00001

CENTERS FOR MEDICARE AND MEDICAID SERVICES
Medicare Evidence Development & Coverage
Advisory Committee

May 1, 2013

Centers for Medicare and Medicaid Services
7500 Security Boulevard
Baltimore, Maryland

00002

Panelists
Chairperson
Rita Redberg, MD, MSc
Vice-Chair
Art Sedrakyan, MD, PhD
Voting Members
David C. Beyer, MD, FACR
Mary A. Blegen, RN, PhD, FAAN
Steven Gutman, MD
David Howard, PhD
Pamela R. Massey, PT, MS
Jan Nowak, PhD, MD
James Rizzo, MD, MS
Amy E. Sanders, MD, MS
A. Oliver Sartor, MD
Eric C. Stecker, MD
Sandra L. Wong, MD, MS
CMS Liaison
James Rollins, MD, PhD
Industry Representative
Martin D. Marciniak, MPP, PhD
Guest Panel Members
Barbara Conley, MD

Dorothy L. Rosenthal, MD, FIAC
16
Invited Guest Speakers
17 Barbara Conley, MD
Sreelatha Meleth, PhD
18 Dorothy L. Rosenthal, MD, FIAC
Katrin Uhlig, MD, MS
19 Nedra Whitehead, PhD, MC, CGC
20 Executive Secretary
Maria Ellis
21
22
23
24
25
00003
1 TABLE OF CONTENTS
2 Page
3 Opening Remarks
Maria Ellis/James Rollins, MD, PhD/
4 Rita Redberg, MD 4
5 Introduction of Panel 9
6 CMS Presentation and Presentation of Voting Questions
7 Chuck Shih, PhD 12
8 TA Presentation
Nedra Whitehead, PhD, MS, CGC 16
9
Presentation by Invited Guest Speaker and
10 Panel Member
Barbara Conley, MD 30
11
TA Presentation
12 Katrin Uhlig, MD, MS 43
13 Presentation by Invited Guest Speaker and Panel Member
14 Dorothy L. Rosenthal, MD, FIAC 67
15 Scheduled Public Comments
E. Robert Wassman, MD, FAAP, FACMG 84
16 Margaret Havens Neal, MD, FCAP 90
Bernard H. Berins 94
17 F. Anthony Greco, MD 99
Lawrence M. Weiss, MD 104
18 Catherine Schnabel, MD 110
19 Open Public Comments
Daniel M. Jones, MD 116
20
Questions to Presenters 117
21
Panel Discussion, Final Remarks and Voting
22 Questions 169
23 Closing Remarks and Adjournment 260
24
25

00004
PANEL PROCEEDINGS
(The meeting was called to order at 8:12 a.m., Wednesday, May 1, 2013.)
MS. ELLIS: Good morning and welcome, committee chairperson, vice chairperson, members and guests. I am Maria Ellis, the executive secretary for the Medicare Evidence Development and Coverage Advisory Committee, MedCAC. The committee is here today to discuss selected genetic tests for cancer diagnosis for cancers of unknown primary site, and for cervical cytology findings of uncertain clinical significance.
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
00005
presentations disclose if you or any member of your immediate family owns stock or has another form of financial interest in any company, including an Internet or e-commerce organization, that develops, manufactures, distributes and/or markets, consulting, evidence reviews or analyses, or other services related to genetic tests for cancer diagnosis. This includes direct financial investments, consulting fees, and significant institutional support. If you haven't already received a disclosure statement, they are available on the table outside of this room.
We ask that all presenters please adhere to their time limits. We have numerous presenters to hear from today and a very tight agenda, and therefore cannot allow extra time. There is a timer at the podium that you should follow. The light will begin flashing when there are two minutes remaining and then turn red when your time is up. Please note that there is a chair for the next speaker, and please proceed to that chair when it is your turn. We ask that all speakers addressing the panel please speak directly into the mic and
00006
state your name.

For the record, voting members present for today's meeting are Dr. Art Sedrakyan, Dr. David Beyer, Dr. Mary Blegen, Dr. Steven Gutman, Dr. David Howard, Pamela Massey, Dr. Jan Nowak, Dr. James Rizzo, Dr. Amy Sanders, Dr. A. Oliver Sartor, Dr. Eric Stecker, and Dr. Sandra Wong. 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 available on our website following the meeting. I ask that all panel members please speak directly into the mics, and you may have to move the mike since we have to share.

The meeting is being webcast via CMS in addition to the transcriptionist. By your attendance, you are giving consent to the use and distribution of your name, likeness and voice during this 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

00007

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 conversations 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 these 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 topics during breaks and lunch.

If you require a taxicab, there are 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 this room.

And lastly, all CMS guests attending today's meeting are only permitted in the following areas of CMS single site: The main lobby, the auditorium, the lower level lobby

00008

and the cafeteria. Any person 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. Rollins.
DR. ROLLINS: Good morning. My name is Jim Rollins and I'm the director of the Division of Items and Devices here in the Coverage and Analysis Group.
The MedCAC serves three purposes for CMS. First to get input from experts in the field on the topic, and that information can help us strategize our efforts related to future activities on that topic. Number two, to help disseminate information to the general public. And number three, a more immediate use of the MedCAC is it helps -- I'm sorry -- the more immediate need of the MedCAC, along with the external technology committee, is it helps us to craft national coverage determinations.
I would like to thank the members of the MedCAC, especially the chairperson as well as the vice chairperson, along with participants and presenters for today's

00009

discussion.
DR. REDBERG: I'm Dr. Rita Redberg, I'm a cardiologist at UCSF and I am chairing the committee today, and I'm looking forward along with the rest of the committee to a very informative and interesting discussion on the issues of genetic tests for cancer diagnosis.
I have no conflicts of interest, and I think we'll just go down the line and introduce ourselves and state any conflicts.
DR. SEDRAKYAN: I'm Art Sedrakyan from Weill Cornell Medical College in New York. I'm an associate professor of public health and cardiac surgery. I don't have any conflicts to disclose.
DR. BEYER: I am David Beyer, I am a practicing radiation oncologist from Phoenix, Arizona at the Arizona Oncology Services. I'm also a past health policy council chair for the American Society of Radiation Oncology.
DR. BLEGEN: Hi. I'm Dr. Mary Blegen, I'm a nursing researcher recently of UCSF, and I have no conflicts to declare.
DR. GUTMAN: I'm Steve Gutman, I'm a strategic advisor for a regulatory consulting

00010

firm called Myraqa on the west coast, and Myraqa does provide consulting services for BioDesex, Set Genomics and Foundation Med.
DR. HOWARD: My name is David Howard, I'm a professor in the department of health

policy and management at Emory University, and I have no conflicts of interest.
MS. MASSEY: My name is Pamela Massey, I am a physical therapist retired from the University of Texas MD Anderson Cancer Center, and I have no conflicts of interest.
DR. NOWAK: I'm Jan Nowak, I'm a pathologist for NorthShore University HealthSystem in Illinois, and I have no conflicts of interest.
DR. RIZZO: I'm Doug Rizzo. I'm a professor of hematology, oncology and bone marrow transplantation at the Medical College of Wisconsin, and I have no conflicts.
DR. SANDERS: I'm Amy Sanders, I'm an assistant professor of neurology at the Albert Einstein College of Medicine in the Bronx, and I have no conflicts.
DR. SARTOR: I'm Oliver Sartor, professor of oncology and medical director of

00011

the Tulane Cancer Center.
DR. REDBERG: Would you state if you have conflicts?
DR. SARTOR: Oh, I'm sorry. No conflicts.
DR. STECKER: I'm Eric Stecker, a cardiologist and electrophysiologist at Oregon Health and Science University, and I have no conflicts.
DR. WONG: I'm Sandra Wong, a surgical oncologist at the University of Michigan. I have no conflicts.
DR. MARCINIAK: I'm Martin Marciniak, I'm the industry rep, I'm a vice president at GlaxoSmithKline.
DR. CONLEY: I'm Barb Conley, I'm a medical oncologist. I head the cancer diagnosis program at the National Cancer Institute, and I have no conflicts.
DR. ROSENTHAL: Dorothy Rosenthal, a cytopathologist from Johns Hopkins Hospital. I have no conflicts.
DR. REDBERG: Great, thanks very much, and now I would like to turn the mic over to Chuck Shih from CMS for a presentation and

00012

discussion of our voting questions today.
DR. SHIH: Good morning, everybody, and welcome to everybody for joining us here today at CMS. My name is Chuck Shih, I'm an analyst in the Coverage and Analysis Group at CMS, and welcome to today's MedCAC meeting on selected genetic tests for cancer diagnosis.

Just a quick overview of terms before we go through the questions. FISH refers to fluorescence in-situ hybridization. CUP or CUPS, cancer of unknown primary site. ASCUS, atypical squamous cells of unknown significance. And LSIL, low-grade squamous intraepithelial lesion. This slide is meant to just give a sense of what the outcomes of interest are to CMS. For FISH testing they include histologic confirmation of higher-grade cervical intraepithelial neoplasia on biopsy, overall survival, mortality, evidence of harms of anti-tumor treatment, quality of life, and other potential outcomes that may be discussed by the panel today. Similarly for CUP testing, tumor recurrence, overall survival, mortality, avoidance of harms of anti-tumor treatment,

00013

quality of life, and others. Moving on to the questions, question number one: How confident are you that existing evidence is sufficient to confirm the clinical validity (defined as how reliably the test results are associated with the presence of the disease or target condition) of each of the following? A, DNA- or RNA-based testing (CUP testing) to predict tissue of origin for CUP. And B, FISH testing for cervical cancer/pre-cancer in patients with ASCUS or LSIL. For scoring on question one, it will be from one to five, one being low confidence, five high confidence. If the answer to either part of question one is at least in the intermediate range, that's a group mean score of 2.5 or more, move on to question two. If not, we'll skip two and three and move to four. Question two: How confident are you that there is sufficient evidence to determine whether genetic testing of tumor tissue affects health outcomes, including benefits and harms, for patients with cancer whose anti-cancer treatment strategy is guided by the results of

00014

each of the following? A, DNA- or RNA-based testing to predict tissue of origin for CUP. B, FISH testing for cervical cancer or pre-cancer in patients with ASCUS or LSIL. Scoring for question two will be on a one to five scale as well. Assuming an intermediate score of 2.5 or greater, we'll move on to question three. Question three: How confident are you

that there is sufficient evidence to conclude that genetic testing of tumor tissue improves overall health outcomes, including benefits and harms, for patients with cancer whose anti-cancer treatment strategy is guided by the results of each of the following? A, DNA- or RNA-based testing to predict tissue of origin for CUP. B, FISH testing for cervical cancer or pre-cancer in patients with ASCUS or LSIL. Scoring for question three will also be on a one-to-five scale.

Question number four. Please discuss whether the evidence as presented may be generalized based on each of the following factors: A, regulatory status of test, FDA approval or cleared versus laboratory-developed

00015

testing. B, site of testing, e.g., university medical center or commercial laboratories versus community-based laboratories. C, patient subgroups within the Medicare beneficiary population, e.g., age.

Question five. Please identify and discuss any evidence gaps in assessing outcomes of interest to CMS for both, A, DNA- or RNA-based testing for tissue of origin for CUP, and B, FISH testing for cervical cancer or pre-cancer in patients with ASCUS or LSIL.

Six, this is the final question. Please comment on whether CMS should encourage development of additional evidence relevant to coverage determinations for, A, DNA- or RNA-based testing to predict tissue of origin for CUP, and B, FISH testing for cervical cancer or pre-cancer in patients with ASCUS or LSIL.

Thank you.

DR. REDBERG: Thank you, Dr. Shih, for setting the stage, and that's very helpful in terms of what we will particularly be paying attention to in the presentations so that we can inform the voting questions.

00016

I would like now to introduce Dr. Sreelatha Meleth, a senior research statistician in social statistical and environmental sciences, as well as Dr. Nedra Whitehead, a senior genetic epidemiologist, also from RTI, to go through the TA presentation, and you have 30 minutes together.

DR. WHITEHEAD: I'm Nedra Whitehead, I will be doing the presentation, and Sreelatha is here as well. I want to thank you for the opportunity to discuss the technology

assessment of genetic and molecular tests to identify the tissue of origin in cancers of unknown primary site. The technology assessment was funded by a contract from the Agency for Health Care Research and Quality. Dr. Meleth and I are responsible for the content of the report.

A multiprong approach is used to identify the tissue of origin for tumors of unknown primary site, including microscopy, staining, imaging, and most recently molecular and genetic tests. Until recently, genetic testing was used in cases primarily whenever differential diagnosis included a cancer that

00017

was associated with specific chromosome rearrangement, and site of genetic testing was used to determine if the rearrangement was present, to include or exclude the cancer associated.

Recently developed tests focus on a pattern of gene expression or microRNA expression rather than specific chromosome or gene rearrangements, and these tests examine pattern level, the levels of the expression as well as the presence or absence of multiple messenger RNAs or microRNAs. Statistical software analyzes the patterns to predict what the tissue of origin is of the tumor.

We focused on these five key questions. What tests were available, their analytic and clinical validity, the validity of the statistical algorithm for the test, and the evidence that the tests had clinical utility and that they were relevant to the Medicare population.

The analytic framework we used for our technology assessment. For the first four key questions we used studies that had patients of any age group that had cancer of unknown

00018

primary site. For the key question on Medicare analysis we examined which studies had patients that were 65 or older and in the core Medicare population. We included both studies that used genetic or molecular tests for the identification of tumors, of tissue origin, as well as studies that compared those methods, the genetic or molecular test to other methods.

And for the studies of clinical utility, we used descriptive studies that looked at outcomes among patients that had a genetic or molecular test, as well as compared directly to patients that did not have a genetic or

molecular test.
We defined a test as commercially available, because that was a criteria, that the test had to be commercially available in the United States, if we could identify a mechanism for a physician to order a test or for a laboratory to order a kit or the test itself. And we looked for these tests using primarily Internet searches through Google with the limits that the page had to be in English and it had to have been updated within the last year.

00019

We used multiple databases to search for studies. We included systematic reviews, controlled trials, observation studies and case series. The search was limited to English studies that were published after 1990, and we conducted strategies using MeSH headings and text wording for each of the individual lab tests that we identified.
We graded the studies and synthesized the evidence qualitatively based on the methods described in the Evidence-Based Practice Methods Guide for Medical Test Reviews. We conducted meta-analysis on the question for clinical validity, which was the only question for which there was sufficient data available to do a meta-analysis.
Then we assessed strength of evidence using the evidence-based practice center domains of risk, which are risk of bias, consistency, directness, precision of the evidence, and we used the Simon criteria for whether or not the algorithms were developed using statistically valid methods.
And these criteria are how the tests are normalized, the statistical classification

00020

that was used, whether it's a supervised or unsupervised classification, with a preference for supervised classification, and how much risk of bias there is in the validation methods.
We evaluated whether the body of evidence was applicable to the Medicare population by examining the study populations that were reviewed in terms of age, race, gender, and what the diagnosis was that were included in the study population.
We retrieved 840 citations, we ended up including 41 cases on CUP. We also identified eight articles that looked at FISH testing for Ewing's sarcoma. Those were

examined separately and I won't talk about those any further today. Generally the quality, in fact most of the studies were rated as good, some as fair, and one as poor, and it was excluded from further analysis.

We identified four genetic or molecular tests to identify the tissue of origin, the Pathworks tissue of origin test, CancerTYPE ID test, miRview mets and then the later version of the mets2 test, and

00021

chromosomal analysis.

The slide shows the analytes that are actually measured, the panel size, how many different analytes are measured for each test, the laboratory methods used to identify statistical and analytic methods. Also the number of tumor types that are identified by the tests, but there is some variability in how the different manufacturers classify tumor type versus tumor site. For the most part, this represents tumor site only. And finally, the reported score, the way the laboratory reports out the results.

From this point on I'm going to focus primarily on the three molecular tests for our primary analysis. All three of the molecular tests identified the ten most common primary sites according to a recent review of autopsy series. There was at least one study on the analytic validity of each test, and this is the ability of the test to measure the actual topology that it is designed to measure. Some of the studies of clinical validity and clinical utility recorded assay quality measures as well, and we considered those

00022

whenever we looked at the evidence on analytical validity.

Most of the publications reported different measures of analytic validity, so even though you have multiple studies you often only have one report of any given measure of analytic validity. The exception with this was there were two separate independent studies that reported interlaboratory correlation for the Pathworks TOO test, and in both cases the interlaboratory correlation was about 90 percent for that test.

There was sufficient information in the literature to determine whether or not the algorithm met the Simon criteria for the CancerTYPE ID and the miRview test, and all the criteria were met. For the Pathworks TOO test,

there is sufficient information on the internal and external validation measures that were used to determine those criteria were met, but there was not sufficient information reported to assess the dimension reduction criteria. The classification rule was clearly supervised but there is not more detail available on that, and so it was, we were unable to determine whether

00023

or not those criteria were met. Most of the evidence that was available was available on clinical validity.

There were multiple studies, independent studies for each test. There was, as I said, enough information here to do a meta-analysis, and overall we found that the degree of accuracy for the tests were very close, from 85 to 88 percent, with very tight precision levels. There were two studies that addressed, directly examined and compared IHC staining to one of the molecular tests. One study used CancerTYPE ID and found that 78 percent of the predictions from CancerTYPE ID were correct, compared with 68 percent from IHC staining, and the other used Pathworks, Pathworks TOO test. In that case they had multiple pathologists examine the results of the IHC staining, so even though they only had, they had ten tumors that were looked at by ten different pathologists. The Pathworks TOO test called the right tumor type in 90 percent of the cases, and the pathologist called the right tumor type in 64 percent of the cases.

There is very little information

00024

available on clinical utility, which is how well the test actually works in practice, and how much effect it had. The quality of the studies here was not as good as the quality of the studies on clinical validity.

In looking at the ability of these tests to actually make a diagnosis, there are multiple studies that found that in most cases, between 57 and 100 percent, there were -- they were able to make a diagnosis in 57 to 100 percent of cases, sorry about that, in these studies, and in the larger studies it was usually over 90 percent.

There were a few studies that examined whether or not the diagnosis of the tissue of origin test matched another source of diagnosis. In some cases that was cases where the diagnosis, the tissue of origin was later found, the primary site was later found, and

somebody just compared it to the clinical pathologic characteristics of the tumor. They found that they matched in 48 to 88 percent, but as I said, there's a wide range of confirmation standards there.
In five studies they looked at whether

00025

or not the test changed or resolved a diagnosis, it did in 44 to 81 percent.
One study surveyed physicians who had used the test and asked whether or not it proved to be clinically useful, and they found in about two-thirds of the cases, and there's two numbers here because they reported on it twice, and they reported slightly different numbers each time.
There were very few studies that reported on the usefulness of the test for treatment decisions. Of those who did, four studies looked at whether or not the test changed treatment, and it did in 26 to 81 percent of cases.
None of these studies are well controlled. They either have no control group and they're strictly descriptive, or else they're comparing historical cases or in some cases self-selected people who weren't willing to have their treatment assigned based on the test.
One study found an increase in site-specific treatment compared to empirical treatment, and four studies looked at treatment

00026

response and found a wide range of 41 to 74 percent that responded to treatment based on the TOO test. One study compared that to empiric control and had 17 percent, but again, there is no adjustment for differences in the population as controls for that.
In looking at outcomes, there's even slightly less information here, and they had the same caveats as the ones on treatment decision. There are no well-controlled studies, they either have no control group and they're only describing what happened in their population, or else they're comparing to a historical cohort.
Of the studies that have looked at survival, compared to empiric treatment it's about a three-month increase in survival. The survival overall for all the patients who had a TOO test ran from 13 to 21 months. One study calculated as a projected survival the patients who were still living and found, again, about

three to four months, and adjusted for quality of life at about three months. One study reported on how many of the patients had stable disease, and it was like 33 percent.

00027

The tests are primarily done, actually, in patients in the Medicare population. Of 19 studies of clinical utility, there were almost 2,400 patients who were over 64. The studies included both sexes. Very few of the studies report on race, and so we were unable to address that. And the tests, the clinical validity studies used virtually every type of cancer, and so there was a wide range of diagnoses.

In summary, the ratings of the strength of evidence were that the analytic validity for CancerTYPE and miRview were insufficient simply because we had only one study of each one of the measures of analytic validity. We considered the evidence of the Pathworks TOO analytic validity to be high. The clinical validity evidence was rated as high for all three tests.

The rate of the evidence that the test consistently predicts a tissue of origin in test patients is moderate, and that that is a confirmable diagnosis, the diagnosis is actually true and confirmed by another method is low. That the test is useful, there's a

00028

variety of ways of how that was addressed, but that was also rated low, and there's insufficient data to elicit treatment change or treatment response. Low evidence on looking at survival, or estimating the survival among patients that have genetic tissue of origin tests.

Some of the limitations of the body of evidence that went into this technology assessment is that it is difficult to determine the true site in CUPS, which makes it difficult to know how accurate the tests are in actual clinical use, that's just a problem with the actual question.

There is no well-controlled studies of the effect on treatment decisions or on health outcomes, and in almost all of these studies the test manufacturers are listed as coauthors and/or are funding the study.

Some of the strengths is there are multiple well-designed studies that test for the accuracy of tissue of origin tests by testing tumors of known primary site. There

were several creative study designs designed to look at the accuracy of these tests among, in

00029

prediction in true cases of CUP, and some of the recent studies directly compared the diagnostic success of molecular tissue of origin test with the traditional IHC staining. Some of the strengths and limitations of our review is we used standard evidence-based practice in our methodology, we used a framework that has been used in other genetic testing, which is the CDC phase framework. We used a rigorous search which captured published studies, conference abstracts and early publication studies as an attempt to get at whether or not there might be some negative studies that weren't ending up being published in the peer reviewed literature. And we were able to use meta-analysis to look at the identification of tumors at known primary site. Under the limitations of the findings here is that the manufacturers are constantly looking to update and revise and improve, or change the test, and so over the course of time the test changes a little bit and then you never quite nail down exactly what test is being done at the moment. And there's a

00030

rapidly evolving literature, I think half of our studies we found in our second search. In conclusion, the likelihood is genetic tests of tissue of origin tests are moderately accurate when tested on tumors of known primary site, the accuracy of prediction in CUP cases is still unclear. Additional and more rigorous studies of clinical utility are needed, and studies that are conducted and funded independently of the test manufacturers are needed.

DR. REDBERG: Thank you, Dr. Whitehead, that was very helpful. Next I would like to introduce Dr. Barbara Conley, who is the associate director of the cancer diagnosis program at the National Cancer Institute. You have 20 minutes.

DR. CONLEY: Thank you very much, it's a pleasure to be here. I head a unit at the NCI that is concerned with the discovery and development of tests that will improve the life of cancer patients, so it is especially nice for me to be able to address this topic. Having said that, I will say the views are my own and not necessarily those of the people

00031
that pay me, and I have no conflicts of interest.
So, molecularly guided treatment, of course, is something we all pin a lot of hopes on if we're treating cancer patients. The idea is you want to get it right the first time because the treatments sometimes are toxic and sometimes if we don't get it right and the disease progresses, you don't get a second chance. You also want to avoid unnecessary toxicity, of course you want to improve the survival while improving the quality of life, and even if you cannot cure the disease, you would like to convert cancer to a chronic disease that doesn't kill the patient.
Now as we've heard, the cancer of unknown primary site is, depending on how you look at it, a reasonably common or uncommon situation. It's three to five percent of adult malignancies. It usually is metastatic on presentation, and usually when cancer is metastatic on presentation we don't think of it as having a cure. The median survival, therefore, is anywhere from two to 12 months, and it's very difficult to predict if we don't
00032
know anything about what the cancer will do, which we usually learn from studying cancers at particular sites, so we don't really have any correlation, any reliable correlation of either traditional pathologic histologic features or genetic characteristics with response to treatment or with survival.
60 percent of the cancers of unknown primary sites can be classified as adenocarcinoma, 30 to 35 percent totally differentiated adenocarcinoma or completely undifferentiated adenocarcinoma, five percent are squamous cancers, and two percent are a class called neuroendocrine. So one of the open questions that we wrestle with is given that cancers of unknown primary site present with metastatic disease, should we expect the biology and the prognosis to be different compared to carcinomas where the primary is evidence, or is this a fundamentally different kind of tumor or is it something we expect to behave like your regular breast cancer or colon cancer that just happened to metastasize? So if we did know the tissue of origin, would we expect CUP to do the same, better or worse if
00033
we treated it like the tissue of origin?

So, people have been studying this for some time and they've really done a great service to people who are treating these cancers in daily practice, and 20 percent of these tumors seem to have a favorable prognosis and these people survive a bit longer, and if you just kind of look down the list here, they may be tumors where you might assign treatments that are generally thought to be fairly effective, although maybe not curable. The rest of them, 80 percent are unfavorable and we don't know what to do with them, but of course metastatic disease of the liver is bad no matter whether you know the primary or not. Brain mets are not usually good. Multiple lung or pleural metastatic disease is bad. Lytic bone disease where the bones are more likely to break is bad, and squamous cancers in the abdominopelvic area where they tend to infiltrate vital structures and cause a lot of pain are quite bad, so that doesn't give us a good framework for help. So then we have to define clinical utility, okay, what is clinically useful? The

00034

gold standard, I submit, would be that you have better outcomes than what's currently available, that would be clinically useful for a test. That would mean that survival is improved. Okay. If we're going to say that survival is improved, by how much does it have to be improved for us to say a test is clinically useful?

