yeah, but see, the

parentheses is preceded by outcomes, not by management.

DR. HENDERSON: Okay. Dr. Guadagnolo.

DR. GUADAGNOLO: So, that makes sense assuming that survival and recurrence risk and all of those outcomes are wholly encompassed in 1(b). Are we assuming that all of that goes into 1(b) in terms of clinical utility, is that the assumption?

DR. HENDERSON: I'm not --

DR. GUADAGNOLO: Because, I mean, health outcomes are cancer recurrence, you know, as well.

DR. HENDERSON: Right. I thought that was really covered more in 3.

DR. BURKE: No, no. 1(b) is sensitive, how accurate is the test.

DR. HENDERSON: Right.

DR. BURKE: Simply how accurate the test is. You have to establish accuracy at some point.

DR. GUADAGNOLO: Right. That's 1(a), correct?

DR. BURKE: No.

DR. GUADAGNOLO: So you're saying 1(b)

is --

DR. BURKE: Clinical validity.

DR. GUADAGNOLO: Clinical validity, so then -- but then health outcomes in 2, which would then include survival, recurrence risk, cancer-free survival, sometimes I think it's dependent on physician management.

DR. BURKE: Of course it is, but 2 is different than 1(b) because 2 is how you actually use the test, so the test has a certain accuracy, but then how you use the test can many times be different than the actual, you know, clinical validity.

DR. GUADAGNOLO: Right, but it should include both, though, it shouldn't just be management-related health outcomes, but actual natural history of disease-related outcomes.

DR. SCHEUNER: I think that's encompassed by the phrase that precedes it, where it says that using the molecular

pathology test to estimate prognosis affects health outcomes, blah, blah, blah, so maybe, does that help? Because again, it brings us back to dealing with this prognostic issue and not the predictive, we're focusing only on the

prognostic component.

DR. HENDERSON: Okay. So the fact, I mean, again, the classic example is you find out your patient has a very high risk, you're more likely to put them through a toxic therapy. If you find they have low risk of recurrence and you put them through, you're going to withhold that toxic therapy, and that's the point you're making, right?

DR. SCHEUNER: Yeah.

DR. HENDERSON: But that differs from, say you find out that this particular test tells you you're going to respond to this therapy, you're using it because the therapy is going to work, that's a predictive test.

DR. RAMAMURTHY: Right, so I think some of these issues are being conflated here. I want to come back to 1(b), make a quick qualifier to add to the bottom of the comments

list. When you say clinical validity of these various biomarkers, these are gene names, and these genes have mutations at various sites, not all of the mutations enjoy an equal level of clinical validity, and therefore the qualifier should be for proven clinical

validity somewhere in the comments section, because --

DR. HENDERSON: Now you're talking about 1(b) specifically.

DR. RAMAMURTHY: Yeah, I just wanted to add the qualifier I was waiting on.

DR. HENDERSON: And what specific wording do you want to insert here?

DR. RAMAMURTHY: That not all biomarkers associated with given genes, by biomarkers I mean the mutations or variations, not all variations for each of those genes have completely proven clinical validity, they have varying levels of validity.

DR. BURKE: Varying levels of accuracy.

DR. RAMAMURTHY: Right.

DR. HENDERSON: But what you're doing

is making a comment.

DR. RAMAMURTHY: Yeah.

DR. HENDERSON: What I'm trying to do, the comments will go into the record if there's a comment, but what I'm trying to get at is whether you want to add a comment like we did before with your comments on 1(a), whether you

want to add a comment or whether you want specific word changes. If you want specific word changes, I would strongly encourage you to keep them to a few words, three, four, five words, rather than three or four sentences.

DR. RAMAMURTHY: No, I'm satisfied with a comment at the end just for the record to reflect that this is not a monolithic quantity.

DR. HENDERSON: All right, that is a comment that we will add as a proviso.

DR. RAMAMURTHY: Right, and I want to quickly then jump onto 2, because you have done something within 1(b) and 2 here that is very related. So in question 2, your suggested language that's still under consideration says including change in management and benefits and

harms thereof. And that language you just discussed hasn't been fully voted on yet, or discussed yet. And there what you have done, if I'm to understand, is even apportioning therapy based on the mutation status would be considered a change in management, and therefore you are roping in potentially the predictive capabilities of said biomarker in

the estimating prognosis of overall health outcomes, if I understand correctly.

So in a way you are conflating the predictive nature of the biomarker and using that to discuss overall prognostic value vis-a-vis health outcomes. Did I understand that correctly?

DR. HENDERSON: Well, it doesn't seem to me that you do. Now -- because this is why what Maren added was so important. I realize that the exercise that we're going through here is somewhat artificial, a lot of people have commented on that. I mean, a lot of people on the panel have indicated that they're rather uncomfortable with this issue between prognostic and predictive, but I think Maren

made the point here that what we're doing with question number 2 is saying having prognostic information might affect treatment outcome, independent of any predictive information, and it's my understanding that is what CMS is asking us to look at.

DR. ROLLINS: Right, the prognostic. Now it's possible that in the discussion section you might say something like it's very

difficult to separate the predictive ability of the test from the prognostic ability of the test, and this is in the discussion section, and make that point very clearly. But in terms of voting, we're looking at the prognostic ability of the test.

DR. HENDERSON: Okay. Dr. Kamrava.

DR. KAMRAVA: Is it possible in question 2 for the word prognosis actually to be taken out, so that in question 1(b) you can actually have an assessment of what we believe the impact on the prognosis is, but then 2 actually becomes how is it actually being used. It may be used as a predictive, it may be used as a prognostic, but I think if you take the

word prognosis, so, you know, how confident are you, blah, blah, the molecular pathology test to affect health outcomes, and take prognosis out.

DR. HENDERSON: In question 1(b)?

DR. KAMRAVA: In question 2, and leaving 1(b) as is, and using that as a true assessment of prognosis.

DR. GRANT: Could I just make a comment and refer back to the TA, which I

think, you know, one way to consider 2 is it's really about the chain of evidence. I mean, prognosis, a test doesn't do anything directly to health outcomes, and so what the TA did point out, which is the case for most diagnostic tests absent some randomized controlled trial that definitively answers the question, is that it links together these pieces of evidence. And I think part of this is just directed at, is there a chain of evidence that's plausible and likely to occur in practice when you have a prognostic test and make management decisions, those management decisions have been shown to affect the balance

of benefits and harms ultimately leading to then getting to the question, is it positive? But that's essentially, I think, the way I would look at it.

DR. HENDERSON: You're talking about the ACCE, is that what you called it before, A-C-C-E?

DR. GRANT: I'm talking about the chain of evidence. I'm talking about the chain of evidence, you know, there was that nice little analytic diagram there, you do this, you

do that, you get a test, you do this, you do that, you're going to do all those things. So the question is, number 3, when you do all that, in other words, are those in fact connected, are you really going to make a decision, and we've heard commenters say no, it's not going to make a change in the decision-making, so that answers that question. And for 3 it's when you do make those decisions, and you don't treat, you know, with adjuvant chemotherapy alone and it lowers the recurrence, yes, that benefit is improved.

DR. HENDERSON: Okay. So now, we'll

have an opportunity to come back for further discussion and clauses, comments, provisos after the vote, but are there other things where somebody wants to specifically change the wording of a question for some reason or another? Mark.

DR. GRANT: I have one other comment. It's that in the technology assessment, outside of some data on analytic validity, I did not see evidence presented regarding prognosis for MLH1 testing or ALK testing. So given the fact that there's no evidence presented for us to

appraise, I think it's sort of hard, absent unless we're going to use our own expert opinions, to vote on those.

DR. HENDERSON: So, why wouldn't it be just as reasonable to give it a low vote, a low confidence, which is the same as saying you don't believe there's evidence?

DR. GRANT: Absence of evidence isn't the same, so there's that argument.

DR. HENDERSON: Well, there are two choices we have here. One is to vote one, which is the lowest vote you can give which

basically, that takes it out, we don't discuss anything beyond that. Well, that would come on question, you would vote, 1(a) would be whatever, it might be three, four or five, but question 1(b) would be one, and that ends the discussion, we wouldn't go on to questions 2 and 3, right?

DR. ROLLINS: Right.

