00001

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

January 27, 2010

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

Reported by:
Paul Gasparotti

00002

Panelists

Chairperson
Clifford Goodman, Ph.D.

Vice-Chair
Saty Satya-Murti, M.D., F.A.A.N.

Voting Members
Phyllis Atkinson, R.N., M.S., GNP-BC
Catherine Eng, M.D., F.A.C.P.
John Cox, D.O., F.A.C.P.
Josef E. Fischer, M.D.
Daniel F. Hayes, M.D.
Nora A. Janjan, M.D., M.P.S.A.
Karen Kaul, M.D., Ph.D.
Karl Matuszewski, M.S., Pharm.D.
Maren T. Scheuner, M.D., M.P.H.
Steven Teutsch, M.D., M.P.H.

Industry Representative
Peter Juhn, M.D., M.P.H.

00003

PANELISTS (Continued)

Guest Panel Members
Elaine K. Jeter, M.D.
Elizabeth Mansfield, Ph.D.
William Pao, M.D., Ph.D.

Guest Speaker
Andrew N. Freedman, Ph.D.

CMS Liaison
Louis Jacques, M.D.

Executive Secretary
Maria A. Ellis

00004

TABLE OF CONTENTS

00005

00006

PANEL PROCEEDINGS

(The meeting was called to order at 8:20 a.m., Wednesday, January 27, 2010.)

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 the evidence, hear presentations and public comment, and make recommendations concerning whether the results of pharmacogenomic testing affect health outcomes of patients with cancer when used as a guide for certain drug treatments.

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 employers' financial interests. Each member will be asked to disclose any financial conflicts of interest during their introduction. We ask in the interest of

00007

fairness that all persons making statements or presentations also disclose any current or previous financial involvement in development or marketing of testing supplies, kits or

testing equipment that provides testing services intended for clinical or research use in pharmacogenomic testing of human tissues, including tumor tissue from patients with cancer, or that develop policies for such testing. This includes direct financial investment, consulting fees and significant institutional support. If you haven't already received a disclosure statement, they are available on the table outside of this room. We ask that all presenters please adhere to their time limit. We have numerous presenters to hear from today and a very tight agenda and therefore cannot allow extra time. 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

00008

panel please speak directly into the mic, and state your name.

For the record, voting members present for today's meeting are: Dr. Saty Satya-Murti, Phyllis Atkinson, Dr. Catherine Eng, Dr. John Cox, Dr. Josef Fischer, Dr. Daniel Hayes, Dr. Nora Janjan, Dr. Karen Kaul, Dr. Karl Matuszewski, Dr. Maren Scheuner, and Dr. Steven Teutsch. 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 scores will be available on our website following the meeting. Two averages will be calculated, one for the voting members and one for the entire panel.

I ask that panel members please speak directly into the mics, and you may have to move the mics since we have to share. If you require a taxicab, there is a sign-up sheet at the desk outside of the auditorium; please submit your request during the lunch break. Please remember to discard your trash in the trash cans located outside of this room. And lastly, all CMS guests are only

00009

permitted in the following areas of the CMS site: The main lobby, the auditorium, the lower level lobby, and the cafeteria. Any persons found in any other 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. Louis Jacques.
DR. JACQUES: Thank you, Maria, and good morning. Just as a brief comment, this meeting today is a follow-up from a meeting that we had in the MedCAC in February of 2009 where we asked the committee to make recommendations to us about the desirable characteristics of evidence related to diagnostic genetic testing. This particular meeting is taking some pharmacogenomic test contexts that are somewhat more developed than some others and essentially saying gee, compared to what we recommended a year ago, in fact does it look like the evidence generation is in fact consistent with what the MedCAC's recommendations had been.
Just as a reminder, we do not have an

00010

open national coverage determination today on any of the drugs that we're talking about or any of the test platforms that will be discussed today. So, I would nonetheless advise everyone to, you know, heed the MedCAC's recommendations carefully.
Now I'll turn things over to Dr. Cliff Goodman.
DR. GOODMAN: Thank you, Dr. Jacques. I do need to reiterate the importance of our speakers staying on time. We have the little light system, as Maria Ellis mentioned, and I will not hesitate to give you a warning when there are about one or two minutes left, and we certainly need to stick to that, and of course our panelists are cognizant of the importance of getting to the point as well. We've got a lot of ground to cover today, we need to make sure that all the questions are heard, all the invited speakers are heard, and all the other speakers that have signed up to speak are heard as well.
I will start off and introduce myself, and have a brief statement about potential conflicts of interest. I apologize ahead of

00011

time if mine tends to be a little longer.
Cliff Goodman, vice president of the Lewin Group. I should mention, the Lewin Group is a subsidiary of Ingenix, which is a health information and analytics firm. Ingenix is a wholly owned subsidiary of United Health Group. United Health Group has a bunch of subsidiaries, one of which happens to be United