Or the toxicity is lessened. We don't give them a bad regimen that's the cause of a lot of toxicity because they're thought to have a different kind of tumor, but by how much is this clinically useful? The current situation, you have a patient who presents with metastatic disease, they don't, there's no obvious primary by the usual clinical workup that you do. They get a biopsy, they get looked at in pathology with IHC and other tests. Meanwhile the clinician is looking at them and seeing how well do they perform their activities that they normally do, how old are they, what other diseases do they have, and what are their personal preferences for treatment of this most likely incurable condition. We look at how many sites there are, where the metastatic

00035

sites are, and any guidelines that might be available such as from the NCCN or ASCO, and make a treatment decision, okay, which you want

to be the best decision up front, okay? So what we want to know is will using a molecular tissue of origin test result in a better outcome for patients. So we assume, in order to use this, that if a primary site can be suggested, then beneficial treatment can be given, okay? But we just heard that no studies show definite improved outcomes even with current procedures or with molecular tests. Currently diagnosis is not actually done in the most optimal way, because even the laboratory developed tests or the IHCs are not standardized across sites, so which one, or even how they're done is not standard across sites. And then of course, we've heard that in 80 percent of patients the current treatment, even if we know the primary, we're not getting the benefit we wish we had. And the second assumption is that molecular profiling can give guidance when other studies are not optimally informative.

So, does the molecular test guide the

00036

treatment better than the current IHC tests, does it add benefit to the current diagnostic regimen, all of these things that we normally do, or does it allow CUP patients to be eligible for clinical trials in patients with known primaries where everybody's treatment, hopefully, can be improved.

So, in our division they're working on how do you validate, how do you look at clinical utilities of molecular markers no matter what we use them for. So basically we do want to define the setting and the desired utility of whatever marker or assay you're looking at. The magnitude of the outcome or treatment effects for a positive assay must be sufficiently different from those of a negative assay so that a clinician and patient would accept a different treatment for those two groups.

If you have two groups, a marker positive group and a marker negative group, but both are going to benefit from the same treatment, you're not really going to change your mind, you're going to still give the same treatment to both, and of course the estimates

00037

of that magnitude must be reliable. A randomized clinical trial is what we have all thought for many years is the best kind of evidence, and this should be maybe prospective, but how do we stratify carcinoma

of unknown primary site in such a trial since we can be, since there are a number of cancers this could possibly be as the primary, and there are good and poor prognosis groups there. And then, since there's only three to five patients in the whole group of cancers of unknown primary, and we may have to subset it to good prognosis/poor prognosis, will standard of care change during that time that we're trying to do the trial?

Then there's something called a prospective-retrospective study. Studies have already been done and there are enough patients with tissue that we can take the data, the outcome data and the molecular data, and use that as a sort of clinical trial without enrolling patients, but do enough trials exist, since I'm not sure that very many trials would have taken carcinomas of unknown primary patients.

00038

Thirdly, can we do a registry, and how do we do that so we understand whether the outcomes would be better using the test or not. So a little bit on retrospective analysis design. These generally are used as hypothesis generating studies and are generally used as convenience samples, meaning these are not patients who have the same eligibility criteria, these are patients that happen to exist in a tumor bank somewhere in our hospital laboratory, so the prospective designs that one can use are various kind of biomarker stratified designs, positive and negative biomarker stratified designs, adaptive analysis designs, did he do better with the biomarker than without the biomarker, so-called biomarker strategy designs, sequential designs using one or more of these, and then hybrids.

So in order to go this route, there has to be an indication of treatment, first of all, and it's best to use the most efficacious treatment first, and the patient of course has to be fit for treatment, although with some of the targeted treatments even patients who are immuno, you could see a very poor performance

00039

that could benefit.

Carcinoma of unknown primary is uncommon, it's heterogeneous, the randomization would be difficult because of the heterogeneity, the patient characteristics themselves are heterogeneous, and one has to define the magnitude of benefit that will

justify using a new test in this group. So if you're going to design a prospective trial, there are certain advantages to that, you're starting from scratch, so you would probably need the fewest patients for that kind of trial, and by designing it ahead of time, you guarantee you will have sufficient power to show a treatment effect. Disadvantages, though, you have to know what the marker is, and you have to turn around to make it essential. In retrospective-prospective designs, you can maximize the pool if you have enough studies, you don't have to know the marker, you can test various cut points and markers, you can refine the assay while the trial is ongoing, and you can look at marker positive and marker negative groups. However, since you

00040

didn't prospectively design the trial, your power to detect the difference might be compromised. Not everybody who was treated may have a sample, and results therefore may not be generalizable. So when I speak of biomarker stratified design, this is kind of what I mean, and hopefully you can see the black writing. You assess the biomarker in everybody and then you randomize, you separate into groups, you randomize biomarker positive and biomarker negative to either the new treatment or the standard treatment. In this situation, what do we do with tissue of origin or CUP, do we lump all of them, do we say okay, we're going to do the good risk group, the poor risk group, those that look like colon cancer, those that look like something else, how are we going to do it, it will take some consideration. So -- but the design does allow the assessment of a new therapy in positive and negative patients, but it might not be practical if you have more than two evaluated therapies. So enrichment design I will talk about next, I'm going to actually talk a little bit

00041

about that enrichment designs are very attractive because they only take the patients who one thinks are going to benefit from the treatment, and they could be beneficial when the stratified design is not preferred because you wouldn't want to give that treatment anyway and you measure the biomarker on everybody, but the randomization is restricted to the biomarker positive group. You need to have an

accurate test, of course, and you can't answer, then, whether the treatment is better in the biomarker negative group, and you can't answer if the biomarker is prognostic rather than predictive. So when I say that, it's prognostic when all patients do better if they had that particular characteristic no matter what the treatment was, or is it particular to a certain treatment, only those patients who get that certain treatment will do better. So this is an example of an enrichment design that may be applied to a tissue of origin type test to assess the biomarker, and if they don't have, if it doesn't predict the tissue of origin, it's off the study. If the patient has the tissue of origin predicted,

00042

they would be randomized to treatment A per guidelines for the tissue of origin that was predicted, or for guidelines -- sorry -- or treatment B for prediction of tissue origin. That's just one example that could be done to generate the evidence, but you would not know the effect of what the treatment was if the tissue of origin was not predicted, because they would be off the study. And the question, again, is you need one trial for each tissue of origin.

Now in this era of predictive and precision medicine, is it better to find the tissue of origin and treat according to guidelines for that particular tumor that we have now, or should we concentrate on predictive tests for all tumors known or unknown, in other words, find the treatment that that patient is likely to respond to? Can we do that today, probably not, but it is a philosophical question.

So, the conclusions that I came to is the evidence for the clinical utility may be very difficult to obtain with a randomized controlled trial or with a

00043

prospective-retrospective study. However, it doesn't hurt to try, and there may be some instances where we do want to try some of these. A registry might provide some advantages, you could have concurrent controls in an experimental group in a registry, you could get wider participation, more different kinds of patients, patients that are found in the community. However, they would still have to probably be good performance status patients to really be able to discern the benefit.

I thank you for your attention.
DR. REDBERG: Thank you, Dr. Conley. That was really helpful in laying out the big picture of the questions and challenges in the field.
Next I would like to introduce Dr. Katrin Uhlig, who is attending physician in the Department of Nephrology at Tufts University, who's going to talk to us about the technology assessment commissioned by AHRQ.
DR. UHLIG: Good morning, thank you for inviting me. My name is Katrin Uhlig, and I will speak to you about the technology assessment on fluorescence in situ

00044

hybridization or other in situ hybridization testing of uterine cervical cells to predict pre-cancer and cancer. This technology assessment was prepared by the Tufts evidence-based practice center under contract with the Agency for Health Research and Quality, and the opinions presented in this presentation are those of the review team.
You can see here the list of contributors, and I would like to point out that amongst our review team we had a cytogeneticist as well as a gynecological oncologist, and we were in close contact and deliberation with our task order officers from AHRQ as well as the liaisons from CMS.
Cervical cancer has decreased in incidence secondary to widely adopted screening. Screening detects precancerous lesions and cancers in early stages when they can be effectively treated. Almost all cervical cancers are caused by infection with high-risk human papillomavirus genotypes. In particular, HPV genotype 16 and 18 alone are responsible for about 70 percent of cervical cancers.

00045

So, where does genetic testing come in? Cervical cancer has genetic changes that occur early in the process before being apparent under the microscope and this is why genetic tests are being developed to enhance early detection and triage of women with abnormal screening tests.
Last year we saw new guidelines for screening of cervical cancer. The 2012 United States Preventive Services Task Force updated its screening recommendations for women, and recommended in those age 21 to 65 a

Papanicolaou, a Pap test every three years; it had been recommended yearly. A consortium by professional societies issued recommendations that were similar, again recommending in women 21 to 65 years age a Pap test every three years. However, in women age 30 to 65 years, they actually recommended as the preferred strategy co-testing with Pap and HPV screening, and that would allow to extend the screening intervals to every five years. And let me just point out that testing that is recommended here is a screening test only for genotypes, and

00046

this is different from FISH testing for HPV. Now, once a screening test turns out to be abnormal, then a woman is being referred to a colposcopy to obtain a tissue biopsy, and at the same time or at a later stage may be treated with ablative treatment, which means excision of the abnormal portion of the cervix. So the goals of screening are to detect most of the high-grade lesions that are subsequently found on histology, to sort of enhance the prediction to be able to pick up those that are bad on histology, while minimizing the referral of women who will turn out on subsequent histology not to have high-grade histologic lesions, and then would undergo these procedures unnecessarily. Adverse events of colposcopy, biopsy or treatment are pain and bleeding, and possibly with subsequent pregnancy, cervical incompetence with fetal loss and prematurity, as well as the costs associated with the procedure. Cytology obtained from scraping the cervix is staged according to the Bethesda system for interpretation of epithelial cell

00047

abnormalities. A normal finding is one of NSIL, which stands for negative for squamous intraepithelial lesions. The next level of abnormality would be a report of atypical squamous cells, which could be one of ASCUS, atypical squamous cells of undetermined significance, and a later revision of the Bethesda system that also added the category of ASC-H, which stands for atypical squamous cells cannot exclude HSIL, and as you see on the bottom, HSIL is a higher grade of abnormality. The next highest grade is LSIL, which stands for low-grade squamous intraepithelial lesions, and these are reports of cytopathic

changes from human papillomavirus, and then the next highest level of known noninvasive precancerous lesions is that of HSIL, high-grade squamous intraepithelial lesions, and these encompass moderate and severe dysplasia, carcinoma in situ, or cervical intraepithelial neoplasia, CIN2 and 3. A cytology diagnosis can also be one that is reported as showing features suspicious for invasion, or actually of squamous cell

00048

carcinoma. So again, when normal cytology findings are seen, then the woman is referred, for a normal finding the woman returns back to be re-seen for the guideline recommendations. This, a finding of normal cytology is seen, but the co-testing algorithm is followed. And when there is a positive HPV screening test, then the current recommendations are to rescreen her earlier, not in three but in one year, or to go for specific HPV genotype testing. If the cytology is that of ASCUS with a negative HPV test, that's reassuring, and the woman is rescreened according to the routine guideline screening recommendations. However, if there's a finding of ASCUS with a positive screening test for HPV, that woman will be referred for colposcopy, as well as a woman who has a finding of LSIL on cytology would be referred for colposcopy. And again, if the woman has high-grade squamous intraepithelial lesions, she would be referred to colposcopy. The algorithms that are bolded here are those that represent clinical dilemma,

00049

because here we are either concerned about missing abnormal lesions on histology, as in NSIL with positive HPV, return to testing in one year, or in the second bolded, in the two last bolded options, the concern is one of referring to early and unnecessary procedures, because even though ASCUS and LSIL are worrisome, only a portion of the women who go on to have colposcopy will actually have abnormal histology that warrants treatment. Here you can see how the histological changes that are detected on biopsy are graded. They're categorized as cervical intraepithelial neoplasia according to the depth of involvement and the atypicality of the cells into three degrees of severity. CIN1 is considered a low-grade lesion;

it refers to mildly atypical cellular changes in the lower third of the epithelium. It may represent HPV and cytopathic events. Now CIN2 is considered a high-grade lesion, it refers to moderate atypical cellular changes that are confined to the basal two-thirds of the epithelium, with preservation of epithelial maturation. CIN3 is also high-grade lesion,

00050

but it refers to severely atypical cellular changes encompassing greater than two-thirds of the epithelial thickness, as well as full thickness lesions on histology.

Let me talk a little bit about the benefit, what the purpose is for predicting, and why CIN3+ is used as a surrogate outcome rather than apparent or invasive cancer. This is because few studies have sufficient numbers of cancer cases to assess cancer risks directly. The absolute risk of CIN3 including the rare cases of cancer, as combined as CIN3+, is considered to be the best measure of risk of incidence of cervical cancer. In many studies this is combined with CIN2, and as CIN2 because there are not enough numbers for even CIN3 invasive cancers.

Let me shift gears and talk to you about the test of interest for our technology assessment. It's in situ hybridization and I'm using FISH and ISH interchangeably unless specified. In situ hybridization is a technique that is used to detect and localize the presence or absence of a specific genetic sequence in cells using a complementary

00051

polynucleotide sequence, which is called the probe, and the probe is directly tagged with a fluorescent compound or it is indirectly visualized with antibodies that are then linked with the chemical tag and visualized under direct light. So you can have two methods of detection of this probe, either with UV light as in FISH, or with another method in chromatic in situ hybridization such as ISH.

Here you can see two pictures of a cell nucleus that is being subjected to FISH testing. As you can see here are two red dots, two green dots and two yellow dots, and here you can have an abnormal FISH test with three red dots, two green and two yellow. And as you can see here, the third red dot indicates an additional lesion, an additional region, while if there is only one dot, because you have two chromosomes, would indicate a deletion of a

genetic area. Most commonly used ISH tests for cervical cancer detect a gain of 3q26, which encodes telomerase RNA components abbreviated as TERC, which is activated early in the progression to cervical cancer, or a gain of

00052

8q24, which encodes myelocytomatosis oncogene abbreviated as MYC, which is a common site of HPV DNA integration, specifically for HPV 18. FISH probes can also detect the DNA for high-risk HPV genotypes, including HPV 16 and HPV 18. Currently FISH tests are being marketed and directly advertised by commercial laboratories for women with abnormal screening tests. This is our analytic framework which follows the ACCE model, and you can see here our first question was in women with, who are eligible to undergo screening for cervical cancer, what ISH tests are there currently available commonly examined in research studies that have been, have looked at their ability to detect abnormalities in cervical cells, which was our key question one. Key question two dealt with the analytic validity of these most commonly used ISH tests with regards to how chromosomal abnormalities or high-risk HPV genotypes correlate with the ISH test results. Key question three related to the clinical validity, how ISH tests correlate or

00053

are associated with subsequent histological findings for pre-cancer or cancer. And key question four related to the clinical utility in how ISH tests impact the clinical outcomes through diagnostic thinking, evaluation and management, and how they might impact the harms. We conducted searches in medical databases, MEDLINE, Scopus including EMBASE. Our last search date was in July of 2012 without language restrictions, using key words for our tests of interest, in situ hybridization, and for the disease, cervical cancer, pre-cancer, neoplasia and CIN. And we accepted all studies that had at least ten women with cervical tissue. Our horizon scan showed that the most commonly used approach are those related to TERC, as well as MYC, HPV 16 and 18, and therefore we focused our review on TERC, MYC, HPV 16 and 18.

For analytic validity we asked what is the association between ISH tests and reference tests, and expressed the agreement between tests as percentage with concordant results,

00054

and we found no studies looking at analytic validity for TERC or MYC, and 14 studies looking at agreement for FISH tests for HPV with reference tests for HPV either by polymerase chain reaction or Hybrid Capture 2. Agreement ranged from 35 to 100 percent and here you can see the percent agreement between ISH tests for HPV and reference tests in 14 studies.

I'm sorry, this is blinking. Is my time up already?

DR. REDBERG: No.

DR. UHLIG: How much time do I have?

DR. REDBERG: You have 20 more minutes.

DR. UHLIG: Okay.

So, you can see a lot of variability in the agreement between ISH tests and HPV reference tests, which we attributed to a lot of variability in the actual probes. Again, most of these probes, FISH test probes included probes for HPV 16 and 18, but also other probes for other high-risk HPV genotypes, and you can see there is clinical heterogeneity in terms of the reference tests here, some of them being

00055

Hybrid Capture or various types of polymerase chain reaction tests. We did not pool these data because of their clinical heterogeneity.

In terms of assessing the quality of reporting in the studies that we examined for analytic validity, we found deficiencies in reporting which we thought were most likely because the studies were not designed to specifically address analytic validity. The studies did not expressly describe laboratory procedures in detail because ISH tests and reference tests, most often the PCR assays, are well established in general, if not in particular for cervical specimens. Many of the reference tests were commercially available kits that probably included positive and negative controls, even though the studies didn't specifically state that.

This is really the core of our report, that focused on the clinical validity, and here the question, again, was what is the association between FISH test results on cytology and cervical intraepithelial neoplasia

high grade, or cervical cancer on histology? And again, based on the clinical dilemma, we

00056

were interested in stratified results by cytology findings for NSIL, ASCUS or LSIL with or without HPV, and our reference test was histological finding of CIN2+ or CIN3+. We accepted data on sensitivity, specificity, and conducted meta-analysis, five studies were available for any test outcomes here, and graded the quality of the studies according to the QUADAS 2 instrument. We found ten studies, all of them specifically used FISH, eight studies examined FISH testing for TERC, and you can see there that actually there was a fairly large number of individuals that were included in these studies. In particular, there was one study from China with 7,700 individuals. However, only, because we were interested in stratified analysis, we could really only include about a total of 600 with LSIL and 660 with ASCUS from this particular study, because we were interested in the results stratified by the cytological findings. Out of the TERC studies, five, the majority also used probes for, most of them used probes just for TERC, but three also

00057

combined probes for TERC and MYC. Three studies used FISH test for HPV, and unfortunately, not all studies gave us information on the HPV status according to the screening test for the patients. We found that in general, CIN3 results were consistent with CIN2 results, and I will show you, I will focus more on this in 2+ results. Here you can see how the evidence maps out across the different FISH tests and the different cytological stages. There was, the largest pocket of evidence was on TERC tests in women with LSIL for outcomes of CIN2+ or 3+, seven studies for CIN2+ with about a thousand patients in five studies for CIN3+, 900 patients, and for everything else there was really at most three studies, some of them significant in size. Here, this includes the Chinese study with about 600 patients, and for HPV FISH tests, fairly small number of studies and overall small sample sizes. Small sample size is important because it reduces the precision of the estimates. I'll walk you now through each of

00058
these rows. For FISH for TERC in women with LSIL, seven studies looked at CIN2+ with a thousand patients, and here is the first plot for those seven studies showing you sensitivity and specificity point estimates from the individual studies, along with a 95 percent confidence interval, and the summary estimate of sensitivity was .76, for specificity it was .79. And you can see here the meta-analysis, and you can see that the ROC curve fits pretty well but there are really only seven studies or only seven dots, and this is our summary estimate in the range of .75 or so of sensitivity and specificity for CIN2+, and the results for CIN3+ were similar, even though we only had five studies.
For FISH for TERC in women with ASCUS for the outcome of CIN2+, two studies with 790 patients, and because it's only two studies we did not meta-analyze them, so these are just the point estimates from the two individual studies with some variability here as you can see, both for sensitivity and specificity.
This is the forest plot for those two studies. This is the large Chinese study with
00059
over 600 people, very precise estimates, and this is one other study, this was done exclusively in women who were HPV positive, so you know, with two studies it's hard to say how consistent they are, but at least the confidence intervals are overlapping, and the same for specificity.
Now we're moving on to FISH tests for HPV in women with LSIL. Only three studies, each one with very small sample sizes, only a total of 38 patients. Again, no quantitative pooling. This is the range of the point estimates from these three studies and they're not very far apart, but again, with these small numbers there's a lot of uncertainty about these estimates.
Here you can see huge variability in the specificity, and for CIN3, again, it's less people with also a lot of variability in the estimates for sensitivity and specificity.
This is the forest plot for those three studies for the outcome of CIN2+; again, wide confidence intervals due to small sample sizes and a lot of variability, even though the confidence intervals overlap.
00060
FISH test for HPV in women with ASCUS,

only one study provided results for the outcome of CIN2+ with only 12 women, and found perfect sensitivity but poor specificity, and, you know, with two studies, there was a lot of variability here in the estimates.

And finally, one study actually combined in its FISH test probes for TERC or HPV and found a higher sensitivity but a lower specificity than the pooled estimate for TERC alone.

Two studies examined different test strategies and combined FISH tests with, for different probes or FISH test along with a hybrid capture HPV test, and they're interesting in that they show some principles for the testing. So this study here by Voss looks in women with LSIL for the outcome of CIN2+, at the sensitivity and specificity of FISH tests for TERC or MYC, and this is kind of -- well, this is a little higher than our pooled estimate and this is a little lower than our pooled estimate from the meta-analysis for the cytology stratum. And when you look at a FISH test that combines probes for TERC and MYC

00061

just like this, but adds in a probe for HPV and then the test is positive, if any of them is positive you can see that you push up the sensitivity but you lower the specificity.

And if you compare that against hybrid capture tests for human papillomavirus you can see 100 percent sensitivity and the benchmark for these tests is about 90 percent, the industry benchmark. But this hybrid capture test is not a FISH test, it has very low specificity, which is the whole reason we're looking at FISH tests to improve our predictive accuracy. So this principle here is if you add more probes, you'll pick up more abnormalities but you'll reduce specificity, and that's kind of in between the, compared to the hybrid capture test.

This is another study that looked at different test strategies, here looking at FISH for TERC and FISH for TERC or hybrid capture, so this combines the FISH and hybrid capture tests and compares it also to hybrid capture.

Again, hybrid capture has very high sensitivity, low specificity in this particular study, but if you add in the FISH test along

00062

with the hybrid capture you push up the sensitivity a little bit more, but again, you lower the specificity, and the FISH test by

itself has lower sensitivity but higher specificity. This is shown again here in this Chinese study for ASCUS, same principle, FISH test more specific and hybrid capture more sensitive, unless you combine it along with the two of them. So in summary, sensitivity and specificity estimates often had wide confidence intervals, even when we were able to combine estimates in meta-analysis, indicating considerable uncertainty about the tests to identify women with CIN2+ or CIN3+. Again, the largest pocket of evidence was for FISH test for TERC in women with LSIL for the outcome of CIN2+, seven studies with the pooled estimates of .76 for sensitivity and specificity of .79. We thought overall the strength of evidence was low, and one of the major flaws we thought was that the majority of studies did not stratify women based on HPV results. Now granted, that is because the treating

00063

guidelines have evolved to now include co-testing with HPV, which wasn't the case before last year. We thought there were a number of limitations to the evidence, and another important limitation really was that we thought most of these studies were done in convenience samples rather than true screening context, so we were unable to really use the prevalence of the high grade CIN to calculate positive and negative predictive value. Sample sizes were generally small, leading to imprecision, and there were generally few studies for each test outcome pair with the exception of TERC for LSIL. Reporting on items used for risk of bias assessment was often incomplete and another bothersome thing was that the threshold for test positivity varied across studies and point estimates were heterogeneous, so how a test was reported as positive varies across different studies, how many cells had to be abnormal for how many FISH signals. There were panels of HPV probes for the FISH test for HPV 16 or 18 that had

00064

considerable overlap but also irreconcilable heterogeneity, and therefore our confidence in the test performance of FISH was low, and we thought it was unclear how FISH adds to the evaluation of women tested according to current

guidelines, which as I said, now recommend co-testing at least in women 30 to 65, and these studies were mostly done before the updated screening guidelines.

We found no data for women with normal cytology and positive HPV screening tests, and we found no studies examining the association of FISH tests with clinical outcomes.

Our key question four dealt with the clinical utility and harms for ISH tests in cervical cytology, but again, there was no study comparing patient care strategies resulting from different tests, thresholds or combinations of ISH and non-ISH tests, or that examined testing strategies that included ISH tests.

Our conclusion was that the current evidence is insufficient to support routine ISH testing for TERC, MYC, HPV 16 or 18 in women with LSIL, ASCUS or NSIL on cytology, with or

00065

without HPV infection.

We identified the following evidence gaps. We found a lack of standardization of pre-analytic issues, thresholds, probe sets, controls, procedures. Meanwhile, there has been a nomenclature update both in terms of the Bethesda system for classifying cytology findings, which now divides ASCUS into ASCUS and ASC-H, and most of the studies that we looked at we think were before this division came into place, so ASC-H wasn't a specified category.

And now also for the histology staging, a new system is being endorsed by the LAST group, which stands for lower anogenital squamous terminology, which suggests now to triage CIN2 as an equivocal finding into either high-grade or low-shift lesions, and there's going to be a shift in terminology from CIN2 HSIL and LSIL, and this new initiative recommends to stratify CIN2 as high grade and low grade with the use of immunocytochemistry with P53 staining.

For the new testing recommendations, again, this is an evolving field and we don't

00066

think that the current studies were able to consider the use of HPV screening tests. But furthermore, we anticipate that there will be, the evolution in the HPV test in that not only will they be able to screen for high-risk genotypes, but also may be able to give you a more specific answer for specific genotypes

such as 16 and 18 in one go and, you know, that will totally change your a priori probability in how you stratify and triage the women. And finally, there were no clinical outcome studies.

We identified the following research needs. Really there is a need to standardize ISH techniques and thresholds. ISH tests are an emerging technology and should be looked at as add-on tests after Pap and HPV co-testing. They need to be studied in larger samples in which it then would be possible to compare clinical validity for different test combinations.