DR. HENDERSON: So rather than eliminating this, is there some reason, some value to eliminating it in terms of the message we send to CMS, compared to just voting a one? Don't you achieve the same thing? Does anybody

else have comments on this?

UNIDENTIFIED PANELIST: Yeah, it's better to send a message to CMS by voting than by deleting the question.

DR. HENDERSON: Okay. Anybody else have a comment? Okay. Any other specific changes or comments? Okay. So then, before we go on to vote, any other discussion that would likely affect, or that you think is very important in affecting a vote? Usually when we vote, do you usually have a discussion before

each question, or do we go right to the voting?

DR. ROLLINS: Go right to the voting.

DR. HENDERSON: Okay. So now is the time if you have something in the discussion on any of these that you think will potentially influence your fellow panel members or whatever that you want to make prior to the vote, you make it now, and then provisos that you want to make afterwards, qualifiers like we've already heard in the discussion, that each panelist will be asked to make after the vote. Yes, Dr. Zuckerman, did I not make myself clear?

DR. ZUCKERMAN: Yes, you did. I guess I just want to clarify, and I guess I don't

want to do this necessarily in the wording of the question, but that we're looking at this chain of events as has been described without knowing in terms of evidence how treatment was affected. So we're saying there's this correlation between having a test and outcome, if there is one, but we don't know to what extent that's because people were receiving a treatment that was or wasn't effective, and in addition to that we have very limited, if any,

information about complications from the testing. We have evidence regarding recurrence, progression-free survival and overall survival, but we don't have -- I just want to clarify. I don't believe we have any information about other kinds of complications that might have occurred, other kinds of harms. Am I correct?

DR. HENDERSON: Yes, I think. Again, I think that's why the split between question 2 and question 3.

DR. ROLLINS: It might be difficult to include all the possibilities, for that reason we just used some examples, but that does not mean that you need to restrict yourself to only

the examples that we gave, because in the discussion section you can go into more detail explaining your numbers.

DR. HENDERSON: Okay. Any other discussion before votes? So with that, we'll proceed with the voting.

So what we're going to do here is, there's several different ways we could do this, but I've decided that we're first of all

going to vote on question 1(a) for all the tests. So question 1(a) is, for each molecular pathology test listed above, how confident are you that existing evidence is sufficient to confirm the analytical validity of the molecular pathology test to estimate prognosis for Medicare beneficiaries with that cancer type?

So we're really only talking about analytical validity here, plain and simple, so we're going to later talk about prognosis, but that's not really the operative part of this question as I understand it. The operative part of this question is when you do the test, is it reproducible and does it give you an answer to the question that it was designed to

answer.

DR. BURKE: You mean is it reproducible.

DR. HENDERSON: So you're saying analytic validity is totally reproducible.

DR. BURKE: Yes.

DR. HENDERSON: Okay. Dr. Burke just clarified the comment that with the evidence

here, is it reproducible? In other words, if the labs do it in the lab, different labs repeatedly, will you get the same result from the same specimen? Okay. So, does everybody have their voting machine? And you indicated you want them to speak it out loud as well as a vote?

MS. ELLIS: Yes. So what we're going to do is, all the panel members that have the electronic voting devices, when it's time to vote you will push the number to correspond with your vote. After all of those have been tallied it will show up on the screen. What we also need is, for the record and for the web, we need each individual panel member, including the nonvoting panel members, to state your vote as well. Okay? And don't forget also, in your

folders there is a green score sheet that you will fill out as well and turn in to me. So, we can begin voting.

DR. BURKE: So, are we voting for BRAF?

MS. ELLIS: We're going to start at number one, BRAF.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: So our first one, then, for question 1(a), is BRAF for colon cancer. Has everybody voted? I think when we go through these, I'll ask people to give their votes out loud.

MS. ELLIS: You need to do each one as we go along. We need one more person, someone is missing.

DR. HENDERSON: One person hasn't voted, so you can't cheat, you don't get away with that.

MS. ELLIS: Could everyone push the button one more time just to make sure that your vote went in.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: If you can, can you just take your device out of the sleeve to make sure, and just, everyone push the button one more time? We're just waiting for one more vote to register. Thank you.

DR. HENDERSON: Let's start at that

end, Dr. Berger.

DR. BERGER: Based on the CAP data it sounds very reasonable to me, and I would give it a five.

DR. HENDERSON: So what's your number? So, five. Oh, I'm sorry, we're supposed to start at this end, so Dr. Burke.

DR. BURKE: Two.

DR. FISCHER: Two.

DR. GRANT: Four.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: Five.

DR. SCHEUNER: Four.

DR. ZUCKERMAN: Four.

DR. RAMAMURTHY: One.

DR. BERGER: Five.

DR. HENDERSON: So then, the KRAS, so start with Dr. Burke. Oh, we have to --

MS. ELLIS: Yes.

DR. BURKE: Two again. Oh, I thought you were ready.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: Okay, so we have the votes for question 1. You can go ahead and start getting the votes.

DR. HENDERSON: Okay. So now we're voting for KRAS, so now the individual votes for KRAS, Dr. Burke.

DR. BURKE: KRAS, two.

DR. FISCHER: Two.

DR. GRANT: Four.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: Four.

DR. SCHEUNER: Five.

DR. ZUCKERMAN: Four.

DR. RAMAMURTHY: Four.

DR. BERGER: Four.

DR. HENDERSON: Okay. So now the next one, again, the same question, this is analytical validity for MSI, or microsatellite instability. So first of all, vote on your

machine.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: Okay, Dr. Burke.

DR. BURKE: For microsatellite instability, three.

DR. FISCHER: Three.

DR. GRANT: Four.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: Four.

DR. SCHEUNER: Five.

DR. ZUCKERMAN: Three.

DR. RAMAMURTHY: Four.

MS. ELLIS: If you guys could please speak into the mic, as it is being webcast, and we do have a transcriptionist.

DR. HENDERSON: I think, Dr. Ramamurthy, you were a four?

DR. RAMAMURTHY: Four.

DR. BERGER: Four.

DR. HENDERSON: Now again, question 1(a), analytic validity for MLH1 promoter methylation. Vote on your machines.

(The panel voted and votes were

recorded by staff.)

MS. ELLIS: Okay, we have everyone's electronic votes.

DR. HENDERSON: Okay, Dr. Burke.

DR. BURKE: Two.

DR. FISCHER: Two.

DR. GRANT: Four.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: One.

DR. SCHEUNER: One, because I didn't see any evidence.

DR. ZUCKERMAN: One.

DR. RAMAMURTHY: I have two.

DR. BERGER: Two.

DR. HENDERSON: And finally, cancer of the colon, Oncotype Dx. Vote with your machines first.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: We have all the electronic votes.

DR. HENDERSON: Okay, Dr. Burke.

DR. BURKE: Two.

DR. FISCHER: Three.

DR. GRANT: Four.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SCHEUNER: Three.

DR. ZUCKERMAN: One.

DR. RAMAMURTHY: Two.

DR. BERGER: Four.

MS. ELLIS: Okay, we're ready for the next one.

DR. HENDERSON: Okay. Again, analytic validity, question 1(a), for MammaPrint, breast cancer, with the machines first.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: There we go.

DR. HENDERSON: You got them all, okay.

DR. BURKE: For MammaPrint, three.

DR. HENDERSON: Okay, Dr. Burke, you start.

DR. BURKE: Two, or, three.

DR. HENDERSON: Oh, three, okay.

DR. FISCHER: Two.

DR. GRANT: Five.

DR. GUADAGNOLO: Three.

DR. KAMRAVA: Four.

DR. SALIVE: Five.

DR. SCHEUNER: Five.

DR. ZUCKERMAN: Two.

DR. RAMAMURTHY: I have five.

DR. BERGER: Four.

DR. HENDERSON: Okay. Now, again, question 1(a), analytic validity, breast cancer Oncotype Dx, vote with your machines.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: We are waiting on one panel member. Thank you.

DR. HENDERSON: Okay. So now, Dr. Burke?

DR. BURKE: Oncotype Dx, three.

DR. FISCHER: Five.

DR. GRANT: Five.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: Four.

DR. SCHEUNER: Five.

DR. ZUCKERMAN: Two.