Health Care.
I have no personal financial interests to declare but I do want to disclose that as a salaried employee of the Lewin Group, as a salaried employee of the Lewin Group, I am involved in projects having to do with preparations and papers to some organizations that are supported at least in part by some of the industry companies involved in this field. We've done white papers on such subjects as the value of laboratory medicine, how to determine that, the relationship between laboratory medicine and health care effectiveness research. I've done work for the Association of Community Cancer Centers, the National Conference of Cancer Networks, both of which are supported in part by some of these

00012

companies. I should also disclose that we're currently under contract to the HHS Office of the Assistant Secretary for Planning and Evaluation, to do a study on genomic data sharing. Those are my disclosures. Dr. Satya-Murti.
DR. SATYA-MURTI: Saty Satya-Murti. I'm a neurologist and I have been a contractor medical director for a number of years. I consult for industry, academic societies, and on this panel as well. I have no conflicts of interest.
MS. ATKINSON: Phyllis Atkinson, gerontological nurse practitioner. I have an independent health care practice and I have nothing to disclose, no conflicts.
DR. COX: John Cox, medical oncologist, practicing with Texas Oncology in Dallas, Texas. No personal conflicts.
DR. ENG: I'm Catherine Eng, a medical director at On Lok Senior Health Services, and a clinical professor of medicine and geriatrics at UCSF. I have no conflicts of interest and no disclosures.
DR. FISCHER: Josef Fischer, a

00013

practicing surgeon at Harvard Medical School. I have no conflicts, no contracts, no disclosures.
DR. HAYES: I am Dan Hayes, I'm a medical oncologist from the University of Michigan Comprehensive Cancer Center. I serve on the AEB and have stock options for two diagnostic start-up companies, neither of which has a product being evaluated at this meeting. I receive research funding from Veridex, which

is the diagnostic part of J&J, but none of their products are being evaluated here. I am cochair of the ASCO college of medical pathologists review committee for many markers, including HER2, which is being discussed here. I'm also on the National Comprehensive Cancer Network committee for breast cancer and breast cancer markers, which has reviewed several of the markers being reviewed here today, and I'm also on the National Academy of Biochemists guidelines panel, which again has reviewed several of the markers being reviewed today. And finally, I participated in research involving at least three of the markers being presented today, but

00014

I have no specific financial conflicts of interest related to today's discussions.

DR. JANJAN: My name is Nora Janjan. I'm a radiation oncologist from M.D. Anderson Cancer Center. I do consulting work with Ikon and Accuray, but none of my consulting consists of anything related to pharmacogenomics, and I have no personal financial conflicts of interest.

DR. KAUL: I'm Karen Kaul, a clinical professor of pathology at University of Chicago and an anatomic pathologist and molecular pathologist at North Shore University Health System. I do periodic speaking in the area of molecular diagnostics, but have no conflicts of interest pertaining to the things being discussed today.

DR. MATUSZEWSKI: Karl Matuszewski. I'm a pharmacist by training. I'm a vice president of Gold Standard, that's a subsidiary of Reed Elsevier. I am also the editor in chief of Clinical Pharmacology, an online drug information resource.

DR. SCHEUNER: I'm Maren Scheuner, I'm a clinical geneticist. I work at the VA

00015

Greater Los Angeles, I work at the RAND Corporation. I have funding from the VA, the CDC and NIH, and have no conflicts, no disclosures.

DR. MATUSZEWSKI: I forgot to mention, I also have no financial conflicts or disclosures.

DR. TEUTSCH: I'm Steve Teutsch. I'm a retiree of Merck and have certain stock options, but I'm unaware that they have any conflicts with this. And I was on the EGAPP program which did evaluate irinotecan and

UGT1A1, and I chaired the Secretary's Advisory Committee on Genetics, Health and Society, but not involving these specific issues.
DR. JUHN: Peter Juhn, president of Therapeutic Resource Centers and Medco Health Solutions. I'm the industry rep on the panel today, and my conflict is I'm employed by Medco Health Solutions and we do have a set of services that provide testing, pharmacogenomic testing to various clients.
DR. JETER: Hi, I'm Elaine Jeter. I'm a pathologist and I am a contract medical director for Palmetto GBA.