You've seen here how the combination of different probes change the sensitivity and specificity, so if you have a large sample, you may need to do lots of different analyses based

00067

on how you combine different probes, and that will allow you to compare clinical validity for different test combinations. There needs to be consideration of the impact of newer HPV tests and an interesting area would be to examine FISH or ISH in terms of its role to detect adenocarcinoma, which wasn't at all covered in the literature that we looked at.

Thank you for your attention.

DR. REDBERG: Thank you, Dr. Uhlig, very helpful. And now we will have Dr. Dorothy Rosenthal, who is a professor of pathology, oncology and gynecology and obstetrics at the Johns Hopkins School of Medicine. You have 20 minutes.

DR. ROSENTHAL: Good morning, everyone. I would like to thank the organizers for inviting me to participate in this, I've learned a lot. And I also want to thank Dr. Uhlig and her team that did a really superb job of exploring the literature, I also learned a lot from that, and that is the discrepancy between good science and what's in the literature, I think we can all agree to that.

At any rate, let me make sure I have

00068

all of my electronics in gear here. Much of what I have on my Power Point Dr. Uhlig has covered, and so I'm going to just point out a few of the most important features that I consider from a clinical standpoint. I will make a disclaimer right off the bat, I am a morphologist. When I first started studying as a medical student, we had 48 chromosomes, at least that's what was known, and I just

remember this Scientific American article that came out and said there are 46, so we went, oh, my goodness, and that was the beginning of the end, or the beginning, I should say.

I'm going to skip over all my disclosures and just bring us to where we are now with HPV and cervical cancer, cervical neoplasia. The main point that I want everybody to recognize is that most of the lesions that we see as cytopathologists and also as histopathologists are transient infections, and the most difficult, challenging part of this entire spectrum of natural history of cervical neoplasia is what lesions are going to progress, even to high grade, and which lesions are going to go from high grade to

00069

invasive squamous carcinoma. Just because a patient has a high-grade lesion confirmed by histology does not necessarily mean that she will ever go on to invasive cancer, and of course the major question is why don't all of these patients go on to invasive cancer. Whoever finds the answer to that question will probably get the Nobel Prize.

Now in the United States we have lots of good news. We've had, while not really organized screening programs, a lot of programs in effect that have screened large populations of women that have greatly reduced cancer of the cervix in women, but that's not so around the world, and even though our question today is for the American population, and CMS, everything that we study as scientists, please understand is a global issue, and hopefully will affect everyone around the world.

The Pap test has been lauded as the most important clinical and epidemiologic screening test that has ever been developed in the United States, it's drastically reduced the cancer incidence of the cervix, and so in the United States you have very few new cases per

00070

year and only about 5,000 annual deaths per year from the disease, at least half of which occur in women who have never been screened before, or have been screened only quite a few years ago. The test, as all screening tests, must be inexpensive, noninvasive to the patient, acceptable to the patient and the clinician. Unfortunately it has a very high false negative rate, and that is one of the reasons that the repetitive nature of Pap testing has been such a success. We also have

a slow growing biology of the disease and so those two combinations, frequent testing and slow growing lesions, has been one of the things that we have been able to capitalize on. However, it's a subjective test, the Pap test is subjective, and we need individuals who are highly trained and dedicated. There have been some improvements. The Bethesda system for a standardized terminology is one thing, but it hasn't really thoroughly defined what each lesion is. It's given us names for the lesions, but understanding what these lesions truly are and how they behave is something we're still

00071

discovering. We've also invented computerized scanners which have been helpful, but they still haven't given us the solutions to what are these lesions going to ultimately do to the patient. Liquid-based Pap tests have been a very helpful addition to our menu of testing but they really have not improved our pickup of precancerous lesions and cancer as much as we had hoped, and now we're doing HPV testing to triage patients to colposcopy. Our terminology we've gone over, I'm going to skip over this, except to just point out to you that in any population of women, the vast majority of them are going to be negative, and something I really want to emphasize to you is that if you have a false negative test and a false negative portion that we talk about, that's going to affect only a very small portion of your population, those in the SIL categories, squamous intraepithelial lesion category, or your cancer category, so around five percent of your population could be affected by a false negative. A false positive rate in a test is going to affect 95 percent of your population, and so this is where the

00072

financial burden comes in when you're looking at the performance of a screening test. Our definitions I will not belabor anymore, we talked about false negative and false positive. Here's our pyramid that came out of the ALTS trial. The 15,000 cancer cases, I'd say about 12,000 truly in the most recent data analysis, and if you look at the green base of this pyramid, this is where the problem really is, our ASCUS population and now perhaps if you want to add in the LSIL, that's a huge base. And in the ASCUS population, approximately 50

percent of those women will have an HPV positive test if done by a Hybrid Capture 2 test, which is a cocktail of HPV subtypes. In the low-grade squamous intraepithelial category, approximately 70 percent of those patients will have a positive HPV test, and so you can see that even though they have a positive test, it's not going to indicate their propensity to go on to cancer, so these are essentially false positive tests from a standpoint of disease and progression.

So if we have a population of 50

00073

million Paps per year and you do all the math, all the bean counting, you have an economic impact on the health care system of somewhere between 3 and 4 billion dollars a year, and to quote one of my favorites, Everett Dirksen, a billion here, a billion there, pretty soon we're talking about real money, as well as a psychologic burden both on the clinician and on the patient.

We know that human papillomavirus now is an essential but not only cause of our cancers of the cervix, and as I indicated, most of these are spontaneously resolving infections, especially in young women who are sexually active. The prevalence declines greatly after the age of 35, which is probably a combination of factors, including less sexual activity with age, as well as hopefully some degree of immunity, although the immunity produced by the virus, which is really only a cutaneous virus, is still not clearly defined.

And of course, the main question is why don't all of these infections go on to cancer, there must be something in the individual that precludes the development of the cancer.

00074

We have a variety of tests and again, the false positivity of these tests is one that we really have to consider, and want to make sure that it's not driving us to do more follow-up, management type of tests than what should be done. So that in any of these screening tests, negative predictive value is the most critical parameter that we can think about.

Now also when you're doing your screening tests, you want to consider the socioeconomic and cultural aspects of the testing. First of all, from a standpoint of the biology of the HPV, we know from longitudinal studies that most of the time if

it's going to clear, the HPV infection will clear between eight and 24 months, which means if you're going to repeat the HPV test in a shorter time period than eight months, you are going to perhaps be testing the same infection over again.

The other problem is that many of these young women have multiple HPV subtype infections if they've had multiple partners, and so generally we're not separating those

00075

multiple infections out. We know that a lot of these infections don't cross-immunize, so there's lot of questions, confounding factors in any of these studies that we're not teasing apart.

Patient compliance is a major factor. One of the reasons I went into pathology was a lack of patient compliance, rather than clinical medicine. We also have to consider cultural factors. Even in the United States where we're such a melting pot, we have to really consider what women are going to come in for a pelvic exam and Pap test versus a lot of other women who will stay away. Those women who don't come in and are screened and then develop their cervical cancer, we don't know about.

And then also, what are the availability of treatment options. I had an incredible opportunity years ago to go to South Africa, and I said why don't you have a major screening program, this is to the minister of health there, and he said if we screened every woman who is at risk for cervical cancer here, there's no way we could take care of them, we

00076

don't have the infrastructure for medical care in this country. And so you don't want to say to a woman, oh, guess what, you have cervical cancer, but, you know, the bad news is we can't take care of you, and that does happen in some well developed countries too, unfortunately.

So, the reason I'm standing in front of you is to help answer the question, does the Medicare population have an opportunity to benefit from the FISH test which we were asked to take a look at. First of all, before we take a look at the Medicare population, let's look at the distribution of cervical cancer in the United States. This is from the SEER data, this was published in 2011 from data that was gathered in 2010 over the period of time that you see up on the screen.

And if you look at these two charts really carefully, first of all, the chart on the left talks about new cases, new cervical cancer cases, we're not talking right now about pre-cancers, so this I want to really concentrate on for just a second and emphasize to you that if we have a new cancer, that means that these patients have been followed, that

00077

preceding this new cancer we're going to pretty much guarantee that she's had precancerous lesions over a decade, and this is where the Pap test may have come into play and picked up her lesion early. Or these new cancer patients may be patients that have denied their treatment, they have denied coming in for their Paps. Either way, a new case, you can bet was preceded by premalignant cervical neoplasia.

And so if you look at the age groups, most of your bell-shaped curve is going to be in the population of 35 years to 54. If you get up into the Medicare age group, 65 and over, I'm not talking about Medicaid, just Medicare at the moment, under 20 percent of your population are going to be brand new cancers.

Now swing over to the right-hand chart, and stage IA1 to IIA and then IIB to IVB. IA1 to IIA are staged at less treacherous cancers, if you will, and IIB to IVB are those that have metastasized and are going to most likely be lethal unless there is some very very traumatic treatment for the patient. So you want to get the patients early if you possibly

00078

can, and in the younger age groups, most of these women when they're picked up as cancers are in the early stage group, and it's only when they get older that they're in the higher age group for the most part. And so there again, look at those that are over the age of 65, most of them are, or at least 50 percent of them are in the higher age category, and so I really want us to consider our Medicare beneficiaries.

And then a very interesting chart to look at, I found it very fascinating, especially when you look at the proportion of males and females. I think it has something to do with death rate of males, women living longer than males are, but anyway, if you look at the Medicare beneficiaries and you see how many are females, which of course is our target in today's topic of FISH for cervical cancer,

you'll see that the vast majority of our Medicare beneficiaries are not really that affected by cervical cancer because most of it is in the lower age group. And then those women that are in the Medicaid population, while you don't want to ignore them, they're

00079

younger than age 65, which is the screening cutoff, the recommended screening cutoff, and that's a very small proportion of women. So once again, look at our female Medicare beneficiaries. The grand total is in the gray bar for the age over 65, 22.5 million, those that are disabled 3.8. The orange bars on both the aged and the disabled are the younger group, and the older ones, up to the purple, are a very very small percentage of them. So that we really, the older these women get, the less likely they are going to be beneficiaries of a FISH screening test. Now the screening guidelines we already talked about, thanks to Dr. Uhlig, and I think we can just go right on beyond that. The intervals for screening are only if there have been negative tests in the past, so anytime there's a positive the interval is shortened, and this is something that we really have to pay attention to. The thing that disturbs me a lot as a cytopathologist is we are assuming in any of these guidelines that the Pap test is perfect, it's not. We talk about the false negative rate but it's really

00080

not a consideration when people start talking about intervals and extending them, and so you really have to know where your Pap test is being sent, and I will say no more about that. Okay. Again, the technology assessment was wonderfully done, and thank you, Dr. Uhlig, for presenting it so nicely for us. And so here we are with the four questions and I wrote this during Passover, and if you know Passover at all, we have four questions for Passover, so I thought my goodness, here we are, you know, pass the matzoh. At any rate, these are the typical questions that we're asking for our technology assessment, and the main thing that I want to really stress is the clinical validity and the clinical utility. The literature search Dr. Uhlig went through very very nicely with us and I'm not going to belabor that anymore. The TERC, the telomerase component is fascinating. As we all know, telomerase

enables cells to either become immortal or to have a life span that is finite, and so one of the things about cancers is that they theoretically become immortal, although they do

00081

have a life span of their own, but they seem to be immortal. And so any test that can show a gain in TERC is going to indicate that this patient has a tumor that is going to develop and continue growing beyond the constraints of normal tissues, and so if you find a high-grade lesion, either high grade intraepithelial noninvasive or cervical cancer, with a gain of TERC, you know this patient could be in trouble unless you completely eradicate the cancer.

What I look for, and I'm going to continue to look for in any articles that come out, is normal tissues in the same patient that has the gain of TERC, have those been sampled, because we know there must be something within individual patients that enable them to harbor a lesion that is growing and is going to develop into cancer, and I'm not seeing evidence that these patients have other tissues within an area that's nearby, there's tumor that has been sampled, and I think that could be a fascinating study for anybody sitting there in the audience.

DR. REDBERG: Two more minutes.

00082

DR. ROSENTHAL: I'm just about ready to go, thanks.

The studies that Dr. Uhlig talked about, I really can't add anything else to her analysis. They are not very well done from an analytic standpoint and I don't think we have evidence yet. And then when we get to the real struggle, ASCUS and LSIL, the sensitivity and specificity is way beyond what has been considered acceptable. The sensitivity and specificity of one is of course wonderful, but most of the studies don't have it, so they don't really qualify as a good screening test at all, which would be in the .9 range for both sensitivity and specificity, and the same thing for the clinical validity.

And then the CIN2 issue is something that is undergoing a lot of change. Dr. Uhlig mentioned this. CIN2 has always been pretty much of a fence-sitter between low grade and high grade, and is going to probably result in pushing it down into the low grade, which means we're going to have even fewer studies with a

high-grade endpoint, which is unfortunate. And the same, the TERC has very very

00083

few studies that we could really assess the analytic validity and clinical validity of the FISH study.

I'm going to go right on ahead.

There's nothing that would describe the harms of the tests to the patients. The research gaps, Dr. Uhlig also clarified very very nicely, but something that we really haven't talked about is the effects of vaccine on the natural history of HPV, infections we're going to have to really pay attention, so this is a moving target unfortunately, as most of the scientific research these days is.

And so in answer to the question, are we ready yet, I really don't think we are. I think it's a fascinating test that we might want to use as an adjunct. As a screening test, I think there's not enough evidence, and it's also a very expensive test and a very work intensive test unless it's automated.

And so all of these things, lack of automation, lack of analytic clinical validity and clinical utility, as well as the subjectivity of finding the abnormal cells and then counting whether or not there's

00084

overexpression of TERC, we're not there yet, it's a long way to go. Plus, the people who would be benefitting from it in the Medicare population are really quite low in number compared to other ways in which we could spend our Medicare dollars.

So, that's all I have, and I will answer questions later on. I think we have a break now.

DR. REDBERG: Thank you very much, Dr. Rosenthal, and thanks to all the morning speakers, I think we know a lot more about cancer and genetic testing. We will now take a 15-minute break and return at 10:15 promptly.

(Recess.)

DR. REDBERG: I'd like to welcome everybody back, and we are going to start the second half of the morning with our public speakers and I will introduce the first one, Dr. Robert Wassman, who is the chief medical officer of Rosetta Genomics, and each speaker will have six minutes. Dr. Wassman.

DR. WASSMAN: Thank you very much. I want to thank CMS, the committee and the chair, for allowing me to address you today with my

00085
thoughts. I want to focus my thoughts around three general themes, one of which is that diagnostic improvements and improvements in care happen incrementally. Two is, they are founded on basic science that has to be very solidly doing it. And third, in the end, the important thing is the patient experience an improvement in the outcomes for those patients on an individualized basis, not just on a statistical basis. And as identified, I'm the chief medical officer at Rosetta Genomics, which provides one of the tests under question in the CUP category here today.

I would like to start this by pointing out that, you know, historically the challenge of cancer of unknown primary has always been the same, it has been to advance our diagnostic acumen such that we can reach the point where we can avail these patients of the choice of the best known therapy at that point in time based on a correct diagnosis. The difference today is that with molecular profiling we now have the best single diagnostic test possible, or historically possible for these patients. And when we consider it, cancer of

00086
unknown primary, all patients who present with metastatic cancer are essentially cancers of unknown primary to begin with, until we gradually whittle away with successive uses of technologies that evolved over time to a diagnosis. However, prior to molecular profiling, a significant percentage of these patients were left in limbo without a diagnosis, and today a significant number of them, over 90 percent of cases do reach a conclusion, so their oncologists and their pathologists have these remaining questions answered for them or substantially answered for them, to allow them to make choices about therapy based on the current best guidances we have for therapy. For some of those diagnoses, that's going to result in potential for very significant differences in what the expected outcomes based on patients with these cancers are, a generic therapy versus therapies that are specific for those cancers.

Now, our approach to molecular profiling CUP is one that uses a very unique class of biomarkers, microRNAs, and these are based on a very large biology, basis of basic

00087
biology, and they have been shown to be highly

sensitive, highly predictive of tissue of origin, highly reproducible, which is very important in a clinical diagnostic, and importantly too is they're foundational to the biology of cancer. They are things that underlie the whole process that evolves into cancer, and therefore are not secondarily removed from the question at hand with the patients.

When that science is translated to a clinical test, it results in a clinical tool that performs, is very very respectable. When we deliver a single call on a patient coming out of the algorithm of our test, which happens in over 82 percent of patients, there's a 90 percent sensitivity for that single call. Two-call results are more categorical, we had a slightly lower sensitivity as a result, but overall this performance is far superior to the tests that were available previous to patients, and allows us to fingerprint the cancer very effectively.

Now, the challenge in this area is that there is no gold standard that's

00088

identified, so we have to use different approaches to identify what the accuracy of the test is. What I'm going to be looking at is comparing the correlation between known primary metastases and primary tumors, and that correlation, as we see here, is very very close.

In addition, this study also highlights something that's very important thinking about the clinical utility of the test, and that is that they were unable to find this hypothetical CUP as a distinct entity. It's been looked for and hypothesized about for a long time, that CUP is somehow different than other cancer. And when they looked at it and looked at the individual cancers, there was no distinct biological marker identification for CUP as a separate class. Therefore, no justification is apparent in the scientific literature to consider treating CUP as anything other than the tumor which you identify based on its primary origin, which is how we treat all other cancers today.

The current platform has been based upon this basic biology, it has been based on a

00089

first generation test and then on a large training set and a 500-patient clinical validation. In addition, we've extended that

to three separate studies at three separate institutions showing concordance with the final clinical pathological diagnosis at those institutions. Most recently, in 84 patients from the Greek Center that several papers were referenced this morning to, in which they showed a 92 percent concordance between our final clinical pathological diagnosis and all of the testing that could be done on that patient.

So fundamentally, you know, patients with CUP really face a really challenging cancer experience, in part because there isn't some of the doctors confident in their treatment, and this is really adversely affected by the limitations of our diagnostic capability historically.

DR. REDBERG: You have to wrap up.

DR. WASSMAN: Thank you. Do I have a minute more?

DR. REDBERG: About 30 seconds.

DR. WASSMAN: Okay. I just want to

00090

close by pointing to the fact that we regularly experience cases like this, where we see patients who have doctors believe they know the diagnosis, the patient does not respond to therapy, and the doctor says well, maybe I don't know what the diagnosis is, and we define the very distinct diagnosis. And the question to be asked of this patient is, had this test been available to the patient sooner, or had we thought of doing this test sooner, to identify this patient did not have breast cancer but in fact had melanoma, would that patient have had guidance to go on newly approved drugs that are targeted to a mutation which could have been subsequently tested, and had a significantly different outcome in their case.

Thank you very much.

DR. REDBERG: Thank you, Dr. Wassman.

Next is Dr. Margaret Havens Neal, who's a pathologist with KWB Pathology Associates in Tallahassee, Florida.

DR. NEAL: Good morning, I am Dr. Margaret Neal, I am a pathologist in a large private practice group in northwest Florida, we do anatomic and clinical pathology,

00091

I'm a fellow of the College of American Pathologists, I have no conflicts.

The College of American Pathologists appreciates this opportunity to speak before MedCAC on these important issues. The CAP is

well qualified to make comments here as they are the worldwide leader in laboratory quality assurance, and a leader in laboratory accreditation. They serve over 18,000 board certified pathologists and are supported by a wide range of very expert scientific committees.

Cancers of unknown primary are a very important clinical entity. The American Cancer Society estimates that over 32,000 new cases will be diagnosed in the United States in 2013. Currently determining the tissue of origin is very important in these poorly differentiated tumors because treatment is often dependent on the primary site. In recent years we have become much more adept at identifying primary tumor origin by the imaging that's more sensitive, and our routine histologic exam has ever more powerful immunohistochemical stains that are more specific, as well as our

00092

molecular testing.

The available literature suggests that microarray-based gene expression may help identify some of these unknown origin sites in cases where other traditional ways have been exhausted. If used in the way where the other traditional ways have already been explored, and most likely in the hands of a pathologist who is evaluating those questions, the real volume should be low.

These CUP tests are laboratory developed tests, they are proprietary, and for that reason quality assurance issues do arise. They are not part of the FDA process, there are no proficiency tests out there that could be used for these tests, and developing those are very difficult in a proprietary environment. Test validation and independent peer review may have bias associated with these proprietary tests. And in addition, these are expensive and labor intensive tests, and resources might be better used elsewhere.

Currently there are no prospective trials that indicate that microarray-based gene expression is more accurate than conventional

00093

methods for determining primary site. We have one available prospective trial which has a limited number of patients, 252, and a very wide variety of disease sites, 26, with the survival differentiation measured in months. So at this time there is insufficient data to indicate that treatment based on microarray

gene expression results in clinically significant improved survival or improved prognosis.

Currently there is no role for FISH HPV testing in cervical cancer screening. The multi-organizational consensus group that recommended guidelines for cervical cancer screening specifically recommends against using non-FDA-approved tests and does not endorse primary HPV testing. FISH HPV testing is still in clinical trials and there are no standardized algorithmic guidelines for its use.

FISH HPV may have some role in head and neck carcinomas, but again, we do not have literature to support that there is clinical significance.

The CAP would like to thank MedCAC for

00094

this opportunity to offer comments and to hear the excellent discussion from this morning. We support accountable, high quality and cost effective patient care testing, and look forward to future research in these areas that can focus on patient benefits with effects on treatment decision and outcome.

Thank you very much.

DR. REDBERG: Thank you, Dr. Neal.

Next is Bernard Berins.

MR. BERINS: Good morning. My name is Bernard Berins and I have no conflict other than the fact that after I received the results of the Rosetta Genomics miRview mets2 test on my tumor, I bought 500 of their nine million shares of stock.

On July 15th of this year I will be 74 years old. I have been married for over 45 years, have three grown children, three grandchildren, and I have practiced law full time with the same firm in New Orleans for over 50 years. A little more than a year ago as I approached my 73rd birthday, I remember reflecting on the fact that I had reached my 70s with no major health issues.

00095

Over three years ago my oldest daughter, then 38, was diagnosed with triple negative breast cancer in her right breast. After hearing the diagnosis, it took us no time to spring into action. As is the case with any issue affecting my family, I took the lead in identifying the experts in the field in order to find treatment for her situation. Shortly after discovering the lump in her right breast

she underwent a lumpectomy and started off the schedule of chemotherapy followed by radiation treatments.
Almost two years ago, one year after her initial cancer diagnosis, we learned that this aggressive cancer had returned, now presenting with a tumor in her left breast.
Again, we sought expert advice and did our research, and agreed that a double mastectomy followed by more chemotherapy and more radiation would be the best way to reduce her risk of another recurrence.
A little more than a year ago I heard the words again, this time it was you have cancer. While it is not something you get used to, my family and I have been through this

00096

before, and waited the results of the tests so I could start treatment as quickly as possible. Unfortunately it was not as forthcoming this time. Little did we know it would take more than eight months to confirm it since the original diagnosis, cancer of unknown primary origin, or CUP, was not a clear conclusion. What we did know was that the cancer had already metastasized to the lymph nodes in my abdomen, groin and neck, while the primary site was not known.
This time I was facing a battle and had little or no information to guide my doctors and me. Frankly, I had never heard of unknown primary cancer, nor did most of the people with whom I discussed my illness. Many thought I was joking. Who knew that in this day and age that a patient could be diagnosed with metastatic cancer, yet the primary site of the disease is not always obvious. In fact, as I have learned since, I was one of the approximately three to five percent of patients, cancer cases diagnosed as CUP annually in the United States.
Of course my initial concerns grew

00097

deeper when I learned that generally speaking, the best method of treating a cancer is by using a chemotherapy medication that attacks the primary cancer. Without identifying the primary cancer, an oncologist, even one who specializes in unknown primary cancer, sets the treatment plan on some medical assumptions and perhaps a little guesswork.
In June of 2012 on the advice of my local oncologist, I commenced visits to M.D. Anderson Cancer Center in Houston and have

continued almost every ten weeks since. As part of my care, my M.D. Anderson oncologist sent a tissue sample from my biopsy to bioTheranostics in San Diego, California. bioTheranostics molecular tumor profiling determined that I had a 55 percent probability of breast cancer, a 31 percent probability of salivary gland cancer, and an eight percent probability of squamous cell carcinoma.

DR. REDBERG: One minute more.

MR. BERINS: Based on the bioTheranostics probabilities, the M.D. Anderson oncologist started me on a cycle of chemotherapy with two medications attacking

00098

primarily breast and salivary gland cancers while my family and I hoped the specialists would offer more insight, but many questions remained. Later I, in this very serendipitous occasion, in browsing Facebook in the middle of the night during one of my bouts with insomnia, I discovered that a cousin, Kenneth Berlin, was the president and CEO of a company called Rosetta Genomics. I had never met or talked to Kenneth in my life, so I was interested in my long lost relative and curious as to what Rosetta Genomics was all about. So I typed in their website and much to my surprise, I learned that one of their products identifies primary cancers that were previously determined unknown.

Shortly I reached out to Mr. Berlin, and arrangements were made through my local oncologist to send a tissue sample to Rosetta Genomics for this testing. Their test reflected to a 90 percent certainty that my primary cancer was breast cancer. The confirmation came at the best time possible, just days before learning that my original chemo regime was not as successful as expected.

00099

In January 2013 scans showed that not only had the metastatic cancers in my abdominal lymph nodes grown, but also showed evidence of metastatic cancer on my spine.

DR. REDBERG: Thank you, Mr. Berins.

MR. BERINS: Thank you.

DR. REDBERG: Our next speaker is Dr. Anthony Greco, Sarah Cannon Cancer Center and Research Institute.