DR. RAMAMURTHY: Five.

DR. BERGER: Four.

MS. ELLIS: Okay, next?

DR. HENDERSON: Okay. Now we're going on, so this is now question 1(a), analytic validity for ALK in non-small cell lung cancer. With the machines.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: We're ready.

DR. HENDERSON: Okay, Dr. Burke.

DR. BURKE: ALK, one.

DR. FISCHER: One.

DR. GRANT: Four.

DR. GUADAGNOLO: Three.

DR. KAMRAVA: Four.

DR. SALIVE: Four.

DR. SCHEUNER: Two.

DR. ZUCKERMAN: One.

DR. RAMAMURTHY: I have four.

DR. BERGER: Four.

DR. HENDERSON: Okay. Now going on to question 1(a), analytic validity, EGFR in non-small cell lung cancer. With the machines.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: We have the votes.

DR. HENDERSON: Okay, Dr. Burke?

DR. BURKE: Two.

DR. FISCHER: Two.

DR. GRANT: Four.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: One.

DR. SCHEUNER: Four.

DR. ZUCKERMAN: Three.

DR. RAMAMURTHY: Four.

DR. BERGER: Four.

DR. HENDERSON: Okay. And now the last one, question 1(a), analytic validity for KRAS in non-small cell lung cancer, with the machines.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: We have the votes.

DR. HENDERSON: Okay. And now, Dr. Burke?

DR. BURKE: Two.

DR. FISCHER: Four.

DR. GRANT: Three.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: Four.

DR. SCHEUNER: Four.

DR. ZUCKERMAN: Four.

DR. RAMAMURTHY: Four.

DR. BERGER: Four.

DR. HENDERSON: So now, I didn't notice as we were going down; were any of them under 2.5?

MS. ELLIS: Yes. The only one that was under 2.5 was MLH1.

DR. HENDERSON: What was the score?

MS. ELLIS: It was 2.38.

DR. HENDERSON: So for MLH1, we will not vote for question 1(b), 2 or 3, so that's off the books and we don't have to worry about that anymore today. Yes, Dr. Burke.

DR. BURKE: So, I just wanted to, before we vote on 1(b), because we really didn't spend any discussion time on these issues, so I wanted to just say something really quickly about this thing.

DR. HENDERSON: Okay. So, we want to keep moving the voting, so really quickly.

DR. BURKE: Like five sentences. So with hazard ratios, so what we got in the tech

report were hazard ratios, right? And so the issue is, what's the relationship between hazard ratios and predictive accuracy, and the answer is there's no necessary relationship, so you can't use hazard ratios to guide you in terms of the predictive accuracy of the test, right? So what we've got here is, we've got a set of predictors that act as a classifier, it assigns the patients to an outcome class based on the probability of that outcome. In terms of hazards, it separates the patients into groups based on their relative hazard of experiencing the event, i.e., the outcome, and the separation's tested by significance testing. But that separation doesn't provide any information about how accurate the hazard assignment is for each patient, okay? So there's no necessary reason that the particular hazard assignments are highly accurate, it just means that you can group these people into two groups, okay?

So in the CMS information, they ask for predictive accuracy, they ask for sensitivity/specificity, positive or negative

predictive value, they could have asked for ROC

or various other measures. So my problem is that I have no measures of predictive accuracy in the tech report that we can look at, and I think the reason is because very few of these markers predicting prognostic factors actually have any sensitivity/specificity whatsoever, so I took a little look at --

DR. HENDERSON: Okay, can you keep this limited?

DR. BURKE: I'm going to, I'm not going to say what they are, but I took a look at EGFR in KRAS in lung cancer and I looked at all relevant literature for the last three years, okay? And there is literally no evidence of predictive accuracy in these literature, they're just hazard ratios. So my problem is that I can't give these tests, MammaPrint and Oncotype Dx, I can't give these tests any confidence because nowhere in the literature or in the tech report was there predictive accuracy.

Secondly, for MammaPrint and Oncotype Dx, there is some literature on

predictive accuracy but it's low predictive accuracy, okay? So that's going to color my

thinking in terms of my scores.

DR. HENDERSON: Okay. So again, this would have been appropriate, and we're still technically in the discussion hour, I think maybe one or two minutes over, so I allowed you, I felt like what you had to say was important.

But on the other hand, I'm not certain that we need a more full discussion. If there's anybody who wants to make a very very brief response to what Dr. Burke has said, or a different point of view, because in practice most physicians are using hazard ratios. Regardless of the points you make, which are scientifically rigorous, it still remains a fact that most of us do look at hazard ratios and most of us have made conclusions, for example, that one or another of these tests, depending on our area of specialization and so on, one or more of these tests in fact are satisfactory in terms of saying this patient is going to recur sooner, or this patient's going

to die sooner. We do do that, and very respectable physicians who are also experts in this have accepted this, so I just wanted to

make certain that there are two sides to this.

DR. BURKE: But let me just clarify. So if I'm voting on whether physicians actually do this, I would give it a different vote than what the scientific evidence is.

DR. HENDERSON: No, I understand that point and that's certainly valid, I just want to make sure that there are both sides here. Dr. Rollins.

DR. ROLLINS: Yes. From CMS's perspective, when we've written diagnostic NCDs, we have included information about hazard ratios, which were reported in the study.

DR. BURKE: Sure, you can put anything you like in there, they're just not measures of predictive accuracy.

DR. HENDERSON: Okay. I just think we have to take the totality of evidence and different points of view into consideration.

So, unless somebody else has something that's different, we'll -- Dr. Ramamurthy?

DR. RAMAMURTHY: I just want to clarify that. Question 2 refers to the pathology tests being able to estimate prognosis, and as we are going down on 1(b) and

looking at clinical validity, that cloud still hangs over these questions about the markers having predictive capacity versus prognostic, although I just wanted to state that for the record.

DR. HENDERSON: Okay, so I think that's a very important point, it's just underscoring how it's written.

So with that, we will start with the voting, question 1(b). The question again is, for each molecular pathology test listed above, how confident are you that existing evidence is sufficient to confirm the clinical validity of these tests to estimate prognosis in Medicare beneficiaries with that cancer type?

So that's the question, and the first one is BRAF in adenocarcinoma of the colon and rectum. So vote with the machines.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: We have everyone's vote.

DR. HENDERSON: Dr. Burke?

DR. BURKE: One.

DR. FISCHER: One.

DR. GRANT: Three.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: Three.

DR. SCHEUNER: Five.

DR. ZUCKERMAN: Two.

DR. RAMAMURTHY: Two.

DR. BERGER: Four.

DR. HENDERSON: Okay, so a positive score of 2.86.

Now KRAS, again, 1(b), KRAS for adenocarcinoma of the colon and rectum, vote on the machines.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: 2.75. Dr. Burke.

DR. BURKE: One.

DR. FISCHER: Three.

DR. GRANT: Two.

DR. GUADAGNOLO: Three.

DR. KAMRAVA: Two.

DR. SALIVE: Three.

DR. SCHEUNER: Five.

DR. ZUCKERMAN: Three.

DR. RAMAMURTHY: Three.

DR. BERGER: Three.

DR. HENDERSON: Okay. Question 1(b) for MSI, microsatellite instability for adenocarcinoma of the colon and rectum, vote with the machines.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: 3.36. Dr. Burke.

DR. BURKE: Two.

DR. FISCHER: Two.

DR. GRANT: Four.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: Three.

DR. SCHEUNER: Five.

DR. ZUCKERMAN: Three.

DR. RAMAMURTHY: Four.

DR. BERGER: Five.

DR. HENDERSON: So the MLH1 we skip

over because we didn't reach a 2.5, and now we go on to Oncotype Dx of the colon, this is question 1(b), clinical validity for adenocarcinoma of the colon and rectum, with the machines.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: 1.625. Dr. Burke.

DR. BURKE: One.

DR. FISCHER: Two.

DR. GRANT: Two.

DR. GUADAGNOLO: Two.

DR. KAMRAVA: Two.

DR. SALIVE: One.

DR. SCHEUNER: Two.

DR. ZUCKERMAN: One.

DR. RAMAMURTHY: One.

DR. BERGER: One.