DR. GRECO: Thank you very much. I'm a medical oncologist in Nashville, Tennessee. I have been seeing and evaluating these patients since 1976, written a few articles on

them. I appreciate Dr. Conley's review, it was an excellent review, and will knock a minute off my talk today.

This is a difficult problem, to say the least, and I'm going to move -- it looks like I have Dr. Weiss's sides up there, though, just a little technical error. The previous speaker had no slides, you need to go back.

Anyway, the issue of diagnosing the type of cancer is extremely important in patients with unknown primary cancer. That's how we treat patients with cancer. If we don't

00100

have a diagnosis, we're in trouble, we either send them to hospice or we treat them with empiric therapy.

Now, I see them up here but I don't see them over there. There you go.

A couple of the first slides, I don't need. I do speak for bioTheranostics and receive honoraria for giving talks, many of their CME talks.

I won't talk about the incidence of this disease, you've heard about it. The main issue here is that when patients are evaluated, you can't find the anatomical primary site despite the fact that they have metastatic cancer; this is a problem, as you might imagine. We work those patients up rather extensively, you can't find the primary.

What have we done in the past? Well, we've used shotgun therapy for these patients, we use broad spectrum antineoplastic drugs, sort of like broad spectrum antibiotics. Not a good idea now, not a bad idea 25 years ago, we couldn't treat most solid tumors very well.

Now individual treatments for solid tumors is different, kidney cancer, breast cancer,

00101

melanoma, lung cancer, they're different, so knowing the answer to what type of cancer the patient has is not academic anymore.

Now if you want to learn how accurate this test is in unknown primary cancer, I would encourage you to read an article to be published June 5th of this year in the Journal of the National Cancer Institute looking at this particular cancer, type IVR22CR, and the accuracy in unknown primary cancer. That is an important question. There are really only two questions, the accuracy of the test and the outcome of the patients.

If one looks at treatment in large numbers of patients, mainly done by our group,

which is a cooperative group in treating these patients over the past few years, you will see that the median survival of all the patients, grouping them all together because we didn't know what they had, is about nine months. This is pretty rock solid, particularly when you exclude the favorable subsets of patients, those that Dr. Conley went over briefly earlier.

Now we designed the study. Myself and

00102

my associate, Dr. John Hainsworth, had nothing to do with bioTheranostics. We went to them because we needed their tests, because we felt like it was probably going to help us improve the outcomes of these patients. The difficulties in this type of study was outlined beautifully by Dr. Conley and I won't go into that, but we wanted to look at a prospective study where patients were diagnosed with this molecular test and then treated according to what the molecular test said, and we wanted to compare that to historical controls that our own group had developed in nearly 400 patients in just the preceding years. We wanted to see a 30 percent increase in median survival compared to those controls, and we also wanted to compare the more treatable subsets defined molecularly versus those we would expect not to do as well defined molecularly.

So basically this is the study design. The patients had CUP, they had this test done, and then the patients were treated according to the molecular diagnosis. This was published in the Journal of Clinical Oncology. I won't go over the treatments, but they're just more or

00103

less standard therapies for those particular cancers. Remember, all the patients had CUP, but they were being treated individually.

This just shows you a flow diagram. 289 patients in this trial, a large Phase II trial. Some patients didn't have sufficient tissue, but 194 patients had an assay done that we could then give prospective site-directed treatment to. It's important to realize that 115 of these who were molecularly diagnosed with cancers we would expect to do better than the 79, and I'll show you those groups in just a second.

These are the molecular diagnoses. 98 percent of the patients had one diagnosis made in molecular fashion, and you can see that nearly half the patients identified would have

a molecular target available in treatment if in fact they had that type cancer.
DR. REDBERG: One more minute.
DR. GRECO: Okay. This shows the survival of all patients compared to historical controls. You can see -- and keep in mind, this is group data -- it met expectations. The median survival was about three-and-a-half

00104

months greater for the blue curve versus the yellow curve. Probably more important is when you look at the more responsive ones, colorectal, breast, ovarian, kidney, germ cell, et cetera, versus the less responsive defined molecularly in the left-hand column, you can see there's a substantial difference in survival, the curve separates all along with a P value of .04.
So it's clear to me and most of us in this field, this is penultimate data, this is prospective outcome data and you're not going to have any better data than this unless you have the ultimate study, which I can tell you is not going to be feasible in the United States. This test needs to be done when you can't diagnose the patient by immunohistochemistry, so then you can give specific site-directed treatment. Thank you.
DR. REDBERG: Thank you, Dr. Greco.
Next is Dr. Lawrence Weiss, who is chairman emeritus of the department of pathology at the City of Hope National Medical Center, and a senior consultative pathologist at Clarient.
DR. WEISS: My interest was conducting

00105

a study on the comparative effectiveness of gene expression-based cancer classification versus a standard of care in immunohistochemistry, and we recently published our results a few months ago earlier this year in the Journal of Molecular Diagnostics.
My conflicts of interest are given here. Actually, I'm an employee of a corporation called Clarient Pathology Services, but we have an exclusive contract with Clarient. I will also say that I am a diagnostic pathologist, board certified, a member of the College of American Pathologists, and much of my daily practice is in trying to diagnose carcinomas of difficult origin.
So, the standard of the art is right now morphology combined with immunohistochemistry. It interprets histochemical stains at varying sensitivities

and specificities for various organs, occasionally with very significant cross-reactivities. They're not applied in a standardized fashion, different pathologists of different skills apply different antibodies in different ways in different laboratories.

00106

A meta-analysis showed immunohistochemistry had an accuracy of 66 percent. This meta-analysis may have been recent but it was mostly based on older studies, so it's not clear whether it's the same or different. And so, trying to diagnose difficult to diagnose origins is a definite area in pathology that is a significant problem.

So again, the objective in my study was to compare immunohistochemistry to one of these more elaborate molecular studies, and we used cancer ID tests from bioTheranostics. Basically it was a prospectively designed blinded comparative study using retrospective studies' so-called tissues of convenience. We prospectively looked at the diagnosis and only chose cases where we had a diagnosis beforehand, and we did this on the basis of all the information available to us. Many of these patients had been around for many years and were well known by the clinicians for having a primary site.

We had two study arms, one molecular studies, one immunohistochemistry. We

00107

developed the protocol at City of Hope and didn't book any interference from the company even though we needed the company to bankroll the study because we couldn't do, you know, 120, 130 studies without financial support. The cases were, again, selected on the basis of being a challenging primary site, formal and fixed paraffin-embedded tissue available. Because they were challenging cases, they were primarily high-grade tumors, 90 percent were metastatic. We threw in some primary tumors, particularly in lung, so people could just assume it was a metastasis and not a primary from that site. Again, the reference diagnosis was established by clinical correlation, including access to radiology, clinical charts. Many of these patients had been known at City of Hope for many years.

Specimens were blinded, coded and sent to the two separate sites for the diagnosis, and the predictions were analyzed by a

third-party statistician, trying to keep it as unbiased as possible. Here's the basic results. CancerTYPE ID demonstrated an increase in overall accuracy

00108

of 10 percent compared to immunohistochemistry, it was 79 percent versus 69 percent. The accuracy was good for both technologies when there were good antibodies that are relatively organ-specific and organ-sensitive, such as GI, lung and kidney. Immunohistochemistry was not so good when the organ-specific antibodies did not have as high sensitivity and specificity, such as for bladder most notably, but also for breast, and remember that breast is a very actionable tumor.

DR. REDBERG: One more minute.

DR. WEISS: Both technologies made the correct prediction, the same correct prediction in 65 percent of cases. On the other hand, CancerTYPE ID got 14 percent of cases that immunohistochemistry did not, immunohistochemistry only got five percent that the cancer ID did not.

Summary, results from this blinded comparative effectiveness study demonstrated somewhat superior accuracy for the gene expression-based classification. Whether this is due to the greater number of genes that are looked at or that RNA retains its

00109

differentiation better than the protein expression, it did prove better.

It's important to note that the cases selected for this study are not representative, we chose the hardest and the most difficult cases. Therefore, the percentages of accuracy may be lower than some of the other studies.

The results of this study show that a significant number of patients may be at risk for misdiagnosis, and gene expression classification demonstrates its clinical value with improvement of diagnostic accuracy over standard of care.

What we're trying to do is eliminate carcinoma of unknown origin. Over the years we've reduced that, and molecular tests are going to help us do that. I might add that the immuno studies, we gave the pathologists up to 15 blanks, they used a mean of about nine stains, and immunohistochemistry is pretty good, but once you go over nine stains it gets harder to make a diagnosis because you have less good antibodies, so the two technologies

can work hand in hand. And as a pathologist, one needs to be technology agnostic, and use

00110

the best technology that you possibly can for your patients, and I think in some cases it will be immunohistochemistry, and at other times it may well be gene expression profiling studies.

Thank you.

DR. REDBERG: Thank you, Dr. Weiss. Our next speaker is Dr. Catherine Schnabel, who is the vice president of medical, clinical and regulatory affairs at bioTheranostics.

DR. SCHNABEL: Good morning, I'm Catherine Schnabel, I handle the clinical development and medical direction at bioTheranostics, which is the developer of CancerTYPE ID. I appreciate the opportunity to comment today on the significance and clinical impact of these molecular classifiers.

In the current practice of personalized medicine, individualizing care really means knowledge of the lesser attributes of the tumor in order to define clinical subsets that would be responsive to clinical therapies. Tissue of origin plays a fundamental role in this practice because it provides a cellular context that actually

00111

determines which predictive markers will be relevant, and ultimately what therapies will be efficacious. You can see underlying biologies if specific tumors are tissue specific.

The current unmet need is really based on the limitations of standard of care, in particular immunohistochemistry, which is really the cornerstone of tumor classification to date. It is subjective in its approach and its interpretation, and there are numerous studies which document several clinical scenarios where there's a lot of interpathology discordance arising to a definitive diagnosis.

So who are these patients that would be clinically impacted by this type of technology? On the low end are patients that are diagnosed with actual cancers of unknown primary, those are in the tens of thousands, and those are patients that lack site of origin despite comprehensive and exhaustive workups. On the other end are really the ones in the disease state that is really more expanded, which Dr. Weiss represented, the difficult to diagnose cases, and these are patients that actually after workup have a tentative,

00112
uncertain or nondefinitive diagnosis. And these patients may face suboptimal care because no rational approach can be applied, because the tissue of origin has not been identified for these patients, and as more targeted therapies and site-selected therapies are available, these technologies will become more relevant.
So, I would like to make comments about really the collective evidence for these technologies today. These were summarized in great detail by the technology assessment. The only couple of points that I will make is that these technologies have been systematically investigated through pillars of evidence, four pillars of evidence that really underlie evidence-based diagnostics. There are thousands of patients that have been investigated, and several of these studies have been published in peer reviewed journals that actually, where practice changing information is communicated to physicians.
The other point that I would make is that one of the limitations currently in these technologies was that there were numerous
00113
company-sponsored studies. I think we've heard that these protocols have been created and developed in academic centers where involvement of the industry sponsor was minimal, despite what was noted in the technology assessment.
The other point that I will make is that molecular cancer classification has recently been incorporated in consensus guidelines for patients with cancers of unknown primary.
Speaking specifically of CancerTYPE ID, of the three technologies described today, this is one that is based on PCR-based gene expression profiling. The advantage of having a PCR-based platform is that the biospecimen requirement is minimal. The advantage of that is that we all know that with tissue, tissue-based diagnosis, the resources of the biospecimen is precious, and so there's the added advantage of complementing this technology because the cellularity requirements for the technology are very very small, so that allows cells to be saved and tissue to be saved for downstream testing, and more information to be gathered about the tumor.
00114
I would also make the comment that

these tests have a very very precise turnaround time, which is very impactful for these patients, because time to diagnosis is critical for enabling them to get the care in an expeditious manner. The turnaround time for CancerTYPE ID and other technologies in this class is between five to seven days.

This is data we and others have shown, comparative effectiveness. The other thing that the technology assessment noted was that there was low evidence in the area of clinical utility. I would submit and argue that any test that basically goes up to standard of care and shows increased performance characteristics over current standard of care builds a strong case for clinical utility. As I said, we and others have demonstrated that an absolute improvement of over ten percent in diagnostic accuracy, what that means for patients is there would be an increased chance that there would be less misdiagnosis using a diagnosis that is standard and objective.

These are data that were presented by Dr. Greco, and really what was missed, I think,

00115

as well in the technology assessment, is that this is a prospective study where patients were recruited and then treated with site-directed therapy based on the molecular test. These are data that were shown by him, patients that were chosen to have, or selected by the CancerTYPE ID test to have more responsive tumors, showed a statistical significance in overall median survival, which is the hardest endpoint for clinical outcomes and clinical utility.

I will skip this and make concluding remarks. Really what I want to impress upon this group today is that the totality of evidence is really not directional towards a technology that's in its infancy, but really more of an evolving, rapidly evolving arena, and that the technologies that are being investigated today have both the patients and physicians in mind, and that these molecular tests have been routinely integrated and adopted in routine clinical practice, and of course the majority of those patients are Medicare patients.

Thank you for your attention.

DR. REDBERG: Thank you, Dr. Schnabel.

00116

I believe we have one more speaker before the panel is able to ask questions. So, Mr. Dan Jones, from Quest Diagnostics. You have one

minute.
DR. JONES: Hi. I'm Dan Jones, the medical director for cancer diagnostics at Quest Diagnostics, and I also run a pathology group right up the road in Chantilly, Virginia. And I really, without Power Point, wanted to make the perspective that not all health care is as coherent as happens in a regional cancer center. Prior to joining Quest I ran marker diagnostics at M.D. Anderson Cancer Center. Everything was very coherent and logical. Most cancer care is delivered in and out of the reference lab settings in the country, and a significant amount of GYN pathology is delivered there. We need tools to be able to help us in diagnoses where we have minimal information, and right now for the cervical biopsies, colon biopsies, where they see a lesion and they're not sure what they're looking at, they may or may not know what the Pap smear shows, we need such tools as the situ hybridization for HPV, and TERC, FISH and other

00117

tools that are being developed. Because if we don't have those tools, we have a limited ability to do more than just say I'm not sure if this is reactive, or I'm not sure if this is a lesion, and that really doesn't benefit the patient, and it doesn't benefit the clinicians that are treating.
So, I agree with the studies that were shown, we have seen similar studies that were shown this morning, but we have to think about how care is actually delivered for cancer patients in this country, it's not as coherent as we'd like to assume in a regional cancer center where you can do very well controlled studies. I just want to bring that perspective. When you think about these tests, we need these types of tests in the outpatient setting to help make accurate diagnoses.
Thank you.
DR. REDBERG: Thank you. I want to thank all of the speakers on behalf of the panel, and ask if you could now come up to the front row, and the panel members can now start asking questions, and if you'd just signal to me, I will recognize people in order. So,

00118

Dr. Stecker.
DR. STECKER: Thanks to all of the speakers for helping to distill and clarify and give a broad perspective. I had two questions, one for Dr. Whitehead.

The first is about how in the analysis, how your, how the quality of evidence is created. I'm particularly interested in survival data with regard to clinical utility, and this is regarding cancer of unknown primary. I see there are four studies that report on survival. Two have no control group, one has a historic control group, the fourth doesn't report on what the control group is, or you didn't. I note that one of the ones without a control group, Panto in 2003, is rated as good evidence, and so how does an outcomes trial without a control group rank as good evidence?

DR. WHITEHEAD: Many of the studies looked at more than one thing and so they were included if they provided, you know, evidence on a question, but they may have been graded on the overall, you know, the primary point of the study. If you will give me --

00119

DR. STECKER: Yeah, I'm sorry to delve into this. It's PDF page 87 and your page 70.

DR. WHITEHEAD: Okay.

DR. STECKER: And this is one of many things you reviewed, so I understand it may be hard to pull up. In general I was wondering, you know, since, from a clinical research standpoint, survival or outcomes analysis, you know, the primary thing determining is what's the quality of the control, is it randomized with a quality control group, so I was confused about how in a survival study, something reporting on survival without a control group was rated as good.

DR. WHITEHEAD: You said you were on page 87 in the report?

DR. STECKER: PDF page 87, report page 70.

DR. WHITEHEAD: Give me a chance to review this. Okay. That study is actually not -- this is a cytogenetic study, so it is not one of the three molecular tests that I focused on today. The primary purpose of the study was to look at how well the cytogenetic analysis contributed to the diagnosis, and for

00120

that purpose of that design it was well designed, so it got rated as a good study for the primary purpose of the analysis.

DR. STECKER: Because in the table it says outcome median survival, so the purpose you reviewed it for was not survival?

DR. WHITEHEAD: Right. It was graded,

the quality of the study was graded on the clinical validity of the study, which was its primary purpose. It also reported this data on survival and so we included that data here, and that's the reason that it wasn't graded as a poor study, because that wasn't the primary reason it was designed for, it just reported some evidence for something.

DR. STECKER: Fine.

DR. REDBERG: Dr. Sartor.

DR. SARTOR: Thank you. This will be for Dr. Greco, thank you for performing a prospective study. I have some questions about the design. So within the conclusions, there's several. Number one is that against a historic control, that patients with a site-directed therapy did better, and I was wondering a little bit, why not use an active control in a

00121

randomized fashion? It seems like that would have been a more convincing design.

DR. GRECO: Absolutely. It's not feasible. We have a cooperative group that's been studying these patients, really since the early '80s. We pooled the physicians and that study, which is the ultimate study, was not feasible, it could not be done in the United States, they would not put the patients in the trial, so we had to go to the penultimate study, which you heard today.

DR. SARTOR: And so, if I may, the rationale for placing patients in a trial is that they were a priori convinced that site-directed therapy would be better?

DR. GRECO: I guess so. Again, you'd have to ask each individual doctor there.

DR. SARTOR: Sure.

DR. GRECO: Certainly we had data at the time, and Dr. Conley even showed some of the favorable subsets, those patients with colorectal profiles, that was defined not only by immunohistochemistry but also by molecular diagnosis. Those patients looked to as well as their cohorts who had known colorectal cancer,

00122

so you can see why some doctors might be a little bit leery about giving them a control therapy which is inactive against colorectal cancer. That's also true in renal cancer and many others.

So there is an inherent bias that is just part of reality, I can't get around it. There's never been a Phase III trial on unknown primary cancer, so the empiric treatment, which

by the way, basically was developed by my group, not entirely, 90 percent, was based on Phase II data as well, so, you know, the law of the literature is not always perfect. I wish it were.
DR. REDBERG: Dr. Rizzo, and then Dr. Sanders.
DR. RIZZO: I have questions of two people from these studies. To follow up on this question, you're saying that basically because of beliefs, we wouldn't really be able to execute a proper randomized clinical trial to, once we had the potential for directed therapy, directed versus not directed therapy, essentially?
DR. GRECO: Exactly right. People,

00123

you need cooperative groups to do that kind of a study because of the number of patients, et cetera, so you have to have some early indication of whether you're going to have a pool of subjects, and from the cooperative group we had no such assurances, so we had to fall back to the next level of evidence.
DR. RIZZO: Great. Can I follow up on the study you presented with two questions? First of all, you had a historical control group and you compared those with the assay-directed therapy to historical control. Two questions about that historical control. First, what are the years of treatment from which it's taken, and second, you did not present a P value that actually compared the survival, or the survival difference between your assay-directed therapy and your historical control, whereas in the next slide you did present the P value. Can you elaborate?
DR. GRECO: The second one first. There was a statistically significant difference in those two curves that wasn't listed, median survival was the endpoint there, and so that was statistically significant. The

00124

curves were as well, but it wasn't reported on that slide.
The historical control patients were taken from the same cooperative group that did this large Phase II trial. In the six years previously, okay, in 396 patients, we had great details on those patients, but because of size limitations we couldn't include them, a very similar group of patients.
DR. RIZZO: Last question with regard to your presentation, if I may. You presented

a more responsive and a less responsive group, but isn't that what we would generally expect? If you were to compare the survival of patients who had pancreatic cancer to the survival of patients who had colorectal cancer, to a certain degree you've shown us what we already know; is that not correct?

DR. GRECO: Yes.

DR. RIZZO: I'm just trying to understand the value of that comparison versus the real comparison of interest which is, if you have site-directed therapy, is that better than not.

DR. GRECO: Good question. These

00125

patients have unknown primary cancer, we don't know the site of origin, okay? So this test allows us to have the tissue of origin, and therefore the patient would then be treated as such. Without that information, the standard therapy is to give them all the same therapy and it's not effective in a lot of them. Plus there are tertiary, secondary and tertiary therapies that are useful in breast cancer, I can name many, renal cancer, lung cancer, that you wouldn't even know to use unless you have a diagnosis. So that substantiates our belief that it's important to know, those curves are precisely defined by molecular diagnosis. Those are unknown primary cancer patients, yet the curve separates with site-specific treatment. To me that's more important than the primary endpoint, but you have to judge for yourselves.

DR. RIZZO: Thank you.

DR. REDBERG: Dr. Sanders.

DR. SANDERS: I have two questions that are somewhat linked and probably for any of the presenters this morning. Dr. Schnabel characterized survival as the hardest kind of

00126

outcome that applies in these types of studies, and I'm just wondering if a three-month increase in survival in most cases is considered by either oncologists, other treating physicians and patients as a clinically meaningful difference. And then my second question is, if somebody could characterize for those of us who are not oncologists in the room, what the side effects are of these profiles, how they differ between the empiric therapy and the therapies that might be used in the aftermath of the application of one of these tests.

DR. REDBERG: Dr. Greco.
DR. GRECO: I wanted to address the survival because it's a very important question and most lay people when they see that, even a lot of doctors don't understand survival curves in cancer patients, they don't understand, a three-month median survival difference doesn't mean that every patient who gets the treatment lives three months longer and then dies straightaway the next day, those are statistics of the whole group. As I showed from the second curve, there are some patients in there

00127

living three and four years, so how important and how long should a median survival be?
DR. SANDERS: Clearly those are outliers, though.
DR. GRECO: Yes, there are, but some live less, some live more, but your question is a good one. We predefined what we thought was important. In a lot of studies of patients with advanced cancers, that predefined amount of length of improvement is a satisfactory and accepted amount, even though some people say that's nothing, I mean, why do that, that doesn't help anybody. You see what I mean? It does in fact help people, particularly those outliers on the other side of the median survival, which of course is half the patients. Most drugs in this country are approved for median survival differences in advanced cancers of from three to six months, that's it, so this fits that as well.
Now the second question, I think the answer is up for grabs.
DR. REDBERG: Dr. Schnabel.
DR. SCHNABEL: So to Dr. Sanders' point, the other thing that I would point out

00128

is that, and Dr. Conley also brought these points out during her presentation, is that you're talking about a subset of patients that have a vast amount of heterogeneity. It's difficult to demonstrate a clinical margin of three months in a specific tumor type, much less in a cohort of patients that had arguably 20 different cancer types. And so while there are different prognostic variables to consider within, which is a part of Dr. Rizzo's comments, within that patient population you have to realize that three months in that largely heterogeneous population is clinically significant and clinically meaningful.
DR. REDBERG: Dr. Sedrakyan.

DR. SEDRAKYAN: Dr. Whitehead and Dr. Conley, can you comment about the study that has been presented by Dr. Greco that is not part of your technology assessment, and comment about the quality and any other issues that you would like to talk about, sample and convenient sample, whether it's consecutive, please comment about the quality of it.

DR. WHITEHEAD: Some preliminary results from that study are included in the

00129

technology assessment, because we had abstracts that had been presented on that study. We had several concerns about the study from the point of view of addressing whether or not the molecular tests worked better than current standard of care, and that is that everybody's first, there was no randomization on whether people got the tests, so everybody was offered the tests, and were only in the trial if they accepted the test. And the people who were used as controls to compare survival outcomes in that study were people who did not want to have their therapy based on the test.

And the study that used empiric controls, there was no -- I heard today that they just couldn't include the comparison between the two patients, but there was no adjustment for any differences of any workup included in the article that looked at whether or not the patients who were used as empiric controls had similar treatments, similar characteristics, similar diagnoses, or anything that would confirm that those patients were in fact appropriate controls for the study.

DR. MELETH: Also, we could only

00130

evaluate the evidence based on what is published, so if there was information about patients that was not in the article, there's no way to assess the evidence on that.

DR. REDBERG: Dr. Conley, did you want to comment?

DR. CONLEY: Sure. This is Dr. Barbara Conley, asked to comment on the quality of the study presented by Dr. Greco. It is a fact that you cannot do an ideal study if the patients who are supposed to be on it, or the doctors who are supposed to put the patients on it won't do that. The heterogeneity issues are real. I look at that study as promising data.

I do, I think someone in this room commented on the curve showing the difference

between the good prognostic group and the poor prognostic group, but I don't think that -- I mean, that's an interesting graph, but it's too hypothesis-generating to be used for any kind of conclusions, and I'm sure Dr. Greco would agree with that, you know, how do we separate those patients out.

So, I think within the realm of what

00131

we have available and the fact that that same group of doctors have been putting the same kind of patients on the same kind of trials for a quarter of a century, it is probably as good as we would have in the literature at present.

DR. REDBERG: I'm sorry to hear that. I do think a randomized trial with actual controls and removal of bias is still a really important -- I don't know, Dr. Rizzo, if that's what you were getting at, but certainly in our history of cancer treatments, and I was thinking of bone marrow transplantations for advanced metastatic breast cancer which, you know, a lot of people, I mean, this is clearly a very sick and very desperate group of patients for treatments and we want to believe that treatments or tests will work, but there were a lot of women that were harmed until a randomized trial was done showing that there was no benefit and tremendous harm from this test. So it's hard not to think that, you know, we have to really do a randomized trial, and it's certainly a trial of high quality with actual controls and removal of bias to really answer the question and give the best care to

00132

our patients. You know, however good our intentions are, we still need to find some evidence. Yes.