DR. HENDERSON: Okay. Next is clinical validity, question 1(b), for MammaPrint used in breast cancer.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: We are waiting on two

members. Could everyone push the button again, please? Okay.

DR. HENDERSON: Okay, Dr. Burke.

DR. BURKE: For MammaPrint, three.

DR. FISCHER: One.

DR. GRANT: Three.

DR. GUADAGNOLO: Three.

DR. KAMRAVA: Four.

DR. SALIVE: Three.

DR. SCHEUNER: Four.

DR. ZUCKERMAN: Three.

DR. RAMAMURTHY: Five.

DR. BERGER: Four.

DR. HENDERSON: Okay. Next is question 1(b), clinical validity for Oncotype Dx for breast cancer.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: We are waiting for one panel member. Could everyone please push the button again. Thank you.

DR. HENDERSON: So this is 3.875. Individually, Dr. Burke.

DR. BURKE: Three.

DR. FISCHER: Four.

DR. GRANT: Four.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: Four.

DR. SCHEUNER: Five.

DR. ZUCKERMAN: Three.

DR. RAMAMURTHY: Four.

DR. BERGER: Four.

DR. HENDERSON: Okay. Clinical validity for ALK in non-small cell breast cancer, or lung cancer, I mean.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: Dr. Burke?

DR. BURKE: One.

DR. FISCHER: One.

DR. GRANT: One.

DR. GUADAGNOLO: One.

DR. KAMRAVA: Two.

DR. SALIVE: Three.

DR. SCHEUNER: One.

DR. ZUCKERMAN: One.

DR. RAMAMURTHY: Two.

DR. BERGER: One.

DR. HENDERSON: Okay, so we will not discuss ALK in questions 2 or 3. Got that?

Next is the clinical validity, EGFR for non-small cell lung cancer.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: 1.5. Dr. Burke.

DR. BURKE: One.

DR. FISCHER: Three.

DR. GRANT: One.

DR. GUADAGNOLO: Two.

DR. KAMRAVA: Two.

DR. SALIVE: One.

DR. SCHEUNER: One.

DR. ZUCKERMAN: One.

DR. RAMAMURTHY: Two.

DR. BERGER: One.

DR. HENDERSON: Okay. So we won't discuss EGFR further in questions 2 or 3.

Now finally, clinical validity of KRAS for determining prognosis of patients with non-small cell lung cancer.

(The panel voted and votes were

recorded by staff.)

DR. HENDERSON: 2.375. Dr. Burke.

DR. BURKE: Two.

DR. FISCHER: Four.

DR. GRANT: One.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Two.

DR. SALIVE: Two.

DR. SCHEUNER: Three.

DR. ZUCKERMAN: One.

DR. RAMAMURTHY: Two.

DR. BERGER: Two.

DR. HENDERSON: Okay. Well, not a lot of deviation when we look at our voting there. So that is our question 1, and we will not take any of the lung cancer molecular tests forward for questions 2 and question 3.

So now we go on to question 2, how confident are you that there is sufficient evidence to conclude that using a molecular pathology test to estimate prognosis affects health outcomes, including a change in management that might lead to benefits and harms for Medicare beneficiaries whose

anti-cancer strategy is guided by the test's results?

DR. RAMAMURTHY: Mr. Chair?

DR. HENDERSON: Yes.

DR. RAMAMURTHY: Can I confirm Oncotype for colon, I don't know if Oncotype Dx for colon made the cut or not.

DR. HENDERSON: I'm sorry, what?

DR. RAMAMURTHY: Can you confirm the scores for Oncotype Dx colon? Did it make the cut in the previous round?

MS. ELLIS: No, it didn't. The score

for 1(b) was 1.63, so we will not vote on questions 3 and 2.

DR. HENDERSON: So we're only going to -- yeah, okay. So, we're only going to do the first three under colon, we won't be doing either MLH1 or Oncotype, okay?

So I won't read the question again, but just focus on affects health outcomes, not improves, affects health outcomes, question 2, for BRAF in the management of adenocarcinoma in the colon and rectum.

(The panel voted and votes were

recorded by staff.)

DR. HENDERSON: 2.375. Okay, Dr. Burke.

DR. BURKE: One.

DR. FISCHER: Two.

DR. GRANT: Three.

DR. GUADAGNOLO: Two.

DR. KAMRAVA: Three.

DR. SALIVE: Two.

DR. SCHEUNER: Three.

DR. ZUCKERMAN: Three.

DR. RAMAMURTHY: I have three.

DR. BERGER: Two.

DR. HENDERSON: Okay. The next is, again, affects health outcomes, including management, for KRAS in the treatment or management of adenocarcinoma in the colon or rectum.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: We're waiting on one panel member. Thank you.

DR. HENDERSON: 2.125. Dr. Burke.

DR. BURKE: One.

DR. FISCHER: Three.

DR. GRANT: One.

DR. GUADAGNOLO: Two.

DR. KAMRAVA: Three.

DR. SALIVE: Two.

DR. SCHEUNER: Three.

DR. ZUCKERMAN: Two.

DR. RAMAMURTHY: Four.

DR. BERGER: Three.

DR. HENDERSON: Okay. And now next is, again, affects health outcomes, question 2, for MSI or microsatellite instability, for adenocarcinoma of the colon and rectum.

(The panel voted and votes were

recorded by staff.)

DR. HENDERSON: 3.125. Dr. Burke.

DR. BURKE: Two.

DR. FISCHER: Three.

DR. GRANT: Four.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Three.

DR. SALIVE: Two.

DR. SCHEUNER: Four.

DR. ZUCKERMAN: Three.

DR. RAMAMURTHY: Four.

DR. BERGER: Four.

DR. HENDERSON: Okay. Next, affects health outcomes, question 2, for MammaPrint in the management of breast cancer.

MS. ELLIS: I just need one second so that I can get to that slide. There we go.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: 2.75. Dr. Burke.

DR. BURKE: Two.

DR. FISCHER: Two.

DR. GRANT: Three.

DR. GUADAGNOLO: Four.

DR. KAMRAVA: Four.

DR. SALIVE: Three.

DR. SCHEUNER: Two.

DR. ZUCKERMAN: Two.

DR. RAMAMURTHY: Four.

DR. BERGER: Four.

DR. HENDERSON: And next, actually this is the last one for question 2, affects health outcomes in the use of Oncotype Dx for breast cancer.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: Three, okay.

Dr. Burke.

DR. BURKE: Two.

DR. FISCHER: Four.

DR. GRANT: Four.

DR. GUADAGNOLO: Three.

DR. KAMRAVA: Four.

DR. SALIVE: One.

DR. SCHEUNER: Four.

DR. ZUCKERMAN: Two.

DR. RAMAMURTHY: Four.

DR. BERGER: Four.

DR. HENDERSON: Okay, so that's all for question 2, and we have three of them now

that have survived to address question 3. So question 3, which is the clinical utility question is, how confident are you that there is sufficient evidence to conclude that using the molecular pathology test to estimate prognosis has clinical utility, meaning that it improves health outcomes either due to increased benefits and/or reduced harms, for

Medicare beneficiaries with cancer whose anti-cancer treatment strategy is guided by the tests results? Okay.

So question 3, clinical utility of MSI in management of adenocarcinoma of the colon and rectum.

(The panel voted and votes were recorded by staff.)

MS. ELLIS: We're waiting for one more panel member. Thank you.

DR. HENDERSON: 2.12. The next one is clinical utility for --

MS. ELLIS: We need them to state the votes.

DR. HENDERSON: Oh, I'm sorry. Dr. Burke.

DR. BURKE: One.

DR. FISCHER: Two.

DR. GRANT: Three.

DR. GUADAGNOLO: Three.

DR. KAMRAVA: One.

DR. SALIVE: One.

DR. SCHEUNER: Three.

DR. ZUCKERMAN: Two.

DR. RAMAMURTHY: Three.

DR. BERGER: Three.

DR. HENDERSON: Okay. Now we'll go to question 3, clinical utility, MammaPrint for breast cancer.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: 2.25. Dr. Burke.

DR. BURKE: Two.

DR. FISCHER: One.

DR. GRANT: Two.

DR. GUADAGNOLO: Three.

DR. KAMRAVA: Two.

DR. SALIVE: Three.