DR. BEYER: A couple questions. First, for Dr. Schnabel, at the end of your presentation you had made the, kind of tossed off the comment that these tests are the subject of some consensus statements, and I would like to ask you to be a little more specific about whose consensus statements or guidelines or what these were parts of, and were these the specific tests in question that they were talking about.

DR. SCHNABEL: Thanks for the opportunity to clarify my comments. What I was referring to was that there have been several consensus statements really from the opinion leaders and groups that have investigated CUP, that have published guidelines that have

integrated these molecular classifiers in their clinical paradigms. So for instance, you know, investigators and clinicians at M.D. Anderson, we have European investigators as well that have collaborated with our investigators here, that have published diagnostic paradigms that

00133

have integrated molecular classification into their workups.

DR. BEYER: Are there any society statements on this other than what we've heard today?

DR. SCHNABEL: Not currently. These are national and key opinion consensus statements.

DR. REDBERG: Thank you. You had one more question, and then I have Dr. Nowak, Dr. Stecker, Dr. Sartor and Dr. Howard.

DR. BEYER: If I could ask Dr. Greco to come back and comment, I would agree with the other commenters that the second graph, the good group versus the bad group is thought provoking, but merely thought provoking. The first graph that you showed, with the median survival difference, I actually do recognize a median survival of three to four months as pretty good, but I notice that these curves were absolutely superimposable for at least the first six months, there is not a hair between the two. Can you comment on what you think is going on?

DR. GRECO: Yes. This diagnostic

00134

molecular test is not therapy, okay? It unfortunately isn't therapy, it diagnoses the type of cancer you have. You're going to read later how accurate it is, that's going to be published. When you have pancreatic cancer, biliary tract cancer, multiple other serious advanced cancers where we have no effective therapy for it, there's no site-directed empiric therapy that works. That's the front part of that curve, in my opinion. Where the curves separate, and I do believe they separate, is in patients with breast cancer. We had one testimonial here, but I could give you 50 testimonials. Breast cancer patients, the median survival with this molecular test during this study was 28 months. In ovarian cancer with this test in that study, it was over 35 months. So again, can I prove that empiric therapy, which is the therapy given to these patients, shotgun therapy in those very

patients, they wouldn't have done just the same, getting back to the randomized controlled design? No, I can't prove it, but if a diagnostic test will give me the diagnosis

00135

confidently of the type of cancer I have, why would I want to use shotgun therapy, regardless of this study?

So they're two different questions.

Is the test accurate in diagnosing the patient's cancer? Presuming that it is, and the biology is similar to a known primary cancer where you know what the primary is, then you want to give treatment that's effective for that variety of cancer. That's fundamental, I'm not sure that you have to prove that, but I go along with the fact that I would rather have a randomized controlled study. It's just never going to happen, so you're going to have to be left with less evidence. We live in medicine with less evidence. I'd have to go home every day and never treat one of my patients. Maybe I could treat four of them if I had to depend only on randomized controlled data. I want it, but we can't find it. It can't be done in this group of patients in the United States, in my opinion.

DR. REDBERG: Dr. Nowak.

DR. NOWAK: This is a general question. Is there consensus on what is

00136

considered, how you define tumor of origin? I mean, the different tests, I know the Pathworks test gives you a similarity score and that somehow translates into a probability. I don't know about the other tests, whether they give you probabilities or not.

Mr. Berins, when we heard his story, he said his tumor was sent to Pathworks and that had a 50 percent probability that it was breast cancer. That's a flip of the coin. Do we treat him for breast cancer? Maybe it is, maybe it isn't. There was a 30 percent, I think it was, you know, intestinal, but 30 percent is still a pretty high probability that it's intestinal. So is 50 percent sufficient to say this is the tumor of origin and we're going to treat it as this, or do you have to reach 70 percent probability, or 80 percent? I don't know, but that's my question, how do you define that probability, and I suspect it has to be answered specifically for each different type of test.

The second issue that I noticed,

Mr. Berins pointed out the tumor also went to the microRNA assay, and it came back as 90

00137

percent probability that it's breast. So there's one assay that tells us it's 50 percent, another one that says 90 percent. Doesn't that bother anybody? I mean, one of these is right or better, why do we have that difference, and which test should we use? Maybe CMS should only pay for the best test, and I don't know which one that is, but we're calling all of these genetic tests for tumor of origin of cancers, but they're really looking at different biomarkers. They're proprietary algorithms and they give you an index, and in that sense it's a black box. They may all work very well, but it's very difficult to evaluate them, it's very difficult to compare them, and I'll stop at that for now.

MR. BERINS: May I clarify my remarks? The Rosetta Genomics test was not a 90 percent probability of breast cancer, but they believed for that test that it was a 90 percent certainty I had breast cancer. As a result of that, coupled with the bioTheranostics test of six months earlier, my oncologist at M.D. Anderson said that we can just treat it as a breast cancer, and starting in the middle of

00138

January instead of the shotgun approach previously, she prescribed an oral breast cancer chemo, Xeloda, which I didn't particularly care for after a while. But the proof is in the pudding, so to speak, because two weeks ago today I was at M.D. Anderson and the PET scan, for the first time in a year, didn't light up.

Now, I realize the insidious nature of cancer and I realize that perhaps the next time I go in July I'll get a metastatic cancer all over my body, but at this point I feel happy and more optimistic. Frankly, once I got the Rosetta Genomics test, I felt a lot better, my family felt a lot better, and it gave me hope for the future, which so far has proved out, and I hope it continues like that. And I think that frankly every American, whether Medicare or not, should have the opportunity for these tests so that they can not only physically be treated appropriately, but I think emotionally react to it. Having the diagnosis of cancer was not the high point of my life, and the two tests, first at bioTheranostics and then the Rosetta Genomics test was a positive

00139
enforcement on my system of what could be done and hopefully I could survive this thing, at least for quite a few more years.
DR. NOWAK: Mr. Berins, I appreciated your comments, and I'm very happy you took the time to talk to us. I didn't mean to imply that these tests are not useful and I'm very happy that you got a good result, and I hope that continues. In truth, I think these tests are useful. I think our purpose here is to define how they can be best used and how to make them better, and my comments were intended to elicit some of those things.
MR. BERINS: Thank you.
DR. REDBERG: We certainly appreciate your time. I think the issue is really how much we can learn from one person's case, because it's very difficult without scientific data, and that's why we do randomized controlled trials, because we have to have a group of people who, you know, all have the same issue, and then test the intervention and see whether the group that got it gets better, because otherwise we don't know whether you would have, and I don't mean you personally, I
00140
mean the group. That's why we do randomized controlled trials, to see if you get better with the intervention or without the intervention, the only way to test that is in a scientific study.
And I can tell you that there are many many many instances in medicine where something we believed in turned out not to be true, because then we did a study and we learned that the medication we thought was beneficial, I will take it out of -- in cardiology we used lidocaine for MI when I was a resident and fellow, and for many many years. Finally they did studies and found we were actually hurting people and causing damage. And so while -- that's why we try to look at the evidence and we have the technology assessments, but we really are trying to evaluate the scientific evidence so we can offer the best treatments to the most people. Did you want to make a comment, and then I have Dr. Stecker.
DR. SCHNABEL: Just a couple comments to Dr. Nowak regarding the results reporting on each of the classifiers. I think the important takeaway is not basically that each of the
00141
tests have their own output, it's really rooted

in the algorithms and the reference databases associated with the genetic classifier. These are relative measures, and so I think the important takeaway is that the tests were consistent, and in the meta-analysis that Dr. Whitehead and her colleagues have summarized, the key takeaway is that the accuracy is quite consistent across the classifiers.

The other point I'll make is that for the Rosetta tests and the bioTheranostics tests, the biomarkers are published. The gene characteristics have been published, so they're not proprietary from that standpoint, and really what they are are master regulatory genes that are involved in many lineage determinations.

DR. STECKER: Hi, Eric Stecker again. This is kind of a clarifying question for Dr. Conley with regard to the control issue, which I think is very important to figure out here. The foundation of our knowledge in modern medicine is based upon a good control group. I agree that we need to often, I mean for the optimum effect in probably most

00142

patients, we need to either extrapolate beyond or interpolate within the available direct evidence, but it's still based upon well controlled studies, whether directly or somewhat extrapolated.

I was amazed to hear two people involved in oncology research say that historic controls, which is really not even a control group barely, it's called a control group but the lowest quality of control group, but historic controls are the best available for this kind of research for cooperative oncology research. And what that implies to me is that, you know, there is a standard of care, every clinical practice, or every discipline has a standard of care, in other words, what do the vast majority of patients receive for care. The vast majority of patients with cancer of unknown primary are not getting tumor-directed therapy, is that not correct across the country? Therefore, if what you two are saying is correct, the state of oncology cooperative research is such that you cannot test a new therapy against the standard of care, and if that's true, that's an issue.

00143

Dr. Conley, is my assessment correct, or would you clarify it?

DR. CONLEY: Well, it's a little broad

when you say the state of cooperative oncology. We're talking about cancer of unknown primaries here --

DR. STECKER: I mean cooperative oncology groups.

DR. CONLEY: And not all cooperative oncology groups. There was cooperative oncology groups that did this study on the breast cancer and bone marrow transplants, finally.

DR. STECKER: What's different about the cancer of unknown primaries?

DR. CONLEY: Yes, this is cancer of unknown primaries. Not all breast cancers, you know, there are many different types of cancer involved in that, and the reason it's cancer of unknown primaries is because we can't tell what it is. And so, my statement meant to me that this is promising data, and I would love to have more definitive data, but it will take a lot of thinking to figure out a clinical trial that would take account even of the variability

00144

in this class of patients. I think it could be done, but it's going to take some thinking, and may take some novel trial designs such as being, such as are being designed for some of the targeted therapy trials now. It could be done. It's a question, you know, if you are a researcher and you put a grant in on one of these things, the likelihood it will be funded by the NIH this year is fairly low, I think, unless it's brilliant, but you know, we hope for the brilliance.

DR. STECKER: So the issue you brought up about the control group is one very specific to the cancer of unknown primaries, and you feel that it's possible but very difficult.

DR. CONLEY: Yes.

DR. GRECO: I'm sorry if I was taken out of context. I believe in concurrent controls of randomized Phase III trials, of course I do, but Phase II prospective trials aren't useless, you just have to know the limitations, and the circumstances. This is not one disease, this represents probably 50 or 60 different cancers.

00145

It's heterogeneous, we don't know which cancers are which. Plus it's relatively rare, at least it's rare that doctors don't want to treat them as one cancer or another, that's just the way it is. So it's very very difficult to do these

studies. The cooperative groups in the United States have not done one trial since 1979 with these patients, not one trial, that's the public cooperative groups, so this is a very difficult area.

But again, I want to emphasize, I have to say this. Two questions. First, do we have a diagnostic test for what type of cancer it is? Second, if we treat the patients that way, will they do better. Two different questions, don't mix them up. Not all data that's from Phase II trials where we treat patients will harm once we have the Phase III trials. Sometimes it's the other way around, patients are benefitting, and then we prove that with a randomized trial.

DR. REDBERG: But we still need to do the trial.

DR. GRECO: We want to do the trial, but that doesn't mean the patients aren't

00146

benefitting. The way you described it, it always turns out that they're not benefitting. It can work the other way around.

DR. REDBERG: Absolutely, but you don't know, and you can't assume. Okay, Dr. Sartor.

DR. SARTOR: I think I might like to make a little bit of a statement, and I would like to see if this would be an accurate statement. One of the true problems with this disease subset is the incredible heterogeneity that by the very definition of cancer of unknown primary we do have, and I'm looking at one particular manuscript, 20 or 25 different cancers. Ideally you would have to design a trial that would incorporate an appropriate control group for every single subset that is identified against the control, and I think that's the practicality that makes it so difficult.

And I will say as someone who treats a number of patients with kidney cancer, that giving them a cytotoxic regimen if I knew they had kidney cancer, I would consider to be unethical because of the proof that targeted

00147

therapy is effective in that subset, whereas the subset of patients who receive a traditional toxic chemotherapy, we just know it doesn't work.

But the problem here, the real problem here is the huge heterogeneity, where I believe that there is a subset of people where this

test could probably make a fairly large difference. You cited some breast and then you cited colon. Yet unfortunately, when you actually look at the numbers of the breast and colon patients, it was 12 patients with the breast and 28 patients with colorectal, so we end up with a very small subset.
So I think the real problem here in the lack of randomized trials relates to the heterogeneity of the population under study, and the extreme difficulty in doing the proper study. Did I get that right? Okay.
DR. REDBERG: Dr. Howard.
DR. HOWARD: Thank you. On the subject of trials, I actually do not understand that point. It seems like we're not testing the treatment here, we're testing information. So you have one group of CUP patients where you

00148

don't get any information from the genetic test, you have another group of patients who you give their physicians that information and they can do with it what they wish, they can use it how they want. I guess I don't understand why, unless genetic testing is currently the standard of care, why would any physician be reluctant to enroll patients in that trial?
DR. GRECO: Again -- do you see patients yourself?
DR. HOWARD: No, sir.
DR. GRECO: And I understand what you're saying, but there are a number of reasons. Of course we're all humans, doctors are humans, the patients are thinking about themselves, and we, in Europe or some places in the world, we might be able to do the type of study that would be ideal in this setting, but not in the United States. Patients that are thought to have renal cancer, they'd have to go on the empiric treatment or the renal cancer therapy randomly allocated, do you understand, the doctor can't do what he wants or she wants. So if they have renal cancer and they happen to

00149

get the flip of the coin to get the toxic chemotherapy you have to give it to them, that's part of the study.
DR. HOWARD: I still don't understand, but if I could ask how, do most patients -- I guess I'm having trouble reconciling the idea that all patients, or almost all patients with CUP receive empiric therapy was the data that, I believe it was Dr. Wassman, that you put up.

You showed a slide with two bar graphs comparing genetic tests with, I believe, standard pathologic examination, and it looked like standard pathologic examination could provide information that would allow patients to receive cancer-directed, or tumor-directed therapy in a lot of cases. Am I reading your results incorrectly?

DR. WASSMAN: No, but understand that once standard immunohistochemistry, which worked in 69 percent of the cases, makes the diagnosis, these are not CUPS, these are tumors for whatever category they go into. And all I'm saying is that when you use molecular tests, 10 more percent of those difficult to diagnose cases are categorized into a specific

00150

category, and those patients are no longer considered to have CUP, or if they ever were, but, you know, go into that basket of colon cancer, ovarian cancer, and get treated, you know, on those other protocols.

DR. REDBERG: Pamela Massey, then Dr. Stecker, then Dr. Sedrakyan.

MS. MASSEY: I guess this is a generic question, but it's related to this do no harms. We know the problem in terms of the heterogeneity of the group, and so if we address the question the other way, do we have any studies that show that when we have a better test that tells us what the tumor is, that we're doing harm to patients by using that test, is there anything that has looked at it that way?

DR. GRECO: That's why it's redundant. This is not like bone marrow transplantation in breast cancer where you have a mortality rate of 10 or 15 percent and you think you're helping patients. All of these patients get cytotoxic chemotherapy if they're healthy enough because that's the standard therapy. So when you use a test, all you're doing is

00151

deciding if you can, this test, which specific treatment you would use based on their likely diagnosis rather than shotgun therapy. So they all get their chemotherapy, so the harm of doing this, there's no harm in it, you see what I mean, so it's different.

DR. REDBERG: Dr. Whitehead or Dr. Meleth, did you want to respond?

DR. MELETH: There are no studies that have specifically looked at harm, so most of the outcomes that have looked at survival.

DR. RIZZO: Technically speaking, assuming everything works as it should be, then if this assigns the right diagnosis, then you're aligning the harm by way of toxicity of treatment with the appropriate tumor type in terms of that, if you want to think about the harm. At least then you would know, and I think you guys would agree, if that's the case, then the toxicities experienced by the patient are at least more appropriate to the type of malignancy, and the harm that they would otherwise experience when treated along that pathway, and maybe you are avoiding harm in a few malignancies where there is more directed

00152

therapy that is less toxic than a general regimen that's customized, if you will, to a specific tumor type.

DR. WHITEHEAD: If I may clarify something, the charge of our technology assessment was to look at the test and to evaluate the test, so we did not specifically look for articles that compared site-specific treatment to empirical treatment.

MS. MASSEY: Our question is about improving treatment outcomes, and I didn't know if any of those studies addressed harm.

DR. WHITEHEAD: None of them mentioned even the possibility that there might be harm, there was no data on that at all.

DR. REDBERG: Dr. Sedrakyan, then Dr. Stecker, then Dr. Wong, then Dr. Marciniak, and then Dr. Beyer.

DR. SEDRAKYAN: I think this is really an important discussion to reflect on, particularly in the context of sensitivity and specificity of the tests that you presented, and it's an important part of the technology assessment, so it would be important for us to get more clarity around the harms. If a cancer

00153

is identified primary site, do you think in the literature, and this is a question also for clinicians, do you think there would be a more aggressive therapy path because now you know what the primary site is? And is there any documentation in the papers whether the chemotherapy regimen or radiotherapy regimen, or whatever the treatment regimen, is a lot more aggressive, which potentially can lead to more harms? Did you look at that question in the literature?

DR. MELETH: No, we didn't specifically look at that question. One of the

things that I would like to point out is that there is also literature that says that the group of patients that, in quotes, are benefitted by this test will be very small. That is another reason, because the site-specific therapy would be applicable to a small number of patients who were diagnosed with CUPS. But there were no studies that actually identified harms and had any good sort of quality of life data that we looked at. And the, another thing that might be important to point out is that as somebody

00154

said, this is a probability scale that is given, so there isn't a typical sensitivity specificity associated with these tests, so the total probability adds up to one.
DR. SEDRAKYAN: Okay.
DR. REDBERG: Dr. Stecker.
DR. WHITEHEAD: There was only one study that I can recall that actually even looked at what the chemotherapy regime was.
DR. REDBERG: Dr. Stecker.
DR. STECKER: Dr. Greco, sorry to give you a workout here getting up and down to the mic. I just wanted to follow up on what one of my colleagues asked about with regard to what are we actually testing, is it wondering what is the effectiveness of a diagnostic test to tailor therapy for cancer of unknown origin? And so why is it not practical to do, for instance, outside of financial and federal sequester reasons maybe, why is it not practical to do a randomized trial of testing a patient -- my colleague posed test everybody and then randomize to tailor therapy or not. I understand that neither patients or physicians are going to be comfortable having had a test

00155

done and not having that information, but why would it not be practical to randomize to either do this molecular testing or do standard of care and then see, for molecular testing -- you know, do the tailored therapy for the tumor of origin, and see what the outcome is compared to standard of care for cancer of unknown origin, why would that not be practical?
DR. GRECO: You have to have adequate informed consent for patients, and you could do two concurrent studies, you could do empiric therapy for unknown primary cancer patients like we've been doing for the last 20 years, and we could do the study like we did. But remember, even though they're not randomized

controls, they're concurrent controls, they still could be criticized. It's not a randomized controlled trial, it's a Phase II trial with concurrent controls done in another Phase II trial, it still has limitations.
DR. STECKER: But why not randomize, why can't you randomize?
DR. GRECO: You have to have an informed consent from the patients.
DR. STECKER: We do that all the time.

00156

Why in randomized trials --
DR. GRECO: The patients have to sign informed consent for the randomization process and the doctors have to explain it to them.
DR. STECKER: Oh, I'm sorry. You're saying that the patients wouldn't even want to stick around in that trial, they want to know they're getting some experimental therapy and --
DR. GRECO: The experimental therapy is based on the diagnosis for the actual cancer. The standard therapy is shotgun therapy.
DR. STECKER: Right.
DR. GRECO: So they have to agree to either get one or the other regardless of whether their diagnosis is based on the molecular test.
DR. STECKER: Is it correct for me to interpret that you're saying that patients with cancer of unknown origin are very unlikely to want to participate in any randomized study, they would either want standard therapy or, if a study is available, get on the study, no randomization.

00157

DR. GRECO: In this country, it appears that way.
DR. REDBERG: Dr. Wong.
DR. WONG: Sandra Wong, surgical oncologist at the University of Michigan. I want to take this conversation away from trials for just a second and focus on a point of clarification for the panel, and I think this is specific to comments that were made by Dr. Neal and Dr. Weiss.
What I'd be interested in and I think what might inform the panel here is very specifically, Dr. Weiss talked not necessarily about tumors of unknown origin but tumors of difficult to characterize origin. So I very specifically want to know the value added of the genetic test above and beyond standard

immunohistochemistry. In my practice what I see a lot are patients who have gone through standard anatomic pathology and gone through multiple layers of immunohistochemistry, and then had genetic tests, and sometimes the pathologist will comment to me that the genetic test confirmed the immunohistochemistry. That almost seems like the genetic

00158

test may not have been necessary, so I wonder if you could comment on some of the data that were presented here and how much of that were tumors that could have been characterized without the genetic test but the genetic test was done anyway. I think that's an important distinction, and I think that Dr. Weiss almost got to it by saying that these were tumors of hard to characterize origin, meaning that you kind of had a suspicion to begin with, so a pretest probability type of question.

DR. WEISS: So, all I can do is tell you the data we have. Using poorly differentiated tumors, where looking at it you didn't have a good idea where the primary was, using standard immunohistochemistry we got about 70 percent of the time, using molecular tests we got another 10 percent out of that. I know in my daily practice, I have cases where I'm able to do it on five, eight immunohistochemical stains and feel pretty comfortable with the diagnosis. I also know I have cases in my daily practice where I do 10 or 12 stains and I still don't have a good idea of what the diagnosis really is, and I think,

00159

forget about the clinical follow-up or whatever, I think those cases could benefit from molecular testing.

So I think there's, regardless of whether patients show differences on randomized trial, I think better diagnoses can be given about 10 percent of the time on poorly differentiated tumors that come every day to a pathology lab.

DR. WONG: Let me ask the difficult question then. If you get the tests back and then you're able to better characterize it with immunohistochemistry, in other words, would you be able to get to that anyway if the battery of IHC were done de novo.

DR. WEISS: Let's say I do a battery of IHC and it comes out, I think there's a suggestion of pancreatic, and then a molecular test is done and it says 80 percent pancreatic.

Does this help the clinician? I think it does. I'm not answering your question.
DR. WONG: No, you are, but I, maybe Dr. Neal can comment, because I think she presented a slightly different point. And I would be curious just to inform the panel,

00160

because I see this as slightly conflicting in terms of the order that the test was done in and how that influences clinical decision-making.
DR. NEAL: Dr. Neal, and I do share your concerns and think this is an excellent question. I agree that there are many cases that immunohistochemistry does point to a very clear tumor of origin. However, there are a subgroup that we really don't have the techniques at this time, and it's evolving, we're getting more powerful and specific with immunohistochemical tools and so it is evolving. But what is of concern, especially to me, is those cases where the molecular test is ordered before the immunohistochemical tests are even finalized and so the cost issue is certainly significant, as well as, I'm not sure, at least a good portion of these tests have no clinical difference, so that the immunohistochemistry would have told the answer.
Again, there is certainly a subset where at this time immunohistochemistry does not address it, and I believe the molecular

00161

test does have merit. Who's going to be able to develop those algorithms so that we know when immunohistochemistry is at its softest, what they can at this time, and then the potential for the molecular test, I suggest the pathologists might and should be involved in these decisions.
DR. REDBERG: Dr. Marciniak, we'll hear your question, then Dr. Beyer, and then we're going to break for lunch.
DR. MARCINIAK: The question I have is, you know, are we shooting for where the puck is today, are we shooting for where we think the puck is going tomorrow. My first question is for the groups that put together the TARs, so Drs. Whitehead and Uhlig, and then I want to hear from Rosetta and bioTheranostics as well about this.
So we had some good positive information, we talked about the clinical trial this morning, Dr. Wong has helped clarify this

as well. Where do you see the evidence moving in terms of the outcomes? You've got a test, you have a physician, you have a patient, you give a test, does it lead to a better outcome?

00162

Do you see the literature evolving there?

DR. WHITEHEAD: At the moment I think that there is so little literature on whether or not the test makes a difference in the final outcome that I would not want to say which way I think it would go. That's why most of the studies were rated insufficient evidence. And I will say I'm not a clinician, I'm certainly not an oncologist. I think to me, it looks like, you know, they're pretty good at telling you what the diagnosis is, at least based on the evidence we have now. Whether or not they do, you know, enough of a better job of that to make a difference, I don't -- there's very little reporting. There's a couple things in the literature on the difference in cost, which was not part of our technology assessment, or the timing of the diagnosis, which would I think impact the clinical utility question as well, and there's just not enough data in the literature for me to draw a conclusion on.

DR. REDBERG: So at this time insufficient evidence on outcomes and no clear evidence of where we're going. Dr. Uhlig, did you have a comment?

00163

DR. UHLIG: Well, for FISH I think it's an emerging technology and I can see that it would have a role, as has been shown in the beautiful pyramid, in a lot of people with abnormal cytology tests. But this is, as you said, it's an evolving field, and the other diagnostic tests are evolving as well. You know, the HPV test is evolving, and that will level the playing field again, so you will basically have to go back in and reestablish the value added from something like FISH test. So, I think it's very challenging to assess evidence in diagnostic tests, and this is in an area of rapid technological evolution.

DR. MARCINIAK: Looking at the oncologists on the panel in an effort to sort of help clarify, do the tests help you practice medicine better, do they give you insight in your patient population that you wouldn't have otherwise to allow you to address this therapy?