DR. SCHEUNER: Two.

DR. ZUCKERMAN: Two.

DR. RAMAMURTHY: Four.

DR. BERGER: Four.

DR. HENDERSON: I think the grouping of the panel is rather interesting. Now, the last one is clinical utility of Oncotype Dx for breast cancer.

(The panel voted and votes were recorded by staff.)

DR. HENDERSON: 2.875. Dr. Burke.

DR. BURKE: Two.

DR. FISCHER: Four.

DR. GRANT: Three.

DR. GUADAGNOLO: Three.

DR. KAMRAVA: Two.

DR. SALIVE: Four.

DR. SCHEUNER: Three.

DR. ZUCKERMAN: Two.

DR. RAMAMURTHY: Four.

DR. BERGER: Four.

DR. HENDERSON: So those are the first three questions. Now we go on to question 4, and we don't have to have a vote here, but these are discussion questions. So the main thing here is to give commentary that will be helpful, be of help, I mean, to CMS. So again, we want to keep this, each one of the discussions very very focused, and so they

could be kind of yes or no type answers or qualifying answers.

So the first one is, please discuss whether each factor below might change the generalizability of evidence about prognostic

molecular pathology tests in Medicare beneficiaries with cancer. So, in a way we've already voted on the generalizability of the test so this is the place where, you know, you might modify your vote depending upon the answer that you have to each of these questions. So the first one is regulatory status of the test, for example, FDA approval or clearance versus laboratory-developed test.

So this gets to part of the question that one of you was trying to get at, or several of you were trying to get at before, about all tests were not equal. So, would it make a difference to you in terms of the questions whether this was FDA or laboratory-based, FDA or CLIA. It's kind of, I think the distinction is probably important because CLIA is not technically what you mean by FDA-approved. FDA-approved means they've gone through a formal process, so to give an

example, MammaPrint is FDA-approved, Oncotype is a CLIA approval, or not approval, but a CLIA test, it was developed with CLIA guidelines. So, any comments people have on this?

Dr. Ramamurthy.

DR. RAMAMURTHY: I might have a small predictable comment on this, I'll try not to be predictable. Certainly the regulatory status of a test gives us some confidence that it went through a premarket review process, number one. Number two, that if there are potential problems with the test in how it's being applied in the laboratory setting, then there is recourse in terms of being able to identify problems for the manufacturer to issue recalls, adverse events to be tracked and so on and so forth. It's like when you buy a car, if you register with the dealer, then they know if something went wrong with a headlight or windshield wiper then you get a little postcard in the mail and it says you should go to your dealer and get your car looked at.

Having said that, I completely appreciate that sometimes patients cannot wait for very long and that if there is a

preponderance of evidence that the test is being used appropriately, as well as data, then I guess from the physician's point of view,

they might be feeling confident that this particular test is applicable in that kind of setting for treatment of the patient. That's all I have to say.

DR. HENDERSON: So you're saying you would look more favorably at an FDA-approved test generally?

DR. RAMAMURTHY: In general it's a known quantity. There is no impugning of one versus the other, there are good FDA-approved tests, there are bad FDA-approved tests, and there are very good LBTs and there are LBTs that we don't know how good or not they are. But nevertheless, it's about the known versus the unknown.

DR. HENDERSON: Anybody else want to make a comment? Yes.

DR. FISCHER: I just have a question. Is there any legal basis for one approval versus another? In other words, let's take the FDA, which we have not talked about much. But is there a legal basis that FDA approval means

something?

DR. RAMAMURTHY: Right, I mean --

sorry, did I jump back too fast? Yes, there are specific regulations that prohibit marketing of tests without having obtained regulatory clearance or approval, and for that matter even companion diagnostics when they are on the package insert or indicated for a drug, will say the drug, will say drug so-and-so should be given to patients who have been identified with that biomarker by an FDA-approved test, so there are other legal implications, yes.

DR. HENDERSON: I want to clarify that a little bit. Dr. Rollins, you might want to make a comment as to how much --

DR. ROLLINS: The only thing that I was going to say is FDA's mantra is if it's approved, it's safe and effective. Now, can safe and effective be applied to diagnostic tests? I guess, I'm not sure.

UNIDENTIFIED PANELIST: A 510(k) is not safe and effective.

DR. HENDERSON: Okay. The point I want to clarify here is that these tests come

under devices, not under drugs as a general

rule, and I think most of us are much more comfortable in understanding the FDA process for drugs. Now unless there's accelerated approval, most drugs will usually require some evidence of clinical utility, and they require that it has a proved patient outcome in one way or another, pretty much the way you'd define it today. However, they don't require that for these tests or for devices in the same way.

So devices, you know, there's a lot more different regulations. For example, you can show that they're much more likely to accept equivalency to something that has already even been grandfathered, where drugs very rarely do that today. So there's a lot of difference in looking at a drug that has been approved by the FDA and a device, any device including these tests, that has been approved by the FDA. So, I just think it's important to keep that in mind. Joe, and then we'll move on.

DR. FISCHER: I suppose my question was, is approval by the FDA under those circumstances a very good C minus, or does it

really have the weight of the other approval? In other words, is that the second best way of getting something approved?

DR. HENDERSON: Now, wait, I'm not certain that I understand your question, the other approval you were talking about.

DR. FISCHER: In other words, we have ways of approving things, okay? Is there a single way of approving things or are there two ways of approving things under certain circumstances?

DR. HENDERSON: Well, in a way you could say there are three ways, okay? There's the way drugs are approved in general, although even there there are several subcategories, there's accelerated approvals, there's full approvals, and they vary quite a bit actually in terms of tumor type to tumor type although they follow the same principle.

In terms of devices, there may be a full approval called a PMA, or there may be, what's the other one?

DR. RAMAMURTHY: It could be 510(k).

DR. HENDERSON: 510(k) is another one, which is relatively a much easier and less

rigorous way of getting it on the market. And so devices fall under both of these categories, and as I explained before, you do have some evidence of clinical utility with drugs for accelerated approval, but you may not have that -- well, even with accelerated approval they require some evidence of clinical utility, it's a little bit shakier, but for devices you may not have that.

The other point I wanted to make which Dr. Ramamurthy referred to, which I think is important to keep in our minds, and that is that when a test now is approved, a drug is approved and it's linked, it's my impression, although I can't say I'm an expert on that, I went through the approval for the HER2 test, for example, but I get the impression that the FDA approval process is more vigorous there. If somebody wants to argue with that, I would certainly be open to it, but I think you have to distinguish the different ways in which a test is approved, and I would almost count that as a third way.

And somebody, maybe it was Dr. Ross, referred to the process, particularly with the

Dako HercepTest at the time it was approved in 1998. So Dr. Zuckerman and then Dr. Ramamurthy.

DR. RAMAMURTHY: I just want to clarify, I don't want to --

DR. HENDERSON: You're speaking for both of you?

DR. RAMAMURTHY: I just want to clarify, you brought up the issue of 510(k) versus PMA, and I don't want to get into the weeds here, but those particular regulatory pathways are linked to risk classification, so on and so forth, so it's not comparable to what you said on the drug side, the various forms of approval. I just didn't want a faulty comparison between the two different paths of approval for the different centers, they are very different and they have different purposes for how they are done.

DR. HENDERSON: Dr. Zuckerman.

DR. ZUCKERMAN: Yeah, I've followed some articles on that that I won't go into, but I do think that these tests would go under a PMA, would go under the more rigorous standards. Although they are not as rigorous

as they are for drugs, I still think they have certain advantages over the CLIA process, and so I do have more confidence for tests that have been approved by the FDA.

Also, I wanted to mention something that I think was raised in public comment or in the discussion earlier about the shortcomings of postmarket surveillance on these kinds of tests, because although the requirements for reporting are for deaths and very serious injuries, the voluntary system under the FDA is for all kinds of complications and problems, and that's something that CLIA doesn't have, so that is an advantage of an FDA approval, is that there is a system for reporting problems of all types, even though it's underreported usually.

DR. HENDERSON: Okay. Any other questions, or comments? Yes, Marcel?