DR. REDBERG: Do you want to answer that question, Dr. Sartor?

DR. SARTOR: So, you know, I'm going

to speak as an individual because that's all I can ask, and I would say that it's helpful for

00164

me to know the site of origin in some cases but not others, in some cases it can make a dramatic difference. And I've cited the renal cell, for instance, whose therapeutic armamentarium with FDA-approved drugs is very very sustained, compared to the cytotoxic therapy, which is known not to work. Unfortunately, I think for many patients, those potentially who may have pancreatic or biliary cancer, we just don't have very good treatments, and so whether or not I'd use the empiric regimen or a pancreatic-directed regimen, it probably doesn't make a lot of difference. So in summary, I think for some patients that it could make a big difference, but for many it makes no difference. That's just a personal opinion.

DR. REDBERG: Dr. Beyer.

DR. BEYER: I want to somewhat echo that. It's always nice to know and I want to know, but whether it makes a difference, I think is actually one of the questions we're here to answer today. And oftentimes, you know, it's hard to be convinced that it makes a

00165

huge difference, which brings me to the question that I wanted to ask, and I'll kind of throw this out, and I'm not sure if you want to take a stab at it, but it kind of relates to where the puck is going. Are we looking at the wrong thing? Is it interesting to identify where the cancer came from or is it more interesting to identify if that patient has specific molecular genomic targets for which we have specific therapies? Here we're identifying they have, you know, biliary cancer, we don't have a targeted therapy for that. Would it make more sense to be running a battery of tests that will tell us, does the patient have A, B, C or D, where we have specific targeted agents?

DR. WEISS: Right now the answer is both, and until we get better with the molecular pathway business, the organ of origin is going to be more important in the short run. In the long run, you know, you can easily answer that question, that pathways are ultimately going to be more important, but right now I think the answer is both.

DR. REDBERG: I told Art he could have

00166
the last word before lunch, but we will be back here at one o'clock and you can make a comment then.
DR. SEDRAKYAN: I really wanted to kind of reflect on this question of potential harm again, I think some of you wanted to talk about this issue, whether there is a potential for more harm if this tissue of origin is identified and there's more aggressive therapy. And a continuation of that question, if you have this tool, an oncologist has this tool, what's the potential for it to be used inappropriately, which is now suddenly, you become uncertain even when you identify the tissue of origin, you start using this as another confirmation and confirmatory tool, so you kind of spread this technology inappropriately, and what's the potential for that? We've got a lot of uncertainties that you're dealing with in oncology.
DR. GRECO: I think there's a potential for harm if you use empiric therapy in patients where it doesn't work. For instance in pancreatic cancer, Tetracel carboplatin doesn't work, so you don't use it,
00167
so if you have the diagnosis you avoid harm. I could give many other examples, so you actually are avoiding harm by knowing the diagnosis. The other issue I just briefly want to mention is that oncologists use immunohistochemical findings from our pathology colleagues to treat patients. There's never been a randomized trial showing that that works in unknown primary cancer, yet we all use it, so there's something else going on here. I don't know what it is, but immunohistochemistry has never been subject to a randomized trial for unknown primary cancer, yet we use it every day.
DR. SEDRAKYAN: Other comments about this?
DR. WASSMAN: I'm Dr. Wassman, from Rosetta Genomics. The diagnostic incremental information as demonstrated here today improves our ability to diagnose these cancers, to sort out the heterogeneity. The sorting of heterogeneity was foundational to our starting on the genome project, we had a group of disorders that all looked alike, and our ability to sort out heterogeneity of disease is
00168
what leads to diagnosis, which leads to correct

target therapy. That's a specific targeted therapy that is based on a mutation, but a targeted therapy is sometimes based on the tissue diagnosis.

I mean, breast cancer alone is not necessarily in the targeted therapy, but 10 percent of these CUP patients when studied molecularly by either test roughly come out with breast cancer. About one percent of them are male patients like Mr. Berins with breast cancer where it's not being suspected. The difference in therapeutic response of those patients is dramatic.

There's not been a controlled study since this is a small subset of a small population, again, but there is a traumatic response, and oncologists know that if you treat breast cancer as breast cancer, they respond to that therapy.

DR. REDBERG: Thank you, and I would comment that the real value of a test, obviously, is in adding incremental information that we would not have gotten without that test or without clinical assessment, that leads to a

00169

change in management not on its own, but a change in management that leads to better patient outcomes, and that's really the bar that we need to meet in order to get that therapy to our patients.

We are now going to break for lunch and we are returning at one p.m. and will welcome you back, and continue with discussion and questions.

(Luncheon recess.)

DR. REDBERG: I want to welcome everyone back from lunch, and it looks like a beautiful day outside. So, we will start again our panel discussion, and right now, this is questions from the panelists to the presenters, and I would like to, it's on the program that we will have open panel discussions and kind of discuss the voting questions, but it would be, I think helpful, to keep in mind the voting questions and kind of think about any questions you have starting now that you need help or clarification, or want to state opinions about your own views on the voting questions.

And remember, we have three voting questions and three discussion questions, and

00170

then after we finish this discussion, we will formally vote on the voting questions.

And so I think, Dr. Gutman, did you

want to start out?
DR. GUTMAN: Yeah. I'm very struck by the fact that there's actually no analytical or clinical proof here, and I thought Dr. Wong shot the arrow right in the bulls eye, because what we're being asked here clinically is whether, how well you detect when you're dealing with tumors of known origin that might have been hard to define, rather than, the only person who actually knows the relationship of tumors of unknown origin is actually God. So the question I have to ask is whether anybody actually paid any attention to whether the test speeded up the route to the tumor of known origin, if it somehow facilitated getting there faster, better, cheaper or wiser, or was standardized in some way. Otherwise, it seems to me that the clinical validity, I realize in the tech assessment they talk about a moderate constant to extrapolate, I'm not sure I'd be that generous, I'd say you don't know how it came

00171

with 80 or 85 percent concordance with known tumors, so if you asked about an unknown tumor, would it still be 80 or 85 percent? So it seems to me if you have to back-load the whole study, it makes the clinical utility piece even more important than the panelist suggested. And it makes, it strikes me that just because something is hard to do, it doesn't mean it shouldn't be done, that in fact you have to either believe in evidence-based medicine or not, and if you're extrapolating across a chain, you'd have to have some real output information or you have to be able to create a bridge, and I don't know how you create a bridge if you don't know the clinical validity of the test.
DR. REDBERG: That is an important point. What I took is that there's a lot of uncertainty about the diagnosis and the treatment and the prognosis.
DR. GUTMAN: I'm just casting my vote in favor of those who want a little bit more on the clinical validity, and the question actually is for the tech assessment group, were they able to look at faster or standardized, or

00172

better arrival at the tumors of unknown origin? It seems to me that the best you're able to do here is make an association of the accuracy in relationship to tumors of known origin, tissues of origin that are known, and

I'm just asking, was there any attention paid to whether this test might actually aid you in reaching that decision better, faster, more standardized, in a cheaper and more intelligent way, because that would be a value if the test would do that.

DR. MELETH: This is Sreelatha Meleth. No. What we did assess was whether the tests accurately predicted tumors of known origin, the time to predict tumors of known origin versus immunohistochemistry or other methods was not a focus, was not one of the questions we looked at, and from my memory of the papers, there isn't a lot of information in the papers that we looked at that addressed that.

DR. GUTMAN: I'm just addressing that there would be real value to the test if it somehow expedited that.

DR. MELETH: Yeah.

DR. WHITEHEAD: If I could comment,

00173

there were a few papers that looked at, you know, in cases that had been diagnosed as CUPS, you know, and a genetic, or one of these molecular tests were done, and they later diagnosed, found the primary site or they had those cases and they retrospectively went back and tested them molecularly that provided, then looked at them and saw how accurate were those predictions, and that was shown on the slide about diagnosis.

It ran from 48 to 88 percent. My memory is that sort of, they were clustering around 60 to 65, but there aren't a lot, and so the strength of evidence there was rated as low based on those studies.

DR. REDBERG: Dr. Sanders, I think you were next, and then Dr. Nowak.

DR. SANDERS: Sometimes with diagnostic tests that are sort of new and not entirely flushed out in an evidentiary sense, there are appropriate use criteria, and it strikes me that this might be an area where they might be helpful, given that it sounds like sometimes the genetic tests are used before CUP, or before ISH, sometimes after,

00174

maybe sometimes concurrently. I mean, is there any sort of algorithm or decision support tool that you've gotten wind of that maybe somebody is working on somewhere?

DR. WHITEHEAD: I saw nothing like that in the articles we reviewed. In some cases, the studies, you know, that tested --

well, in many cases they were testing bank tissue specimens from databases, and so it would have been an irrelevant point in those studies. There were very few sort of in-process clinical studies that were in the literature at the time we were looking at them. I think there have been a few published since.

DR. SANDERS: Was this something that could be simulated or modeled?

DR. WHITEHEAD: Not based on the data that was available at the time we reviewed it.

DR. REDBERG: I do think that's an important point, not just for this test but for new tests in general, and I guess genetic tests are often in that category, is where do they fit in in terms of what we already know, because obviously we have a lot of tools through clinical and other diagnostic tests

00175

that are currently available, so where does the genetic test fit in, particularly if you're talking about tests that might tell you probabilities but not actual certainties, and, you know, what is their role, and then how is that information used, does it lead to a change in management, and does that management lead to better outcomes and less harms, so overall net benefits.

DR. SANDERS: I'm not so sure that one size fits all here. There may be some unknown primaries where the result will be much more helpful than others.

DR. REDBERG: But the question is will we ever be able to identify those, and it seems like we need to be doing studies in order to do that.

Right now this is just panel discussion, thank you. There's a lot of people that are listening. Excuse me, but you do not have the floor, thank you. Dr. Nowak.

DR. NOWAK: I think the previous comments have been expressed, but I was thinking, my understanding is that the usage of these tests was not in lieu of

00176

immunohistochemistry, but would come into play after immunohistochemistry played out and it would augment that, and there's an issue about how much immunohistochemistry you do before you throw up your hands and say I just don't know what this is. So if the comparison is, is this faster, better, cheaper than immunohistochemistry, I think that's a different issue, but I didn't understand that

the usage of these tests would be that, I thought it would come into play after you had done the more routine kinds of things. And it doesn't mean that they can't supplant immunohistochemistry at some point in the future, but I didn't think that is where they are today.

DR. REDBERG: I think that seemed to be one of the questions, we heard various scenarios where sometimes they were used instead of, sometimes they were used in addition to, and it wasn't clear to me what you do when the two tests give you different answers.

DR. NOWAK: Well, you know, most tumors, most cancers, 95 to 98 percent of them,

00177

the tumor of origin is obvious. We're only talking about those that end up being called cancer of unknown primaries, and those are the ones that pathologists struggle with. But even at that point, I would think that these tests aren't entering into a total vacuum. We do know something, someone has looked at the tissue, there may be some immunostains that are available that are informative, and there's a pretest probability of what this might be. You're not going to surmise that it might be a prostate cancer in a woman. I mean, you know, there are all kinds of things that you already know clinically, you do know something pathologically, you do know something from the immunohistochemistry, so there's a pretest probability, and what you're trying to do is strengthen that probability by doing yet another test.

And from the figures that were thrown around, you know, immunohistochemistry on, I don't know if this is just on cancers of unknown primaries or in all of them, but Dr. Weiss said that in 60 percent of the tumors immunohistochemistry is sufficient to get the

00178

answers, and of the remaining 40 percent which is subjected to this kind of additional testing, you might get another 10 percent on top of that 60. So 25 percent of the time on cancers of unknown primary you're unable to get an answer, if we interpret those numbers loosely.

DR. BLEGEN: I may be asking the same question just using different language, but as I looked at the presentations I would say what is the value added of this genetic test, or

either, and my sense was from the lack of discussion, as well as presentation, that there wasn't much added done for the FISH testing, but there may be, and the question is how much and is it worth it. If it's 10 or 20 percent, that may be; if it isn't, then it probably isn't, doesn't have any value added.

DR. REDBERG: So just to remind you, there's actually two sets of things we're looking at with regard to both tests, and you're right, we haven't had as much discussion about the FISH test and you may have questions about that. But we're looking at both the clinical validity, so how reliable are the test

00179

results for diagnosing the condition, and then we're also looking at whether the evidence from the tests, the genetic testing affects health outcomes. So we have two separate questions to look at for each of these two separate tests, the clinical validity of the tests and then how that test result affects outcomes. Dr. Nowak.

DR. NOWAK: Someone at lunch asked me if pathologists are going to become obsolete, and I suppose in the extreme you could think that. First we take the tumor out, we send it out for the test, it's just, you know, from here to there, and why even look at the tissue. But my answer was that if anything, this affirms what pathologists have been doing, and we look at tissue and we examine it grossly, microscopically, and we assess whether it reflects the tissue of origin. Does it still look like breast, is it well differentiated, moderately differentiated, poorly differentiated, and we try to do that histologically. And those pathologists who are old enough will remember that we used to do something called a Kreyberg stain, to tell whether it was squamous differentiation or

00180

whether it was mucin production, to distinguish squamous cell from adenocarcinomas of the lung mostly.

We've gone beyond that, so in those situations where we can no longer tell histologically or histochemically, we've gone to immunohistochemistry where we're looking at antigens that are expressed on the surface, and as people have pointed out, those antibodies may not have gone through prospective trials and searches for evidence that's available, but I think we've accepted that, we accepted something as an adenocarcinoma if it's PVF-1

positive and P-63 negative on a small biopsy for lung, and we treat. And I think empirically, my guess is that the data will support that that's valid, and you can go back and do those studies, but my guess is everybody would probably agree that that thinking is valid.

So now after having done all of those things -- and so immunohistochemistry looks at expression. Looking at RNA is one step further upstream and as I think Dr. Wassman pointed out, we're just looking at upstream markers of

00181

these things, we're looking at the message RNAs that code for these proteins, or in microRNA assays, we're looking at other regulatory RNAs that determine differentiation and expression of phenotype. So it makes sense biologically that these things should have relevance in determining lineage and differentiation, and while -- so the hypothesis is that that's true and so far the data, while it may be limited in some ways, supports that hypothesis, and it goes along with our history of thinking and how we have treated cancers. So it would not surprise me if at the end all this plays out, and looking at these kinds of markers will be of value. I've lost my train of thought. So I think even outside of specific studies, that historical context tells me something about these molecular markers and this approach to evaluating these tumors, and it is one part of the continuum in how we go about evaluating tumors, it shouldn't be seen in isolation.

DR. REDBERG: Dr. Rosenthal, did you want to comment on that?

DR. ROSENTHAL: Yes, I do. Taking it

00182

as far as you've taken it, I completely agree with you and I'm following your chain of thought. But if you go back, I remember the days when cytology was distinguishing between adeno and squamous carcinoma of the lung and we thought doing a very good job of it, and then in the '80s somebody came along, I can't even tell you who, probably the WHO, said it doesn't matter what you're going to call a non-small cell, and those of us who love our cells said we're making a big mistake. And now what goes around comes around and I can, you know, laugh and say ha-ha, I can do this by looking at the cells, I don't need all this fancy stuff. But the reason we need the fancy stuff is that we

now have target drugs that are based on the genetic mutations in these tumor cells, so it's vitally important that we identify what's going on with this particular tumor, because now we have a real piece of ammunition for this tumor, and the therapy is driving us, as perhaps it should.
That's not to say that we shouldn't continue to do genetic analysis on every tumor that we possibly can, because eventually there

00183

are going to be target drugs for all of them hopefully, if we live long enough. And so just, I don't think we should be doing it just to give somebody the answer of where is that tumor from, and the clinical outcome is going to depend on what drugs we have to address the tumor the best way we possibly can.
DR. REDBERG: Dr. Sartor, and then Dr. Beyer.
DR. SARTOR: So, I think it is important to note that the grouping here is not by molecular mechanism that we would utilize in most targeted therapies, but rather to serve as a source of origin, tissue of origin, and I think that there is a significant link between those. Just for instance, I will mention the b-raf mutation, which is an FDA-approved melanoma for a very specific b-raf mutation. Now it turns out that there are other tumors from other tissues that can express that mutation, and in the New England Journal there was a demonstration that the drug affected those tumors possibly as well. So I think we do have to make the distinction between what these tests show, which is tissue of origin, as

00184

opposed to actionable mutations for targeted drugs, those are two separate issues.
DR. ROSENTHAL: But they're intertwined.
DR. SARTOR: They're intertwined, but in a Venn diagram there's a lot of distinction.
DR. ROSENTHAL: Absolutely.
DR. REDBERG: Dr. Rizzo, did you want to comment on this?
DR. RIZZO: I would sort of mirror those comments. I think right now identifying the tumor of origin gets us partway down the path of refining therapy, and in some cases knowing the tumor of origin, then actually you want to define molecular markers that will help you refine therapy more, and that's what we have now.

What we have coming is the ability to molecularly identify potential targets as we get more sophisticated across any tumor, like b-raf, or the use of Gleevec and the other targeted therapies, so we'll get there. But right now our best pathway to that road is tumor of origin to refine therapy with what we have right now. It's imperfect, however.

00185

DR. REDBERG: Dr. Beyer.

DR. BEYER: I have been sitting here listening to the conversation and thinking that what we're really doing is using organ of origin as a surrogate for what we really want to know, which is what does this tumor, what is the weakness of this particular tumor, what's the target. And insofar as we can identify a valuable target that we can aim at, I think that it's extraordinarily useful. Otherwise, it becomes a little harder to know whether what we're doing is making a difference in the long run or just making us feel better in the short run.

DR. NOWAK: But identifying, simply identifying signaling pathways that are apparent, I think will be insufficient, and I know people would argue for that, why don't we just go and find out what drugs are going to work. It depends on, the same signaling pathway if it's turned on or off has different consequences depending on the tissue background in which it resides. So you can identify aberrations consistent in pathways, and in some cases they activate something and in other

00186

cases they'll actually turn things off, so understand the differentiation of the tissue, and it may not necessarily be the tissue of origin, but it's the single differentiation, so it is the context in which those signaling pathways work.

So my inclination is to think that both types of evaluation will be very important, you will need to know which signaling pathways are there, and you also need to know the differentiation context, so I think both things will be important. And that makes these kinds of answers all the more critical as we start looking at those and the varying pathways.

DR. REDBERG: It sounds like the hope is that better characterizing of tumors in a lot of different ways will help us to target treatment and lead to better outcomes. But it

also sounds to me at this point we don't have any clinical trials that are actually addressing those outcomes questions and that's, you know, we have lots of ways to get more information, but I'm a clinician, and what is really important to me in the care of patients

00187

is does that information actually help me to take better care of a patient so I can offer them that benefit, and I think that seems to me to be where certainly there is an evidence gap, or one of the areas that need to be addressed after we establish clinical utility. And we haven't talked that much about the reproducibility and variability of the actual testing itself, which might be worthy of some discussion. Dr. Sartor, did you have a comment? You still have your card up.

DR. STECKER: That's mine.

DR. REDBERG: Oh, I'm sorry, Eric.

DR. STECKER: I've heard a number of people comment about the compelling rationale for targeted therapy and I share excitement about it, I think it's amazing. In fact I'm at Dr. Druker's institution, so it would be blasphemy for me not to be excited, and I am legitimately excited. But there is no biologic or clinical rationale that is overwhelming, and we've proven that so many times before. You know, I've lived through the antioxidant hypothesis of cardiovascular disease, and the basic biological scientific rationale for

00188

introducing antioxidants to prevent heart disease is overwhelming, but it ends up that they're actually clinically overwhelmingly negative. In my own field, using encainide, flecainide and moracizine to suppress recent PVCs after a heart attack, tremendous mechanism for suppression, but it actually harms people, it's very dangerous to do, and that was demonstrated in a randomized trial. Somebody had said that these are always negative. They aren't by any means always negative. There's a randomized controlled trial which looked at using defibrillators in patients after MI, and people thought it was unethical to do the study because there was such overwhelming evidence for benefit, but people soldiered on, NIH funded a study, and there was actual benefit. But if we hadn't finished that study, we never would have known for certain, and there would have been a lot of debate surrounding it. So

with that said -- well, I also would like to make one other comment. So basically well controlled trials are critical no matter how compelling the rationale for Gleevec is, or

00189

anything else, because the next tumor might be different.

The second point I'd like to make is a little bit contradictory to that because we've been talking about survival differences a lot, but there are other things out there. Our questions are surrounding health outcomes, and Mr. Berins highlighted this point, and some of the other people highlighted this concept. There are a lot of other -- you know, if you get out of the blue, you get a diagnosis of cancer with a terrible prognosis, you're told your doctors have no idea where it is, what to do with it, I mean, that's, you know, that's got to be earth-shattering, even more earth-shattering than just the cancer diagnosis. I can imagine that having some, whether or not it makes a difference in survival, having some idea, some concept that the people taking care of you are doing so in a directed manner would actually improve quality of life. And so studies that incorporate not just survival but also quality of life, I think would be very important, but I would note that there are none of those studies so far in

00190

cancer of unknown primary.

I also note that, I don't want a regimen where you have to do randomized controlled trials or perish, I agree that there are things you can't test, I don't think that this reaches that. I'd also point out that as far as observational controlled trials go, historical controls are deeply flawed. There are a lot of other ways to do observational controlled trials without randomizing people that are much higher quality, and that therefore I would put much higher stock in.

DR. REDBERG: Dr. Sedrakyan.

DR. SEDRAKYAN: I needed some clarification. There was extensive discussion in both technology assessments about clinical validity and analytic validity. I would like you to reflect on how analytic validity helps us here when you were dealing with clinical validity in here, and when you were referring to analytical validity it was really reproducibility and reliability of the test, so can you comment if that's a helpful concept for

us to consider in thinking about clinical validity, why a substantial portion of the

00191

technology assessment addressed that? Was it a precondition before you conclude clinical validity? Some of the clarification would be helpful about that.

And then, I'm reading the technology assessment for CUP and it says that the evidence that the TOO test contributed to diagnosis of CUP was moderate, and then it goes on to say low evidence supported the clinical usefulness of the TOO test in making diagnosis of tumor. This seems to be a little contradictory, moderate level, low evidence, so can you clarify this for me?

DR. WHITEHEAD: So, the first question was the value of analytics, looking at analytic validity, and I state again that we were asked in our contract to look at it. But if you have good analytic validity, or if you have good clinical validity, the analytic validity information may be, you know, only supportive. But if you don't have clinical validity and you don't know until you look at it, then you have the question of why not, and in that context the analytic validity information can become kind of important, you know, is it not

00192

measuring what they say they measured, can you not reproduce it.

So, I think that's probably the reason it was asked for, and I know that's the reason because I'm actually more familiar with doing strictly genetic test evaluations in the AIDS model. So I know that's one of the key parts of that model, is you need good analytic validity to have good clinical validity, and if you don't have good clinical validity, you want to figure out why.

To address the other question, I think this may be a matter of shortening things a little too much in the table, but there's moderate evidence that the TOO test actually provided prediction in most cases. That's what that is. There's low evidence that it adds to the existing diagnosis, that it's clinically useful in reaching a diagnosis over and above what the standard IHC might have been. I believe that's the answer.

DR. SEDRAKYAN: Thank you.

DR. REDBERG: This is Dr. Uhlig.

DR. UHLIG: I just wanted to add to the perspective in terms of clarification, so I

00193
think that there is a distinction between clinical validity and analytic validity, and, you know, for all the technology assessments, did we actually look at the reproducibility issues, because there are pre-analytic things that are dealt with in quality control and so on. I think that, you know, as was discussed before, is that the evidence builds up. You know, if you don't really have analytic validity, you know, that really makes it more difficult to assess the subsequent findings to base it on, it puts into question some of the possibilities of that, you know, the ability to show impact.
But in the case of FISH, you know, what I've heard from people who are doing this test is that there's a lot of subjectivity in actually scoring the tests, so we haven't even gone there yet, at least for FISH for cervical cancer, there may be things that are pre-analytic validity that we haven't dealt with.
DR. SEDRAKYAN: I'm a little more confused now because on a slide that says analytic validity, Dr. Whitehead, you have
00194
coefficient of reproducibility, and you have reproducibility, CT values, and then you have interlaboratory concordance.
DR. WHITEHEAD: So, technically speaking, there are more things reported in that table than would be considered a strict laboratory definition of validity, but they all get at how well the test is being done in the lab, and I used the broader classification. If you look at the actual mock ACCE framework, there's a long list of questions that feed into that that you consider in that, and how well does it actually measure what the analyte is that is part of the test, and that's why it says validity, but it could be reproducibility, it includes reproducibility of the scores, which of course includes both the laboratory results as well as the statistical algorithm. You need a way in the framework to look at how well the process of getting an answer works, and that's what isn't there.
DR. REDBERG: Dr. Uhlig, please go to the microphone.
DR. UHLIG: And I think you had the question of how important is it to have
00195
evidence on analytic validity, and in the case

of FISH we didn't find any study that correlated TERC with sequence analysis, DNA sequencing, and that didn't worry me, you know, inasmuch as it doesn't invalidate the subsequent findings. So I think it depends on your tests, how important each one of these tests are, but you think of them as building blocks.

DR. REDBERG: I heard one of the presenters, all of these are not FDA-approved tests?

(Discussion off microphone.)

DR. BEYER: But to clarify, FDA approval is not required for these particular tests because of the way they're used; is that correct?

DR. SARTOR: I'm sorry, I heard something about FDA-approved, so could you repeat what you just said?