DR. SALIVE: So, one point that hasn't been brought out about this question is that we had, I think, very heterogeneous evidence because of some of these regulatory pathways, and so I did feel it came out from the

25 presentation from the CAP person on really how

1 the labs are doing these tests in practice, but
2 it's also replete in the literature that, you
3 know, you have to dig down very deep to figure
4 out what test was actually done in a paper, and
5 I believe that the TA did a lot of good work on
6 that, but it's not clear always, you know, if
7 there were six studies whether they were
8 always, all six from the, an FDA-approved
9 product, or if there was some mixture of those.
10 And in any case, it affects our
11 ability to use the body of evidence, the
12 totality of evidence because, you know, it
13 limits us in some ways from using international
14 literature, so it's especially unclear on
15 what's going on elsewhere.
16 DR. HENDERSON: Okay. Again, on the
17 FDA versus laboratory studies, any further
18 comments on that topic? Okay. So we will move
19 on then, if you feel comfortable doing that.
20 DR. ROLLINS: Yes.
21 DR. HENDERSON: We'll move on to 4(b),
22 again, are these factors which will change the
23 generalizability of the evidence about these

tests, and that is the type of performing laboratory, for example, university medical

centers, independent commercial laboratories, or community hospital-based laboratories, would this affect the generalizability of the evidence? Now we're not really talking about whether you trust your community laboratories to do this, we're talking about would a series of tests that generated evidence upon which you're basing your vote that you just made, if you found that was all done in community laboratories, would that make a difference, compared to if you found it was all done in university-based laboratories? Dr. Berger.

DR. BERGER: Yeah, I think it's basically (b) and (a) are related to each other, okay? So that as tests get more standardized, I don't think they have to go through the FDA process, the evidence is more generalizable because the formats for doing the test is the same. When you come down to the different types of laboratories you add another layer of variability in there and you really can't make a, without knowing the lab you

really can't say a university lab is better than a community lab.

DR. HENDERSON: So you're saying a

CLIA lab that's based in a community hospital is just as compelling as a CLIA lab that's based in a university.

DR. BERGER: Oh, absolutely, it depends on who the lab is, it depends on the lab director.

DR. HENDERSON: So it's not the laboratory, it's the certification.

DR. BERGER: It's the laboratory itself and how well they develop the process if it's a CLIA lab, and you may have a very well run lab in the community setting, you may have a very poor one in the university setting if people know how to do it, so to speak, or maybe vice versa. So I think it really does depend on the site if it's not an FDA-approved test. If it is an FDA-approved test, it's standardized.

DR. HENDERSON: Okay, anybody else? Dr. Guadagnolo, did you have a question, or comment?

DR. GUADAGNOLO: Just a note of concurrence. I mean, I think that if they are participating in certification in the labs, it's like Dr. Nowak presented this morning, it

didn't matter which type of laboratory in which the testing was performed.

DR. HENDERSON: You say we did not have that?

DR. GUADAGNOLO: Yeah, he presented this morning that it was CAP certified, it sounded like they had done this and participated in the accreditation, and it seems like there is parity.

DR. HENDERSON: Yeah, okay. Any other questions on question 4(b)?

So now, question 4(c), again, does this factor affect the generalizability of the evidence, and that is subgroups in the Medicare beneficiary population, for example, age? So is it, I guess in part you're asking if this were limited to 50-to-60-year-olds, I don't know, but let's say the evidence is based on 50 to 60, would you be as likely to generalize it if it were only under 60, and that is not

usually a Medicare population? Those are the questions you're asking, right?

DR. ROLLINS: Right.

DR. HENDERSON: So, Dr. Zuckerman.

DR. ZUCKERMAN: Well, I am very

concerned about the age issue even though you would suspect that the test would be, the analytical validity would be the same, but certainly in terms of outcome it could be quite different. So when you have older people who are more vulnerable, who have other health conditions, we don't really know how well they will handle treatment and we don't know if they will choose treatment, and we don't know how successful the treatment would be. And I guess I would just hope that CMS would make a stand to really demand better data on patients over 65. I know that not all Medicare patients are over 65, but certainly for cancer the vast majority would be, and, you know, I think it's really unfortunate that we so rarely have subgroup analyses for people over 65.

DR. HENDERSON: Okay. So if I heard you correctly you're saying that, let's say

that you had 10,000 patients in your database and 5,000 were above age 60 and 5,000 were below. You would feel more comfortable not only having that mix, but also doing a subgroup analysis for those or looking at the, to determine at least that the effect was not

significantly different.

DR. ZUCKERMAN: Right, exactly. I mean, it's important to have a substantial number, but I think what's more important is the subgroup analysis and to look at those groups, analyze them separately and see what the complication rates are, what the success rates are, and the various different outcomes.

DR. HENDERSON: I think that's pretty clear. Dr. Salive.

DR. SALIVE: I agree with the last comments. Also I would add that, you know, one of the key issues is really what do these prognostic tests contribute over and above the prognostic factors that are already known, and so I believe the TA had a Table 1 on page three that talked about all the prognostic factors, and, you know, commonly age is the most

powerful factor in determining prognosis in many conditions, you know, in some of these cancers it is but not all of them, I believe. But the issue is really getting prognostic factors that are, you know, contributing above what we already know.

So there is, you know, the staging

system and so on that does perform quite well, so, you know, I believe that the recommendations of the last speaker, you know, with that in mind, you could look somewhat at the issue of do these prognostic factors, you know, are they better at a younger age, at an older age, do they interact with age as a predictor of the ultimate prognosis?

So I think there are definitely questions that remain that, you know, we did not get into that today, there wasn't sufficient evidence yet, but those are important questions that need to be addressed.

DR. HENDERSON: So if I could take what you've just said and maybe add an example and see if this fits with your point, if you're testing a prognostic factor in a population

most of whom are going to be dead in five years, it may have very little value compared to a population of 45-year-olds, most of whom would normally be alive in ten or 15 years. Is that not illustrative of the point you're making?

DR. SALIVE: Right. I think there is definitely a feeling effect from age, where

your prognosis may not be super long, yes.

DR. HENDERSON: Dr. Grant.

DR. GRANT: I'm kind of, it's not really about the physiology of subgroups, it's really more about generalizability of decision-making, which is really implicit in any prognostic or, for that matter, diagnostic test. So, the issue is what are the thresholds that drive decisions, because the first question, I think, as part of an evidence evaluation is to say can this test validly and reliably discriminate among these thresholds, over which you choose one management decision less than you might choose another. And the threshold for decision-making is really determined, well, it's the physician, but

really it's how patients value the relative benefits and harms, and they'll vary for an individual patient. Particularly because elderly folks oftentimes have a lot more to lose with an intervention that in fact has potential harm. You know, a week of bed rest can do somebody in where, you know, in a 70-year-old or an 80-year-old.

And so I think that it's incumbent

upon us, and we really didn't touch upon it here, but in evaluating evidence for these kinds of tests, it's fairly straightforward to extend them to that kind of analysis, to say that over the range of thresholds where decision-making is likely to occur, that is how we value benefits, does this test really add, and so does it give us incremental value, do we have more true positives with the same, you know, false positive or false negative rates, and that's really what these tests are about in terms of driving decisions absent a randomized controlled trial. And I don't think we really pay enough attention to that because really, it can be quite a bit different, I think, in older

folks, particularly frail versus younger.

DR. HENDERSON: Okay. Dr. Ramamurthy, you had a comment?

DR. RAMAMURTHY: No comment.

DR. HENDERSON: Okay, sorry. Dr. Kamrava.

DR. KAMRAVA: I think in certain disease sites we clearly know that the, you know, biology of disease is very different in age. I mean, breast is a great example, I

mean, clearly we manage patients who are HER positive that are over 70, it's different than women that are not, and so I think that does apply to the Medicare population. There definitely are disease sites where we already know that the prognosis should be different in certain subsets, and we should probably think about that.

DR. HENDERSON: Thank you. Dr. Kamrava. Dr. Berger.

DR. BERGER: Yes, just a technical point along those lines as well. So one of the markers that really didn't make the cut today was MLH1 methylation, but more and more we're

going to see more of the epigenetic biomarkers come to the fore, and methylation, background methylation increases with age, and therefore, different lab-developed tests or different test approaches will have different analytic performance, which may or may not be able to distinguish age-related changes from real changes so it will get a little more complicated, but that is another age issue.