DR. WHITEHEAD: The Pathworks tissue of origin test was cleared by the FDA. I have to have the definition in front of me to remember the difference between cleared and approved, because I don't remember, but I

00196

believe, you know, it's not required, it's presented to FDA and cleared is a less, requires a less rigorous evaluation.

DR. SARTOR: Is it a 510(k) problem?

DR. WHITEHEAD: It's different, and at the time we gathered the information on the test, that was the only one of the three tests that had been cleared by the FDA.

DR. REDBERG: Dr. Gutman.

DR. GUTMAN: Yeah, I can clarify. There are two routes to market currently for tests that are distributed to multiple labs, they require regulation by FDA. If the test is set up and used at an individual lab, that's called a lab-developed test. FDA has asserted that it has jurisdiction over this but has not yet exercised jurisdiction, so those tests are brought to market under the oversight of the other CLEO program, which is not a discrete approval or clearance program, and this was, Pathworks was cleared and defined as, I believe it was de novo, which meant that it was novel and it had special controls put into place, but it went through the 510(k) rather than the CMA process.

00197

DR. REDBERG: This is the substantial equivalence test.

DR. GUTMAN: Well, it's substantial

equivalence, but it can be applied to a moderate risk test and it can create its own boundaries as a moderate complexity test, it's sort of an automatic down classification, and then it serves as a predicate for future tests of that same type.

DR. REDBERG: Dr. Sanders.

DR. SANDERS: So, I've heard a lot today about the question of value added, which begs the question of how does one define value, and I've been taught by many recent discussions in health care reform that I should be defining value as quality divided by cost, so there, I've actually said it. I wonder, how much do these tests cost, and is there parity among them, and is there any kind of standardization across the country? So if I'm going to have this test in San Diego as opposed to Cleveland, does it make a difference?

DR. REDBERG: Part of the issue is that cost is not a criteria for a Medicare evaluation so we can't on its own discuss

00198

costs. We can, I suppose, introduce it as part of the value equation, but Medicare does not consider costs in coverage policy. I think we did hear some testimony about cost of the test, at least I thought we did.

MS. MASSEY: I want to follow up on the quality of life, and not the cost of the test, but does the test being given, and then what happens afterwards affecting quality of life, is there a difference in the cost of that quality of life? And you could look at it a couple of ways. If the test defines a treatment strategy that is more costly or less costly, or prolongs life, or -- I mean, there's a lot of ways to look at the costs from that point too.

DR. REDBERG: Right, and as I think we were also discussing, there's a lot of ways to look at quality of life, certainly the functional status, certainty, a lot of things that would come into play here that are not strictly about your diagnosis, but certainly about your treatment and prognosis, because I think that's what the hope is, it's related to treatment and prognosis, and then of course

00199

there are the tradeoffs of treatment versus nontreatment in terms of disability and additional life years.

MS. MASSEY: Right. I mean, if you're being treated for CUP with a treatment strategy

that has very toxic side effects, then your functional status and all of that is going to be impaired, whereas if you're treated with a known diagnosis there's fewer side effects, or they might be more, I don't know.

DR. REDBERG: Right, or you might be living longer but that might be in the hospital in an intensive care unit, so all those are questions certainly important to address in clinical settings. I think I saw, Dr. Marciniak, did you want to add something?

DR. MARCINIAK: No, I was going to chip in much along the lines that you did, Dr. Redberg, so you covered it.

DR. REDBERG: Art.

DR. SEDRAKYAN: Art Sedrakyan from Cornell. I have a question for the panel, particularly those who practice oncology or oncology surgery. If TOO is identified, how likely that you would go after identified

00200

primary site using MRI or any other technologies out there? How resource intensive can these be? Would you do that routinely, would you do it in selected cases, so could you comment about that? And if you can't identify the tumor origin, would you aggressively try to do that? Dr. Wong.

DR. WONG: From a surgical perspective, I would only do it if I thought it would make a difference in what I could treat the patient with, so that's probably not a great question to answer from my clinical perspective.

But I will mention that I think a lot of times, that will have been done up front, to try to determine the origin before or during the undergoing pathologic testing. I think if someone presents with adenopathy, I think the search for the primary begins at that point, and not necessarily after the identification of the TOO, so to speak.

DR. BEYER: I was just going to say that, jumping off from that, when we're talking about tumors of unknown origin, these patients present with metastasis and get biopsies and

00201

get an HME stain and say oh, it looks like a cancer. Lots of things happen before the immunohistochemistry is done, lots of things happen before you even get a test back, and those things probably do already include, you know, imaging studies, PET MRs, CTs, those studies. Also, you know, in a woman with an

axillary node, you're going to do a mammogram. These things have already been done before it gets called a TOO. So I think it probably, the answer to your question is it probably has already been done, we've already looked at the pancreas, we've already looked at the lung, we've already looked at the breast, we still don't know where it is. The true unknown primary cancers are the ones where the primary is probably involuted, and we may never find it.

DR. SEDRAKYAN: Thank you for the clarification.

DR. SARTOR: So, Dr. Sartor, just sort of a minor clarification. There are a lot of tumors that start out as unknown primary, exactly as we've heard, and you then rapidly undergo assessments where the biopsy might be

00202

positive. And so yesterday, I got a report on a fellow who looked like he had a primary in his liver and we believe it to be metastatic, so we scoped him up and down, and that would have sort of been fairly logical, to think that a colorectal cancer could have metastasized to the liver, or from the colon into the stomach. Well, those tests are now negative, so he falls into a true CUPS category, as opposed to a lot of people who start there, and you quickly narrow it down with either mammograms or other methods.

DR. REDBERG: Dr. Howard.

DR. HOWARD: This is kind of a question and kind of a comment, but when it comes to widespread use, I don't see why we couldn't potentially say well, now we are going to use this genetic stuff as the primary indicator and do all this stuff after the genetic test.

DR. BEYER: David Beyer. Conceivably you're absolutely correct, particularly if it was available quickly. You know, if you were able to quickly have an answer, oh, this is breast cancer in the liver, then you just

00203

spared that woman a colonoscopy, an EGD, and God knows what else, if you can quickly answer that question. I don't think we're there yet, but absolutely, that would be a game changer.

DR. REDBERG: Any other comments? Yes, Dr. Conley.

DR. CONLEY: Yeah, this is Barb Conley. I was listening to all of the conversation thus far today. The point of any

and all diagnostic tests when you don't know where the primary is, is to find something that might have a good response to whatever treatment is out there, so to the extent that things would be useful, would be when you find that.

DR. REDBERG: And it sounds to me like those trials we're waiting for, or that information we're waiting for, but currently we don't have any sort of treatment-directed answers on either survival or quality of life. As you see, Maria is now passing out clickers, and that's because we are getting close to the voting questions. So if any of the panelists have any additional comments or questions you want to raise, we can do it now,

00204

and then we'll turn to the voting questions.

DR. STECKER: I'm sorry, I need to raise a question and make a generic comment before we vote. Two and three, questions two and three on our green sheets, the difference, only difference I really see is question two talks about whether there's sufficient evidence for genetic testing of tumor tissue to affect health outcomes, and question three talks about whether there's sufficient evidence to conclude that genetic testing improves overall health outcomes.

DR. REDBERG: Okay. We won't get to question three unless we have scored 2.5 or more on question two.

DR. STECKER: I'm just wondering, I'm trying to make a distinction between the two. Effects could mean, am I voting on --

DR. REDBERG: Effects could go either way, it could be a net harm or net benefit, so that's any effect, and then question three is specifically in the positive category. Are there any other questions or comments? Okay. So, I am going to read over the introduction, and again, I want to thank the

00205

speakers for the presentations in the morning, but the morning was the time for the presentations and the discussion and open public comment, the afternoon is time for panel discussion and then moving to the voting questions.

And so we are, as you know, we looked at two genetic tests today, the DNA- or RNA-based test to predict the likely tissue of origin in patients presenting with a cancer of unknown primary site, referred to as CUP, and

then we also talked about fluorescence in situ hybridization tests for cancer or pre-cancer in patients with atypical squamous cells of unknown significance or low-grade squamous epithelial cells in cytologic specimens from the uterine cervix. And so when we vote, we're going to vote first on the first set of data for the data set for CUP, and secondly on the FISH test.

And just a reminder, to address the clinical validity of the tests, the outcomes of interest of CMS for FISH include histologic confirmation of higher-grade cervical intraepithelial neoplasm on biopsy, overall

00206

survival, mortality, avoidance of harms of antitumor treatments, quality of life and others. And to address overall health outcomes, the outcomes of interest for CMS for CUP include tumor recurrence, overall survival, mortality, avoidance of harms of antitumor treatments, quality of life and others.

And you are going to be voting on a one to five scale, one is low confidence, three is intermediate, five is high, and you can vote any of the integers, one through five.

So the first voting question is, how confident are you that existing evidence is sufficient to confirm the clinical validity, defined as how reliably test results are associated with the presence of the disease for target condition of each of the following? And first we'll vote on the DNA- or RNA-based testing to predict tissue of origin for CUP, and you should use your clicker to vote.

MS. ELLIS: I'm sorry, we just need one minute, we seem to have a computer glitch.

(Pause.)

(The panel voted and votes were recorded by staff.)

00207

DR. REDBERG: So, it looks like the vote was a mean of 3.25, which is pretty close to intermediate. Usually we discuss the vote, but maybe I will finish part B and then we can discuss the vote, okay? So I'm going to, I will ask the second part of the question, I'm not going to read it all again, and then we'll go down and discuss the votes. And so part B is the same, how confident are you that the existing evidence is sufficient to confirm the clinical validity of FISH testing for cervical cancer/pre-cancer in patients with atypical squamous cells of unknown significance/

low-grade intraepithelial squamous lesions?
(The panel voted and votes were recorded by staff.)
DR. REDBERG: And so that came in lower, in the low confidence or low to intermediate. And so now, Art, do you want to start and talk about 1.A, and then 1.B?
DR. SEDRAKYAN: Sure. I was influenced by the fact that there seems to be some utility to these tests when trying to identify the tissue of origin, tumor of unknown origin. And while the evidence is low to

00208

moderate quality, I mean, in some situations it seemed to help to identify this primary site better and there is some evidence for that, so I wanted to err on the side of positive potential for these tests to have, and I voted three.
For FISH, there was very limited evidence presented to us to make any statements and conclude anything, so I was much less confident for FISH.
DR. BEYER: David Beyer. I --
DR. SEDRAKYAN: I voted one for FISH.
DR. BEYER: David Beyer. I was actually fairly convinced that the testing for CUP in carcinomas of unknown primary was able to identify tissues of origin where the tissue was known, and to me that's an important thing to say, you know, that you can identify pancreas as pancreas, you can identify thyroid as thyroid, and I thought that the evidence presented was fairly convincing on that. I think that the ability to then identify a truly unknown, in a patient where it's truly unknown, I don't know that we've answered that question, I think there is still some uncertainty whether

00209

you can make the easy step from one to the other, but I was convinced enough that I scored it as a four.
On the FISH, I was a little less convinced about the ability to meaningfully identify something and make a statement. I gave it a little more benefit of the doubt, but I scored it as a two, however.
DR. REDBERG: Thank you, and just, can you first state, starting with Dr. Blegen, how you voted on 1.A and 1.B, and then tell us the reasons.
DR. BLEGEN: Okay, sure. This is Dr. Blegen, and I scored for the CUP issue a four and for the FISH issue a two, and my sentiments

are similar to the previous speakers, that there just seemed to be sufficient evidence, not overwhelming but certainly sufficient evidence to think that the CUP test would actually improve things, whereas the FISH tests do not look like they would.

DR. REDBERG: Thank you. Dr. Gutman.

DR. GUTMAN: Yeah. I was a little bit more skeptical on the first test, I actually remain uncertain. I agree that when your

00210

tissue of origin is known actually, I believe there is probably 85 to 88 percent concordance. It's not clear to me when the tissue is unknown, whether that falls to 75 or 55 or 15, so I didn't buy it, I actually put a two because I think it's plausible but I couldn't go above that. And I gave the ISH a two also, just because I thought it was plausible, but nice try, no cigar.

MS. ELLIS: Excuse me, panel members. Because we are, the meeting is being webcast, could you please state your vote first into the mic so that everyone can hear you, and then you can do your discussion, so if you would say 1.A and what your vote is, and then 1.B, what your vote is. Thank you.

DR. HOWARD: 1.A I voted a four, 1.B I voted a two. With regard to CUP, I thought the evidence where they tested CUP on specimens with known tumor site, the CUP, I think the test performed very well, was able to identify a high proportion of those, I found that relatively convincing. For the FISH test, the number of studies seemed to indicate that the specificity was fairly low and most of the

00211

studies tested were of intermediate outcomes rather than CIN3, so for that reason I was only able to give that a two.

MS. MASSEY: This is Pamela Massey. I voted a three for 1.A and a two for 1.B, and my reasons are, have already been mentioned by previous speakers.

DR. NOWAK: This is Jan Nowak. So, for 1.A I voted three and for 1.B I voted two. For CUP testing, I think the evidence is reasonably good that the test can demonstrate similarity to known tumors, I believe that the tests do what they say they can do, but I have some questions about what that similarity means and maybe even the degree of similarity as to what that means biologically and what it ultimately means clinically, but I believe the

tests do that. There could be more evidence, so I could have gone to four, but I put down three.
For FISH testing, several things. One is, for me, cervical cancer screening refers to cytology, and once you start talking about looking at tissue, that's histology and that's a different usage, and the discussion here got

00212

a little bit complicated because both kinds of things were included and different kinds of FISH tests were included.
That they might have some relevance for the evaluation of ASCUS or for LSIL, I may acknowledge that there may be some utility there, but the available tests that we have for cervical cancer screening have been clinically validated on thousands, tens of thousands of women through the ALTS trial, and for any test to displace that or to even try to show equivalence will be very very difficult, and I certainly didn't hear anything to say that these tests are anywhere near there. So I gave it a two, and I thought maybe that was being a little generous.
DR. REDBERG: Thank you. Dr. Rizzo.
DR. RIZZO: Hi, this is Doug Rizzo.
My vote was, on 1.A I voted a three, and on 1.B I voted a one. The reasons have largely been articulated. I also, I struggled with the methodologist in me struggling against the biologist/clinician in me and the methodologist kind of won, particularly on 1.B. I think knowing the validity for CUP testing against

00213

the tumors that we already believe we know the answer on is important, I think adding ten percent, if that's really an accurate assessment, adding another ten percent of patients for whom it would be unknown and converting them to a known is potentially valuable inasmuch as that's really a well done study against a gold standard, but I wasn't really convinced about that. Biologically, though, I think that the promise here, it's just not conclusiveness.
I really did not feel that the data to support the FISH testing was sufficient. I think the best category was in terms of LSIL but the others have much less data, I think the appropriate thresholds are not well determined, and I don't think this is an area where we could not get better data.
DR. SANDERS: So, I voted three for --

DR. REDBERG: Dr. Sanders.
DR. SANDERS: Dr. Sanders, I'm sorry. I voted three for 1.A and one for 1.B, and my rationale for 1.A was primarily determined by the evidence that I heard in the technology assessment and that I found fairly persuasive,

00214

that in fact these tests are able to identify what they set out to identify. I also was a little bit on a fence between a two and a three, if I could have been a two-and-a-half I would have been a two-and-a-half, but I was persuaded in part also to go three by the results from the comparative effectiveness study from this morning, and also by the somewhat, the heartfelt comments from the gentleman from Louisiana about the difference that it made in his outlook on life having an answer, rather than remaining in the kind of unknown category, and I allowed that to tip me into being a three.
For question 1.B, I found that the presentations were for the most part pretty cohesively indicating that the evidence base is premature and that there may be potential, but that we're not there just yet.
DR. SARTOR: Dr. Sartor. I had a four for 1.A and a two for 1.B. My rationale for the four was that I was reasonably convinced that a histologic arm could be established, and certainly it's not with high confidence. We heard about the probabilities that the test

00215

would work, but nevertheless the concept that as we go from protein to RNA and other genetic-based testing that we could refine what we currently do is certainly plausible, and I think the data would support that.
With regard to the FISH testing, I did not have confidence and I'm really more of a 1.5 than a two, that we really were detecting what we wanted to detect, and that there were a variety of observational variations that were important, and when you put it in the context of a complex algorithm, I was struggling to find its place in the algorithm.
DR. STECKER: Eric Stecker. For 1.A I voted three and for 1.B I voted two. The three vote for DNA and RNA testing for CUP was based on the concept that I feel pretty confident that it's correctly identifying the tumor of origin in a setting of known tumor origin. Whether that is the same for unknown tumor of origin is difficult and potentially unknowable,

strictly speaking. I think what additional data could convert me to a three or a four or a five would be treatment evidence that tumors identified as

00216

X, Y or Z respond conventionally or fairly conventionally to the usual treatments for X, Y or Z cancer. So if there were some treatment data to validate that, then my vote could have been higher; absent that, it was a three. For FISH testing for cervical cancer, my vote was a two. I think that the tests are heterogeneous, the results are heterogeneous, and so I didn't have any more confidence than a two.

DR. WONG: Sandra Wong. For question 1.A I voted a three, for question 1.B I voted a one. The rationale has largely been delineated by previous speakers. I will add that I think the technology assessments really helped inform this vote. I think the methodologic basis for the tests are at a point now where my votes were a little lower, I think that if the methodology were to be strengthened that those numbers could move up, but I think the current state of the evidence doesn't support a higher number than what I voted.

DR. MARCINIAK: So, Martin Marciniak. My vote on 1.A was a three, my vote on 1.B was a two. A number of my comments have already

00217

been stated. What I will say, it's important for me in terms of the links, and so we have the validity story, but more importantly, how does that link to other things in terms of treatment outcome? I think that as we take a look at the evidence paradigm of importance, with all the evidence discussion, it would be well worth the effort to put some time into that to help people understand whether this gives us a plus or minus sequentially in terms of the value that therapies are providing above that which is already in place today.

DR. CONLEY: Barb Conley. For 1.A I voted a three, for 1.B I voted a one. For 1.A, this was based on what I thought was reasonably good evidence that tumors of known origin could be detected, as well as the technological assessment and some other data coming out that showed that many metastatic tumors do kind of track with primary tumors molecularly. And then for 1.B, I was not convinced and I was a little disappointed that there really wasn't much in the literature to support

a clinical validity assessment at all.
DR. ROSENTHAL: Dorothy Rosenthal.
00218
For 1.A, a two. I think philosophically it's a great idea to be able to tell what the primary tumor is, but I think we need greater analytic validity and some very good, well controlled studies, whether they're blinded or not or randomized, it's something up to the people who design good studies, but we need some.
And for 1.B, I've already stated my views very loud and clear, loudly and clearly, and I voted a one.
DR. REDBERG: Thank you. So, we will go on now to Question 2.A only, because only A had a score of more than 2.5. So, for this the question is, how confident are you that there is sufficient evidence to determine whether genetic testing of tumor tissue affects health outcomes, including benefits and harms, for patients with cancer whose anticancer treatment strategy is guided by the results of DNA- or RNA-based testing to predict tissue of origin for CUP?
It could affect it negatively or positively, benefit or harm.
(The panel voted and votes were recorded by staff.)
00219
DR. REDBERG: Okay. So the votes on this is a low to intermediate confidence, it's a vote of 2.083, and now we'll do the same thing, we'll go down the line and discuss the reasons for the vote.
DR. SEDRAKYAN: Art Sedrakyan, I voted two. And the reason I voted two is that I was really not convinced that we have substantial data, substantial evidence that the outcomes change. The only study that has been presented and discussed potentially in colorectal cancer and in overall population, I looked at that paper thoroughly and I couldn't even find the baseline demographics presented for the groups that are being compared, so I have serious analytic concerns about the studies that show these benefits that have been discussed.
Now one can think of this as even a small study can show you benefits, so it's a huge effect size. I would say the biases can be so substantial in selecting the patients and who is likely to agree to get the CUP, it might be healthier patients, patients who are younger, there might be a lot of factors that can lead to choice to get this CUP or to who it

00220
is offered, and while people who haven't got the CUP might have been those who had hopeless conditions regardless, serious comorbidities and wouldn't be able to go for chemotherapy to start with.
So there were a lot of concerns analytically, and we know talking about statistical issues there were many statistical problems there as well, so I wasn't convinced that there's enough evidence for me to vote higher than a two.
Why I voted two, not one, a major determinant of that was the fact that the harms seemed to be not a big problem here. The determination of TOO, and these patients are getting chemotherapy anyway, I'm convinced that therapy will lead to more harm, or differential therapy based on identification of TOO will lead to more harm. So harm is not the issue, but we just don't have any benefits documented here, and that was critical for me, that we really need to have more data, more evidence before we can vote about the ratio of benefits to harms, which is question number three.
DR. BEYER: David Beyer. I voted a
00221
three on this, and I actually voted twice, I'm not from Chicago but I voted twice. They tell me my second vote is the only one that counts. I started with a two, and I started with a two because the question says how confident am I that there's sufficient evidence, and when you come back to the word evidence, I don't think we saw sufficient evidence. I think we saw some provocative studies and enough to give, to get me off the one and to convince me that there may well be something here. I actually believe that this does affect health outcomes and I believe that this does make a difference, but I don't think the evidence is there and I think as the question is worded, we just don't have the evidence, and I think we really need the evidence.
Having then decided I was going to vote a two, I actually bumped myself up, also out of respect for the fact that it does clearly impact patients to know what their diagnosis is, and I thought that was enough to push me from the two to the three.
DR. BLEGEN: I'm Dr. Blegen and I voted a one on that, and it came down to how
00222
much confidence I had in the evidence that this

affected outcomes, and in terms of strength of evidence I just did not feel that it was there, so I had to stay with a one.

DR. GUTMAN: Yeah. I think the whole thing is, is this an unusual test where the clinical utility would trump the fact that you don't actually have clinical validity and may not ever be able to get clinical validity because of the nature of the beast, and I thought that it's very promising but that there just wasn't enough to go anywhere, so I voted two.

DR. HOWARD: For 2.A I voted two. Like the other panelists, I wrestled with this question. Certainly the potential is there but when we're talking about issues of evidence, you know, if not a randomized controlled trial, then maybe some modeling studies, some type of decision analysis that makes the logical leap between identification of the tumor site, changes in treatment therapy, and ultimately patient outcomes. If I'm having to make those logical leaps in my mind as opposed to actually seeing someone do it, I can't vote with a high

00223

degree of confidence that this is a test that is going to improve patient outcomes.

MS. MASSEY: Pamela Massey, I voted a two, and like others, I was really disappointed to see what there was out there for evidence, because I so much wanted this to be a positive. I think for people who have this disease, that we've heard very articulately how much it can affect one. And we're here because of science and evidence, but there is also something called hope and belief, and that also contributes to the healing process. And, you know, if you don't know what your disease is and you don't have much hope, then that can influence your outcome. So I really would encourage that the scientific community go and look and see if we can build a better case for this. Thank you.

DR. NOWAK: Jan Nowak, I voted two. Like the other panelists, I really believe that this is going to fly, but the question isn't about what I believe, it's about evidence, and so I had to hold back and I gave it a two. I also share Dr. Sedrakyan's comments about some of these effects may be negative and we don't

00224

know that, so the effects may be there but they may not all be positive, so that's important to find out.

DR. RIZZO: Hi, this is Doug Rizzo, I voted a three. I also wrestled, again, with this issue. I found the data to be provocative but I would agree that not sufficient in terms of the evidence. I wrestled with the fact that I believe some tumors definitely respond better to targeted therapy than others and being able to identify those may offer promise for patients, which I think is important, and I hope that we can develop appropriate evidence as this field evolves.

I also factored in the fact that having better assurance that the toxicity is being matched to the appropriate disease and treatment could be very important for avoiding harms in patients. I don't think that there are necessarily harms to patients from these tests, there may be harms from a health system perspective, but I think that's a different question.

DR. SANDERS: Amy Sanders, and I voted a two, and I think that many of the other

00225

panelists have touched already upon my basis for my vote. Primarily it's that there is some provocative suggestion that there may be some effect on outcomes, including survival, that moved me from being a one, which would have stated I had no confidence, to two, meaning that there is some suggestion that there may be an answer to this question at some point, and as others have said, it's unclear whether the predominance will be positive or negative, but that wasn't really the question.

DR. SARTOR: Oliver Sartor, I was a three. The data is in transition. We're not charged with voting on the future nor our wishes, but rather the current evidence, and the current evidence does not need in my mind the level that I think would enable me to vote any higher than a three, just by the fact that I would like to. I think it probably does make a difference in a subset of patients, I think we need to see more data in order to convince me that that is correct.

DR. STECKER: Eric Stecker, I had exactly the sentiments as Dr. Sartor. However, it led me to a vote of one for question 2.A,

00226

but with exactly the same rationale. I have experience with evidence-based coverage recommendations and I'm on a committee in the state of Oregon, although I've never participated in an oncology review, and I would

say that this is by far the lowest quality of clinical evidence, not to say anything about the treatment or the future of it, but the current state of the evidence is by far the lowest quality of evidence that I have ever participated in, and because of that I voted it a one.

I would say that it's exciting technology, I think there's tremendous, as a cardiologist I think there's tremendous potential for success. I would differ with some on the panel in saying I think there's also unrecognized probability of harm. Every day in literally every clinic, I see at least two patients who are on drugs that, I'm managing the risk of those drugs, when anywhere from ten to 20 years ago those drugs were thought to be a slam dunk case based on clinical and biological rationale and that clinical trials don't even need to be done, but

00227

when a randomized trial was done, it was shown to be actually harmful. So I would say in a setting of diagnosing these disorders, in using chemotherapeutic drugs and going down different diagnostic and therapeutic paths, we could end up in a place of harm.