DR. HENDERSON: Cancer's a form of aging, yes. Dr. Burke.

DR. BURKE: Well, to get back to your point, people don't realize that the threshold you set depends in part on how accurate your test is. If your test isn't very accurate, then your mistakes, then your variability is very high, so setting a particular threshold has to take into account that your test isn't very accurate so you have to go a lot either further down or further up to achieve your goals. So it's how good the test is that begins to determine where you can set a particular threshold.

DR. HENDERSON: Okay. So, any other

comments about the subgroup on age? Maren?

DR. SCHEUNER: If I could just make a brief one, I've been thinking about, just as a geneticist thinking about any ancestry, your germline, your background and how that may make a difference with some of these tests. It's just something to think about.

DR. HENDERSON: Okay. So now, for (d) is, again, the degree to which this factor may change the generalizability of evidence, and that is genomic variations within cancers, for example diversity of cancer genomes.

Dr. Ramamurthy.

DR. RAMAMURTHY: So, this is a topic that I'm keenly interested in, because increasingly as molecular pathology testing is becoming more next generation sequencing, larger genome panels and so forth, the earlier questions that this panel tried to address on analytical validity and clinical validity, before long the issue of analytic validity will be solved because really, these next generation sequencing platforms are incredibly accurate and once they are properly set and fine-tuned,

then there aren't very many chances of getting inaccuracies.

The issue is going to be on clinical validity, and this is an issue also, I know the regulatory agency, FDA is also dealing with the next generation sequencing and how to regulate them, mainly because within a cancer genome some mutations are the more causal mutations or, the term has been used as driver mutations, versus some are more passenger mutations. And therefore, to be able to have clinical validity, one might not have sufficient sample size that has all of these mutations, so one

has to almost get to a compounded level of study design, where one looks at an entire panel with varying scores of the passenger mutation having to be present or not present.

So I think as we go forward, this issue will come back for us to study again and again. Therefore, that becomes a factor in generalizability of molecular pathology tests.

DR. HENDERSON: Okay. Dr. Berger, you look like you're on the verge of a remark.

DR. BERGER: Well, I just wanted to

say that that's true within assay, but even larger, as you have a large tumor you might have hemastatic deposits. Setting a standard at some point for how widely things have to be sampled and sequenced to make sure we catch all that genetic variability will be the topic of other discussions, I'm sure.

DR. HENDERSON: Wouldn't that be an important element of the analytic validity, determining, defining how much of a tumor and how many loci it needs to be tested at, isn't that intrinsic in the analytic validity?

DR. BERGER: Yes, and that's what's going to index up to the clinical validity, so

it's all related, but we don't have guidance right now.

DR. RAMAMURTHY: Well, in terms of analytical validity, you heard mentioned today in some of the past presentations that for an example with KRAS, there are known seven mutations that are much talked about, the codon 12-13 mutations, but then an earlier speaker talked about many many many more mutations that are also found in KRAS. In terms of analytical

validity, once you have a next generation sequencing panel, it should be quite easy to identify them all in terms of being able to have the right sequencing apparati and then the software. It's how mutation number 8, 9, 10, 13, 14, how clinically valid and clinically associated would be the problem.

DR. HENDERSON: Okay. Dr. Scheuner, no comment on this one? Okay. Anybody else have comments? I think people are getting worn down.

So that, then, finishes our discussion of the questions that we voted on. So now, the last segment are your final words, your take-home points. And I think that what we'll

do is we'll go down the panel and ask each person to make a final comment, it should be a maximum duration of two minutes, and then if there's still some remaining time, in other words, if you've all kept within two minutes there should be some remaining time, then you will be able to offer last minute kind of impromptu remarks if you have anything to add.

I would encourage you, though, in

these remarks to think about what you think are the most important things that you need to highlight to CMS based on the discussion of the day, taking the presentations, the panel discussions, the questions, the voting process and everything into consideration. What are the things that you believe need to be highlighted, or there may be a point that you don't think has been made all day for some reason or another that you need to be, that you think needs to be added, or finally, I don't know that some of the comments, particularly the ones that Dr. Ramamurthy made where we really tried to make it very clear that those are comments to be added to the voting and put into this part. I don't know if those need to

be repeated, but, you know, things that you feel are provisos or comments, or footnotes, if you will, to the whole voting process are appropriate here as well.

So we'll start with you, Dr. Burke, and moving down the panel, and then come back for any additional closing remarks that haven't been covered in the individual remarks.

DR. BURKE: One thing, Mr. Chairman, you did a great job, by the way, thank you. So, I've been doing predictions for about 25 years now and it's a very tough area, and that's why this was very tough, because it has its own complexities and its own variances, a predictive accuracy of whatever variety is very difficult to obtain.

I mean, 20 years ago I did a TNM and found out it's an ROC that's the same as MammaPrint today, so the point of putting multiple factors together is the way things are going to go, including age and perhaps the history and other factors, but it's going to be very challenging.

DR. HENDERSON: Thank you. Dr. Fischer.

DR. FISCHER: I think the one thing that I see is that ethics and the reasons for giving drugs which may be harmful are loosened up, and we have people getting potentially toxic agents for making money. In actual fact, I mean, you have a situation with carcinoma of the breast which is becoming more and more

defined about who gets what, in which only women, only four to six percent of women should get chemotherapeutic agents, but in putting together the seventh edition of Mastery of Surgery, which is the standard surgical textbook, it is amazing how many women get chemotherapy, and there's only one reason for it, they can charge for it.

I think that to me has been gradually evolving until it seems to be all right to do that, and since these are toxic agents and mistakes happen, I think we have to be on the side of the angels in minimizing those drugs that do not necessarily need to be given, and we see it all the time in clinical medicine.

DR. HENDERSON: Okay. Dr. Grant.

DR. GRANT: First I want to thank everybody, particularly the speakers today, it

was very very informative, and also the CMS folks.

A couple things. The first is, I think the task here presented, at least for me personally, was somewhat daunting given the number of tests, the number of potential

outcomes and all the other nuances that are involved, and diagnostic/prognostic tests are difficult at best. But I think as far as from the evidence perspective, and that's what I will address as my thing, I think, is it's really important to superimpose a decision analytic construct to be able to derive decisions about the potential benefit from a test.

And so it means going beyond, as we've mentioned, the relative risks, and ROC curves, you know, they're good to look at, but they don't tell the story, small improvements can actually translate into better decisions. And I think we do have the tools at our disposal today to extend that conversation a bit to bring it to the end. There's decision curves, there's utility curves, and they are the kinds of things that are actually not difficult to

understand that articulate conclusions, allow us to draw conclusions from evidence in a way that just looking at little bits and pieces don't. So, that's my parting thoughts.

DR. HENDERSON: Okay. Dr. Guadagnolo.

DR. GUADAGNOLO: So, much of what I was thinking has been said, but I guess one thing along the line of decision-making, as we worked our way through the voting list it became clear that the data regarding patient outcomes and what was influencing patient-physician decision-making and the downstream effects of these tests, it all got very thin very quickly, and in the new era of precision medicine there's a lot of excitement for that from all of us. We heard from Dr. Ross, who very passionately spoke about it where we're profiling tumors in cancers differently.

I think unfortunately, the entrepreneurial and enterprising thrust that goes with that is maybe, we're going to have to focus more attention on actually gathering the data about the patient experience and the actual patient-centered outcomes that go with

it. I know CORI does that to some extent, but I wondered if CMS in its interdepartmental discussions within HHS could talk about, whether through NCI or AHRQ, whether there

could be some companion funding to get research incentivized along those lines, because it's not going to happen as quickly if that doesn't happen.

DR. HENDERSON: Okay. Dr. Kamrava.

DR. KAMRAVA: Yeah, I don't have too much more to add other than I think today's discussion, I think, brought out the complexities that we have as clinicians, knowing that, you know, many of our standard treatments aren't working and we all want to do better for our patients, but we're all kind of doing the best that we can with these markers, but the data is just not quite there yet, but we're trying to do the best that we can. And I think the issues of predictive and prognostic and how we use them and how they're developed, it gets very tricky because we're all trying to help patients in places where we know standard things don't work, and hopefully we'll be able to figure it out soon.

DR. HENDERSON: Okay. Dr. Salive.