DR. WONG: Sandra Wong, I voted two on this question. I completely agree with the concern about the harms here. I think that we are letting practice get ahead of the evidence, and I believe that is a potential for large harm. I think there is a lot of gaps in the evidence that need to be addressed before we head any further in terms of practice.

DR. MARCINIAK: Martin Marciniak, my vote was a two. To bridge off Dr. Howard, I think there were opportunities to clarify some of what we were thinking, or some of the things we thought we might want to see, hope gets me to a two. But the tipping point really is how do you clarify that which is happening, whether it be through decision modeling or other tools, to help us understand the evidence better in truth with words, where the public is going to go in the future, so that's what led me to my

00228

vote.

DR. CONLEY: Barb Conley, I voted two. We have some provocative evidence out there but I think there's more development to be done to eventually seek out the subset of patients that may benefit.

DR. ROSENTHAL: Dorothy Rosenthal,

also a two. Nothing more to add.
DR. REDBERG: Thank you all. We will go on to the discussion questions. I'll just comment, I thought that was a very helpful discussion, we really appreciate all the information from the presenters and the panelists and everyone who came to talk this morning. You know, I think these are important questions because clearly, since I finished medical school more than 30 years ago, there are a lot more tests we can do and a lot more possible treatments, and patients in general are getting a lot more tests and a lot more treatments. But I think what we really need to focus on as a clinician is that missing link of how are all of these additional tests and treatments leading to better outcomes for our patients, because I think we have a lot of

00229

because we can, we do, and perhaps with good intentions, but I really feel it's our professional responsibility to inform our patients what we know and what we don't know about tests and treatments.
And, you know, the fact that it might be hard to enroll in the clinical trials suggests that we have certainty, at least to me currently, and I think for some things that may be true, but for a lot of treatments, and I think some of what we talked about today certainly falls into it, we don't have certainty, and that is the rationale for a clinical trial. And while we all hope that a new test will lead to better outcomes, what we hope and what happens are very different, and the only way we can find out what will actually happen is by doing randomized clinical trials. And we certainly, you know, cancer is a very large clinical diagnosis and it's a terrible thing for patients to hear, and we want to offer them treatments that work, but we can really only do that, I think, when we have good quality data to show that it does change outcomes. And I was a cardiology fellow when

00230

we were giving all those type 1A drugs because of the thinking that because they suppressed PVCs they were helping patients live longer, and then we did the study and found out that there was an increased death rate.
And I know that sometimes studies can be positive, but I think what we heard today was that there is a lack of data on clinical outcomes. We talked a lot less about sort of

the laboratory standards, which I think are another issue for genetic testing because of the lack of a requirement for FDA approval, and the importance of having reproducibility and variability, particularly when there is direct to consumer advertising as we heard there was for some of these tests. I think we're asking patients to make decisions that are very difficult for us as physicians to make, and certainly hard for patients to know what to do with that information.

So I think it is, you know, we have a lot of desire to get things to patients quickly, but I think we only want to get things quickly when we know they're of benefit, and otherwise we have to think about the harms and

00231

the benefits and what are we trading off, and I think we had a good discussion of that this afternoon and, today, and this morning.

So the discussion questions that we have, we have three discussion questions, and it's please discuss whether the evidence as presented may be generalized based on each of the following factors: Regulatory status of the tests such as, e.g., is it FDA-approved or cleared, as we talked about, versus laboratory developed tests. Site of testing, university medical center or commercial laboratories versus community-based laboratories. And patient subgroups within the Medicare beneficiary population, e.g., age.

And obviously, I think, just looking at the last one, we heard that 13 of the 14 studies included patients of the age, although I wasn't clear on what the average age was in the studies, although I would assume that cancer patients in general would be older and more likely to be of Medicare age, I think not so much true for the cervical cancer studies.

But at any rate, does anyone have any comments on any of the issues here related to

00232

regulatory status of tests, the site of testing, or patient subgroup data.

DR. BEYER: David Beyer, just a comment. I didn't hear anything presented that would make me think that this was not generalizable. I didn't really hear much discussion about FDA versus non-FDA, and I don't think that the panel was thinking in terms of this test versus that test. I think that, you know, the thinking of the group was fairly generalizable. Probably also, you know,

looking at site of service, I don't think, I didn't hear anything that would make me think universities are going to be able to do this better today than community-based labs. And similarly, I don't think this was age-specific for patients, or Medicare-specific, I think these were generic comments that the panel had to offer.

DR. REDBERG: And I don't know that we heard the data today, but I do wonder about the reproducibility and variability data, because I know that for some of the genetic testing, there were studies where they sent the same sample to different labs, and I didn't hear

00233

evidence either way on that.

DR. SEDRAKYAN: I think if anything, it will add more uncertainty to what we've already talked about, and will probably pull our votes lower than how we voted.

DR. REDBERG: Dr. Rosenthal.

DR. ROSENTHAL: In preparing for today, I have to say that I really learned an awful lot that made the hair on the back of my neck stand up a bit, and that was the laboratory-developed test concept. I've served on a PLIAC committee, I've worked with the FDA on panels, Steve Gutman and I go back to when we were children together, remember, in Poland?

DR. REDBERG: Last year?

DR. ROSENTHAL: Last year. The thought of doing a test so critical to a patient's wellbeing that didn't make it through the FDA and the approval process, I think is a joke. And those of you who know me know that I don't mince any words and I really say this from my heart, but to go ahead and develop these potentially wonderful molecular-based tests is one reason I haven't retired yet, it's so exciting, this is what we have been working

00234

for all my career, all my life. But not to go through a well structured analytical validation, reproducibility, et cetera, et cetera, et cetera, and then a clinical trial. I don't care whether it's not going to be designed, you know, the way it should be because we can't get the patients to sign on because they don't want to be randomized. Modeling, there's all kinds of other ways of getting at a very valid test, a very well controlled test, so that any laboratory anyplace can do it. I think when there's a laboratory-

developed test and you have to send the specimen in and a single laboratory does it, that's a conflict of interest right there, and I'm just really appalled. And I would say that it's time -- I mean, these tests are becoming very very esoteric, and attempting to have centralized laboratories to do these tests because it's economically more feasible. On the other hand, it just, it gets into a real quagmire of people ordering them who have an investment in the laboratory, et cetera, et cetera, and I think we really have to be

00235

very careful about this and look out for the patients.

And advertising to the patients directly, it should be the laboratories who make the decision, along with the clinicians, what test to order, and not the people who have been watching late night television who come in and say I want this test. Okay.

DR. REDBERG: Dr. Conley.

DR. CONLEY: I just want to make a comment on the generalizability, just maybe even a clarification. We have been requesting evidence on clinical utility that is fairly rigorous, but each platform would have to generate the evidence on clinical utility, you know, because one platform does not automatically translate to all such tests of the same kind.

DR. REDBERG: Dr. Nowak.

DR. NOWAK: So, I guess I should comment. I am not shy about putting together a laboratory-developed test and taking responsibility for its validation. Many of the molecular tests that are available now for solid tumors, EGFR, all of these things that

00236

we've used for years, have been made available as laboratory-developed tests, and there are proficiency test results available that demonstrate that the performance of laboratories who do these tests do these things very well.

On the other hand, the test that has gone through the FDA, that does say something, and so that kind of test does have some special status. Maybe that test is not the best test. There are examples of FDA-approved tests, antibody testing for EGFR before you treat colon cancers with Erbitux was not a good test, so just because a test may have been approved doesn't mean that it's a good test or it's a

bad test, or it's the best test. So it is possible for laboratories to develop good tests and validate them appropriately. Several speakers this morning made the comment that really through immunohistochemistry with all the antibodies that we use, that there's variation in how these things are done and how they're performed and how they are interpreted, and I think there's an interesting parallel here.

00237

Antibodies, they became available gradually. One at a time they would come out and as soon as they became available, they were always the best thing under the sun, they had the highest sensitivity and the highest specificity for this tumor or that tumor, and everybody started purchasing the antibody and using them, and then a year later you'd find out, well, it's not really that sensitive, that specific. But the difference is that those things, those antibodies were always being tested and retested by different entities, by different laboratories and eventually we worked it out. And laboratories do struggle to ensure that performance of testing used for these antibodies is uniform across all laboratories, they do make that effort. In this circumstance we have, these are rather complicated tests and they're developed by single laboratories who have a proprietary interest in doing these things only in their laboratories, so they don't benefit from the kind of development that has gone on with the development of monoclonal antibodies for use in cancer assessment, that bias is sort

00238

of there. And there's still -- I kind of made a point this morning that it would be nice to see a direct comparison, and I understand that there's indirect comparisons and the tests seem to compare very well, but nonetheless, if you had things done in multiple places it serves as a crosscheck, it allows for proficiency testing, it stimulates improvements and further development, and we don't quite get that here, and I'm not sure if that doesn't require addressing in some other manner, maybe some better kind of evidence or maybe more evidence. That's the question in my mind, but I don't question that the laboratories who do these things do them well, I think they can.

DR. REDBERG: Dr. Stecker.

DR. STECKER: Dr. Conley brought up two points, one directly, one a little indirectly that I would like to just comment on. The first is that it's very difficult for these highly complex proprietary tests to generate outcomes evidence. You know, is it practical when there's 30 different companies that can't fund a big trial each, is it

00239

appropriate to report outcomes data? I would say in the setting of something that is very targeted for certain disorders and it results in high risk treatments, I would say absolutely you should require that.

Now in the setting of, the indirect point is, you know, just laboratory testing in general, what's the evidence that checking serum potassium is useful? I mean, we check serum potassium level for a million different reasons, so it's completely impractical to have outcomes data with regard to the outcome improvements from checking a serum potassium, but this is very different. This is a complex genetic testing that's guiding, directing therapy, so it's a very different concept.

DR. REDBERG: Dr. Sartor.

DR. SARTOR: I want to elaborate on that a little bit. You know, a lot of the testing that has recently been approved and funded has been tied to very specific therapeutic advance. So for instance, we could look at the so-called ALK inhibitors in a particular subset of non-small cell lung cancer, and the companion laboratory test is

00240

married to that. And I think one of the things that is a bit problematic from a developmental perspective is we would like to see and would really, I think everyone on the panel would prefer to see clinical outcomes being affected by the testing in a very clear manner.

Yet, the amount of testing, the expense of follow-up for conducting such an analysis may in fact be prohibited unless you tie it to some therapeutic advance. So at the same time that we're demanding more evidence, I would like to sort of say with a caution that the evidence we're going to demand is going to be very costly and very time consuming, and it's going to result in very expensive tests as a consequence.

DR. STECKER: Public funding of research would obviate that, so if there were a mechanism whereby CMS directed indirectly

funded research or where NIH would sufficiently fund it, there would be funding trials, or AHRQ would sufficiently fund it, you could fund the trials that were to maximize the public health, which would be high valued things. What you're talking about, things automatically becoming

00241

very expensive is because only the high profit margin/high revenue things will get funded, not only, but, you know, on average.

DR. SARTOR: The expectation of my profit margin --

DR. REDBERG: I really look at it from the point of view of taking care of our patients and unless we're doing trials, even if they're expensive trials, you know, we could be harming patients, there could be net harm. In this case we're talking about treating them with very toxic drugs, shortening what may already be a very short life and making it -- and that's a net harm. So I really look at it from the patient's point of view, but you could argue that we are spending billions of dollars on treatments that we don't know are beneficial, and so to say that is too expensive to study is, I think, really not in that group of patients at all.

DR. SARTOR: I was perhaps thinking of the funding. I actually completely agree with your point. I think that many of the patients who receive CUP therapy, which is a platinum-based therapy today, I realize the

00242

studies that it's going to take. I wish public funding were an option, I anticipate NCI funding cooperative groups, and we have to really set a priority list and this hasn't been a high enough priority. But, I mean, I want more studies, absolutely, I want more studies.

MR. MARCINIAK: For me the question as an economist, you know, we've talked about randomized trials, and I think that's a great way to get to a discrete answer. But we also don't spend a lot of time talking about (inaudible) and so when you think about the pyramid, you think about the CTs, you think about the observational studies and you think about other decision models, there are opportunities for us to use some of those tools better than we currently do, and a good example is with Sierra Medicare. We collect large panels of data across a number of different individuals, and some of these questions might be elucidated from that if the registry part of

the questions were framed a bit differently. You know, did you get a test, yes or no, what type of test did you get, and then you start building out what those pathways look like.

00243

That actually reduces the cost of doing some of this and makes the answer perhaps not perfect, but a little bit more accessible for groups like us who are trying to wrestle with the complex question.

So that would be one way to think about this a little bit differently, just to tilt the paradigm away from what we know to be the high cost ticket where we would like to see more government funding or we'd like to see somebody pick this up, there are things that we could go do more practically with things that are in place.

DR. REDBERG: That does lead into the next discussion question very nicely, and that's to identify and discuss any evidence gaps in assessing the outcomes of interest for CMS, and those were the clinical outcomes that I had read earlier for both the DNA- or RNA-based testing for the tumor of origin in CUP and the FISH testing for cervical cancer in patients with ASCUS and LSIL.

So, I think we've had some discussion on the evidence gaps. It seemed to focus around more clinical outcomes, some laboratory

00244

testing. Eric, did you have some comments?

DR. STECKER: Yeah, just real briefly, I'd just like to second that I feel like a randomized trial, and actually I think randomized trials are overemphasized, I don't think they're at all practical. I echo what a lot of you have said, in many cases once you do it, how is it generalizable, you filter out so many patients in a clinical trial that it doesn't apply to clinical practice anymore, so I would echo Dr. Marciniak's comments completely. It wouldn't have taken much from an observational well controlled trial standpoint to move my vote out of a one up to a three, for instance.

I know that oncology has a tremendous registry like SEER, and I don't know the mechanics of how this works, but with a good registry system you really ought to be able to match patients up who are getting different treatments, and get a stronger idea of whether it helps or not. In the absence of that I don't think it's appropriate to rely on hope

because of what we've already talked about, there's a lot of potential for harm.

00245

DR. MARCINIAK: You know, for us I think it's the level of creativity. I mean, the groups that are doing this type of evidence research are very creative in establishing these subgroups, and that's the reason why the entire order today, when they looked at where the puck was going, were they seeing more in terms of abstracts or other things for public knowledge absorption. And so oftentimes you don't have a lot of information about the pathology at the start, but the simulations actually help get you there, because with, you know, a bright group of people, physicians and others, you can usually toggle between, you know, do you think this will happen or do you think that will happen, and you just build it out and map it out and quantify it. So I think there are a lot of opportunities with groups to confront these things to help bring dialogue better, because as an economist I hope for threes and fours, but oftentimes I get disappointed in the lack of creativity to help frame the conversation that we're having, and so I go back and reflect on the TARs, you know, I got really disappointed looking at questions

00246

two and three because the stuff that is meaningful to me just wasn't there.

DR. REDBERG: Dr. Beyer.

DR. BEYER: This isn't quite as bad as being an orphan drug or orphan disease, but it is a rare disease and it's one that's going to be very hard to get enthusiasm from a lot of the larger clinical trial organizations. Radiation oncology has been very active in setting up a registry, the National Radiation Oncology registry is just getting on its feet. I know ASCO has been doing something with a registry, and registry studies, I think, are going to be increasingly valuable in the years to come.

I don't think it's going to help us answer these particular questions, I think these are too focused, too specific and too granular, and I doubt that the registries are going to be collecting data on these particular questions, it's going to be very hard to do in a registry. I think it will be very hard to do in a randomized clinical trial. I do think some clinical trials of better control groups can be done and need to be done.

00247
I mean, clearly, what I want to see is some convincing evidence that we made a meaningful difference in terms of survival. I want to see convincing evidence that we've impacted the quality of life, that that increased three months of survival is not spent in the intensive care unit but is actually something that patients can enjoy. Or, I would be very convinced if we identify a patient who just plain shouldn't be receiving cytotoxic chemotherapy because we know it's not going to work, I think that's a home run. You know, it can take a lot of different forms, but there are some meaningful outcomes that we can identify.

DR. REDBERG: Art.

DR. SEDRAKYAN: I wanted to comment about simply registry versus registry-based study, just a clarification from our end. Dr. Redberg and I were in a meeting certainly with, organizing the meeting on registry concept, making clear that the concept of registry-based study comparative and matched is different than having the registry address a particular question.

00248
And I agree, Dr. Beyer, that certain questions you can address with a registry, but if you need a comprehensive answer you need to design a study based on the registry but you need to collect the additional information. So let's treat the registry as that general system, not necessarily to address a particular question. I think that's an important distinction in what we're arguing for, for being a representative system even nationally, but something that we can have and as with anything, you can ask a lot of questions and maybe get some preliminary answers, but not specifically targeted to answer these kind of questions, unless you design a study based on the registry.

Another detail I wanted to comment on is this resource use, that I think we need more evidence, and this is an important gap. If the test is applied early or used early, can it reduce the additional workup and resource use, and it wasn't clear to me. If I were to have evidence today that it has potential and document that it can lead to fewer MRIs, fewer tests that are being done, even earlier in the

00249
process, I would think it would have

substantial outcomes even if they were not patient outcomes, but really important outcomes for a health care system in general.

DR. REDBERG: Dr. Wong and then Dr. Howard.

DR. WONG: I do a lot of health services research and so I'm aware of cancer registries and the limitations thereof. I think in the current system of cancer registries there would be no way to collect this type of data. I draw an analogy to the very successful coverage with evidence collection with the PET, where a separate registry was established for the purpose of actually gathering data to determine whether the PET changed decision-making by clinicians. And I think that's the key question here, is whether this test actually changes the decision made to treat, and with what drugs are being used to treat. I think there's a tremendous opportunity here to do that.

I think some of the onus probably falls upon the laboratories doing these tests as well, because at a certain point the test

00250

results get sent out and then it stops, and if data were collected as to what clinicians do with the test results, I think we'd gain a lot from that, and I don't think that's currently being done, I think it's being focused on establishing a diagnosis, but we need data after that as well.

DR. REDBERG: I'm going to comment on what you said and then recognize another few panelists. But I agree with you that the PET registry was a good example of that, and that does lead us to discussion question six, which is what can CMS do to encourage development of additional evidence relevant to these questions, but I would suggest that I don't really think change in treatment itself is a sufficient outcome, because I can change treatment based on no evidence and I can change treatment and it would have no benefit to the patient.

I think there actually has to be some evidence that the change in treatment led to better net outcomes for the patient, because I hear the argument a lot, you know, I changed my treatment. We change our treatments every day

00251

in medicine, and it's not always based on evidence and it's not always good for our patients, and if we want to use the data for

that, I think we should have a real clinical output that is meaningful to the patients in terms of things that help them feel better or live longer.

DR. WONG: Right. I don't want you --

DR. REDBERG: Well, question two and question three, I really think it was does it affect health outcomes one way or the other, and question three was does it improve health outcomes.

DR. WONG: You're right.

DR. REDBERG: So, does it improve health outcomes should be the focus.

DR. WONG: So, I totally agree with you. I don't mean to say that changing the treatment should be the outcome being measured here, but some outcome of interest having to do with the genetic test could actually be easily measured.

One thing that might be of interest based on the discussion here was the value added question, and that is, does the result of

00252

the genetic test actually change the final call? In other words, if there was an IHC call, a pathology result based on IHC, does the genetic test actually change what the final pathology is being called after somebody already looked at IHC, just as an example. I agree that for a genetic test like this, that maybe the change in drug used or not used may not be the correct endpoint. That was the endpoint, you know, whether it changed the decision, that was the endpoint for the PET, but it certainly could be different here. I'm not suggesting that we necessarily go that way, but that is a thought that would allow us to gain more evidence.

DR. REDBERG: Sure, and that's a good example, I think in the PET registry CMS did approach trying to gather more evidence.

Dr. Howard, then Dr. Stecker, then Dr. Rizzo.

DR. HOWARD: I'd like to put in a word for the much maligned historical control studies, and we all know examples of those that led to incorrect inferences. We saw an example of a study like that today, but there were

00253

enough questions about the comparability that I think we were reluctant to base our votes on it.

But in any situation where you have a sharp and sudden increase in the way the

patients are treated, that almost creates like a natural experiment to look at the effectiveness of that treatment, and there have been examples where studies have taken advantage of sharp breaks in treatment patterns to identify the impact of treatment. So an example would be, if you looked at colorectal cancer, cases of metastatic cancer diagnosed between the mid '90s and the mid 2000s, you clearly see for patients receiving chemotherapy, the survival curve starts to creep up as newer therapies are introduced, whereas for patients who are not treated with chemotherapy, survival rates are flat. So I do think there are some really good opportunities here to exploit historical data, but it obviously has to be well done and trying to identify something like a concurrent control group.

DR. STECKER: I was mistakenly called

00254

on, but when you're talking about a natural history experiment, it is a very different thing than what we've seen today. You know, when you're taking a group of patients and there's some natural thing that happens and abruptly therapy changes, and then you can look at the outcomes, that's very different from taking a group of patients from one center and saying how did the average patient follow at a different point in time at different centers. So, I would malign it again.

DR. REDBERG: Dr. Rizzo.

DR. RIZZO: So, I wanted to echo the fact that we don't always have to be confined to thinking about randomized clinical trials, and we should be flexible about study design. As the project director for a large outcomes registry on transplantation, we've been able to use that registry in order to in fact address a CV question right now, and let's not forget that registries help us plan better clinical trials as well by understanding potential effect sizes, enrollments, et cetera. I think the things that we need to understand better about these tests, and

00255

perhaps this is a bit summarizing what others have said, but there's a much better understanding of the incremental improvement in the diagnostic, and that's getting at the value added. I think that we shouldn't underestimate the opportunity that we could make the diagnosis more quickly with these tests, you

could do less testing which, it's possible, we don't know, I think we need to learn that, or these tests could have such a test profile that they're actually more readily standardized across all the clinical and pathological studies that we're discussing, so all of that is useful to have.

And then, I think to look at changes in response rates by therapy assigned, looking at changes in which therapy is assigned, whether that affects the harm of the toxicity of therapy, and of course we need to understand better, that at least for those who have a better response, that that improved their survival, we made the assumption that converting a patient from CUP to a patient with metastatic disease is going to, metastatic disease of a certain histology is going to give

00256

a better outcome, and I think we still have to prove that.

DR. REDBERG: I'll make a comment because we haven't talked very much about FISH testing, that sort of what I got in terms of additional evidence that we need from FISH testing -- well, first of all, I got that we're really doing pretty well with treatment of cervical cancer, particularly in the US, and we're doing pretty well with identification and screening for cervical cancer, but that we, it seemed to be even though we didn't really focus on it, we're doing a lot of colposcopies that were of unclear benefit, because we seem to be doing colposcopy on a very early stage, you know, on the atypical and the CIN, all of those things in your algorithm seemed to lead to colposcopy, which clearly has some harm, anxiety, and probably leads to additional procedures.

So in terms of evidence gaps, it seems to me that we should be looking more in terms of not just the testing, but should we, is colposcopy as they're currently using it really leading to clinical benefit or should we be

00257

doing less of it. I don't know whether we need to be, I don't think that's going to affect the Medicare population very much since most of the incidents are within, at least in the under 65 age group, but that does seem to me to be an evidence gap that was identified today besides what we've already discussed with regard to the FISH testing and how it fits in or adds, if anything, to the testing we already have for

cervical cancer treatment.
But I think -- were there any other comments? Yes, Dr. Sartor.
DR. SARTOR: You know, I think it's also important to emphasize in addition to things like colposcopy, I think there are many instances where cytotoxic therapy has no benefit, and I think we should be open to the possibility that the treatment of patients with standard of care raised more harm than benefit, and that there is a possibility that these tests could eliminate futile therapy. And I was actually thinking for a number of these metastatic cancers of unknown primary that our standard platinum-based regimens bring more harms than benefit, it's only for a subset

00258

where the opposite is true.
DR. BEYER: David Beyer. I just want to make the points from your comment specifically on the applicability of FISH in the Medicare population, and it should not be lost on anybody in this room that Medicare sets the policy, or CMS sets the policy for Medicare for the over 65, but there are a lot of people, a lot of plans that follow Medicare's guide, and while it is Medicare that they're talking about, it has some halo effect.
Those patients who are 55 and having atypia are the patients who are going to be Medicare beneficiaries being treated for advanced cervical cancer. So, you know, if Medicare is looking for a reason to be interested in this further as it develops, I would argue that this makes a big difference for Medicare in the preventative sense.
DR. REDBERG: Absolutely, and there are Medicare beneficiaries that are under 65 as well, because they come in through SSI qualifications.
So, as I say, I think we all learned a lot about genetic testing, both for cancer of

00259

unknown primary as well as cervical cancer. I think we identified evidence gaps, particularly with regard to the importance of clinical trials and clinical outcomes, and understanding how this testing helps our patients. I think in terms of, we heard a lot of discussion back and forth on randomized clinical trials as well as registries, I think clearly we have some examples of the PET registry and other non-CMS-sponsored registries that have helped inform treatment and could be modeled for

getting additional data in this area as well as other cooperative clinical trials. I would like to first of all thank Maria, some of us would not be here without Maria, and some would not leave here without Maria either, so thank you. And James Rollins and Louis Jacques, as well as all the presenters and the panelists, thank you all. I know this was a lot of work to go through everything and we've had a lot of success and we really -- I know for some of you, it was your first time here, I thought everybody really participated, engaged, and thought about the questions really seriously, and on behalf

00260

of CMS and MedCAC, we thank you. The meeting is now officially adjourned.

(Whereupon, the meeting adjourned at 3:00 p.m.)