DR. SALIVE: So, I'm very appreciative of everyone, especially Maria, thank you,

Maria, and all the speakers.

I think one piece to add to what Mark Grant said, I think also clinical decision-making tools have looked at predicting the harms of chemotherapy, and we didn't hear much about that today, but in the geriatrics world that has been coming into some scrutiny, and I think it's a factor that's prominent for a lot of people who get older and may not want those harms, they may not want to be exposed to those harms. We heard a little bit about it but it was, I think, glossed over a bit. So I think there's a lot of room to work in this area and I think, yes, it's just beginning, so thanks.

DR. HENDERSON: Okay. Dr. Scheuner.

DR. SCHEUNER: Not much more to say either, I think that a lot has already been mentioned. I also appreciate everyone's participation.

I found this to be very difficult in thinking about prognostic versus predictive,

and there's a spectrum, and it's really hard because I think in practice we wouldn't just

look at a test for just one issue if we know there's more there. And then I was also, I guess not surprised, but there's just not much data about the utility issue, even just around management, shared decision-making, patient decision-making, so clearly there's a lot more we need to do there.

And I guess the other would be how we value clinical validity in and of itself. You know, I suppose that it has to be tied to a certain outcome, but maybe at some point in time just having information about what will this test predict has value to a patient or a provider in trying to work through things.

DR. HENDERSON: Okay. Dr. Zuckerman.

DR. ZUCKERMAN: I need to agree with almost everyone about almost everything. This is such a promising and exciting field and yet the excitement is so far ahead of the data, and I think that's what's so frustrating, when this could be so helpful if only we knew more. So, you know, it's just, the data are so insufficient. We need to know what the impact

of these tests are on treatment decisions, we

need to know how that affects outcome, and we need to know not just about the usual outcomes but also complications, and that's, we're missing a lot of information in all of those areas.

And I will just say, at our center we talk to a lot of patients who reach out to us to try to figure out what their options are, and they're very confused, they get these kinds of tests, it does influence their thinking, but often it frightens them instead of really helping them, and I don't know why that is, but we need to do a better job, I think, of explaining these tests to them, because clearly the benefit of not undergoing chemo would be a tremendous benefit, for example, if that's the way it works. But if instead we keep hearing from patients where the results are not so clear and so they think maybe they should have chemo and otherwise they wouldn't have, so it can work both ways.

And just, my final thoughts are just that we're talking a lot about individualized medicine and personalized medicine, precision

medicine, but we're still not looking at the effects of major group identification such as age groups, sex, and race and ethnicity, and the possibility that that could affect the usefulness and the clinical outcomes for these kinds of tests and the treatments related to the tests. Thank you.

DR. HENDERSON: Dr. Ramamurthy.

DR. RAMAMURTHY: I thank you, Mr. Chairman, for a very nice MEDCAC, and thank you for this opportunity to participate in this important meeting.

The meeting began with a bit of confusion about predictive and prognostic and all these terminologies, but I think we kind of battled through it and we have kind of come to a very nice understanding around the key issues.

I think this MEDCAC is very very timely. It comes at a time when personalized medicine, the next frontier is here. The White House issued a precision medicine initiative, the FDA has been releasing guidances on co-development and has been approving increasingly co-developed drugs with companion

diagnostics, so for CMS to have this MEDCAC is indeed very timely.

It also means that all the stakeholders, that includes the drug and device manufacturers, patient groups, have to work together to try and figure out how best to deliver value from these amazing new technological advances, because don't forget, after all, the patients are waiting for the next treatment.

DR. HENDERSON: Okay. And the last word, at least for this moment for the panel, is Dr. Berger.

DR. BERGER: It's always tough to be last and find something more cogent to say, but I agree with everybody, thanks to all the speakers and thanks for the opportunity to be here.

I guess the only thing that hasn't been said yet that I'll put out, and one is a suggestion, and I think it would have been helpful to the group if we had penned vignettes that would have walked people through the questions with a case or two example, I think that probably would have helped focus people's

attention to the right places earlier on.

And the other thing I would like to say, and this sort of indexes to what Lakshman was just saying, that the half life of technology development and getting new tests out there, it really is a lot faster going forward than some of the lengthy followup that we use to assess those tests, and that is going to be a challenging issue in the future, that when we have ten years of retrospective data, the first eight years of it probably doesn't matter anymore, and we're going to be stuck with smaller and smaller databases over time.

DR. HENDERSON: Okay. So now, each of you have heard your other panel members. Does anybody have additional comments or items they think need further discussion or that have not been really fully ironed out yet?

Dr. Rollins, do you want to make some comments?

DR. ROLLINS: Sure. First, I would like to thank Dr. Henderson as well as the members of the MEDCAC, the presenters of the TA, as well as guest speakers and public commenters.

It's interesting if you take a look at all four questions, question 1(a), 1(b), question 2 and 3, it turns out that Oncotype Dx was the one that was consistent in terms of the clinical utility all the way back through analytic as well as clinical validity. The other two which also met that were microsatellite instability for adenocarcinoma of the colon as well as the rectum, as well as MammaPrint, although they were not as high as it was for Oncotype Dx, which is consistent with what the TA showed.

In terms of EGAPP, the ACCE model for molecular diagnostic tests, CMS will be using more and more clinical utility in the future when it comes to diagnostic tests. We've approached various lab vendors and they are able to show us plenty of information in terms of analytic validity as well as clinical validity, and some of them feel that demonstrating clinical utility is something that they don't need to provide commercial insurers. I don't know why they feel that way, but as I said, as time goes on we will be using

more and more clinical utility.

Now we've defined clinical utility as improves health outcomes either due to increased benefit or reduction in harm, and I think from a patient's perspective that's what's important to them, so that clinical utility will be something that will be used more and more as we continue to delve more and more into diagnostic tests.

And that's all I have to say.

DR. HENDERSON: So, I think I get the last word.

DR. ROLLINS: Yes.

DR. HENDERSON: I think one of the things that it's important for us to recognize is that the process of discussion which we've had today is very important in changing a lot of practices. First of all, the practice of medicine, and particularly our communication with our patients. I think the more the patients understand some of these processes, understand the way of determining whether these things are going to be of benefit, whether they're reliable, will lead to a different

level of comfort with their medical care, further confidence in the medical profession.

I would actually not say that he's wrong, but would feel that Dr. Fischer's comments about chemotherapy being overused because of the charges, I won't deny that that has an element, but I do think that there's also the problem when you have a cancer patient in front of you who, you know, has all of the feelings of cancer being associated with death, in spite of the fact that the statistics are actually nowhere anything close to what the perception is, that both the doctor and patient feel this need to do anything, that just the perception of danger is sufficient to justify almost any degree of risk, and the potential one-in-a-million chance that I'm going to be cured, even though we don't really quite know what that means in most cases, is enough to justify all kinds of processes and/or therapies. And I think that it's important that patients begin to understand better what goes into making a decision, particularly a decision of coverage, I think they're going to

feel less hostile to the way we fund health care in America.

But the second place where the

processes, this kind of throwing in a new system to most physicians, even to most researchers, the EGAPP process, the ACCE process that was described today, will have a profound effect, I think over time, in terms of the way companies develop these new tests. And I think this is an augmentation in a way of what you just said when companies are coming to you. I think in many cases they just haven't understood it, but when they understand it, not only understand the process, that it's very rational, that it is doable, even though clinical utility is a very expensive determination, which is why we see the least clinical utility data, that they will in fact see the value of doing that.

And the role that health care, that health care insurers have taken in recent years is complementary to the FDA. What the FDA looks at and what the insurers are looking at are somewhat different, and their roles are

different. I think that the public and people who develop these tests are only kind of beginning to wake up to the first, the realization that they are playing this role,

and number two is what's the nature of that process.

So I think that just spending a day discussing these things publicly has in itself a worthwhile impact, in spite of the confusions and the nuances of our discussion, the difficulty in understanding the differences in these definitions, I think struggling with that by itself gives us greater clarity as we go forward, and I think we have accomplished something today in those interactions.

So with that, I think we're prepared, unless somebody has another final comment, to adjourn the meeting. Thank you.

(Whereupon, the meeting adjourned at 3:10 p.m.)