00016

DR. GOODMAN: Any conflicts? I'm sorry.
DR. JETER: I'm sorry. No conflicts.
DR. MANSFIELD: I'm Elizabeth Mansfield, I'm the director of personalized medicine in the Office of In Vitro Diagnostics at FDA. I have no conflicts and no financial interests.
DR. PAO: I am William Pao, an associate professor of medicine and cancer biology at Vanderbilt University Medical Center. I'm also the director of Personalized Cancer Medicine there, in which we are interested in performing mutational profiling of tumors prior to therapy. I'm also a consultant with a molecular diagnostics company which does sell kits for this area of testing, and also has the rights to a patent application for testing certain mutations.
DR. GOODMAN: Thank you, panel. And I would remind the panel, for the sake of our highly devoted court reporter, always please speak directly into the microphone, don't speak faster than at least I can understand, and this also goes for all the people present here today

00017

at this meeting. If you do want to say something, and we do want to hear what you've got to say, especially if it's concise, please do wait to be recognized and come to a microphone. Otherwise, our court reporter may not see you and may have difficulty hearing you, and if it's important enough for you to say, it's important enough for us to have this recorded in the transcript of the meeting.
I believe I will now turn it over for the CMS preparation and voting questions, is that correct, Maria?
MS. ELLIS: Yes.
DR. GOODMAN: All right. So this is

Lisa Eggleston, who will begin the presentation.

MS. EGGLESTON: Good morning. My name's Lisa Eggleston, and once again, welcome to our MedCAC on pharmacogenomic testing for anticancer therapy. I am serving as lead analyst for this project along with Dr. Jeff Roche, and I also at this time would like to acknowledge the director of the Division of Items and Devices, and that is Dr. Jim Rollins, if he would raise his hand. Thank you.

00018

At this time I'm actually going to skip to our Medicare voting scale to apply, because prior to reading the MedCAC questions, I would like to explain the voting scale. The panel will vote on the questions using a confidence scale with a range of one to five, with one being low confidence and five being high confidence.

The first voting question is, number one: How confident are you that there is sufficient evidence to determine whether pharmacogenomic testing affects health outcomes, including benefits and harms, for patients with cancer whose anticancer treatment strategy is guided by the results of testing as described below? Please consider this question separately for each test in the following clinical situations.

CYP2D6 for breast cancer patients who are candidates for tamoxifen.

UGT1A1 for colon cancer patients who are candidates for irinotecan.

HER2/neu for breast cancer patients who are candidates for trastuzumab.

BCR-ABL for CML patients who are

00019

candidates for imatinib.

KRAS for metastatic colorectal cancer patients who are candidates for cetuximab or panitumumab.

Number two. For those items where the answer to question one is at least in the intermediate range, a mean score of greater than or equal to 2.5, how confident are you that pharmacogenomic testing improves health outcomes for patients with cancer whose anticancer treatment strategy is guided by the results of testing as described below? Again, please consider this question separately for each test in the following clinical situations.

CYP2D6 for breast cancer patients who are candidates for tamoxifen.

UGT1A1 for colon cancer patients who are candidates for irinotecan. HER2/neu for breast cancer patients who are candidates for trastuzumab. BCR-ABL for CML patients who are candidates for imatinib. And KRAS for metastatic colorectal patients who are candidates for cetuximab and/or panitumumab.

00020

Number three. How confident are you that these conclusions are generalizable to community-based settings and the Medicare beneficiary population? Number four. Please discuss any important evidence gaps and recommend how they should be addressed. Questions one through three will be voted on using the scale on the following slide, that I will show you again, and question four will be discussed during the MedCAC proceedings.

At this time I will turn the presentation over to my colleague, Dr. Jeffrey Roche.

DR. ROCHE: Good morning, everyone, and thank you very much for your service. Today the MedCAC panel is going to look at a question that has been increasingly important for the Medicare program in general, and that is to find better ways to successfully treat cancer on behalf of Medicare beneficiaries and their physicians.

As we consider this questions that Lisa has just looked at with us, in many ways

00021

it boils down to perhaps a more simple question. Can we say with confidence based on the available evidence that pharmacogenomic testing can guide cancer treatment decisions to improve patient outcomes? When physicians consider a treatment strategy, and perhaps this is a comment that goes beyond anticancer treatments, they must try to balance benefits and harms that might result to an individual patient from the effects of therapy. Given that anticancer drugs can be toxic to noncancer cells or other organs, and also that response to a given drug varies, at times widely among individuals, achieving this balance can be challenging. This slide lists some of what are a whole number of nongenetically-based tools that physicians use to evaluate a patient's

potential response so they can maximize overall benefit to each individual. Today we are going to consider some additional tools, pharmacogenomic tests that may, depending on the evidence that will be reviewed today, stand as additional assessment tools so that physicians can make the following decisions

00022

with confidence, selecting patients for whom the cancer drug will be effective, individualizing treatment regimens for indicating risks, or potentially treatments limiting adverse events for each individual. As we will discuss further and as some of the guest presenters will inform you, pharmacogenomic testing results ideally would help physicians predict from among the central group there of the patients with the same cancer diagnosis which patients are likely to benefit when given agents, which patients would in contrast have more of a combination of toxicity with relatively low benefit. Such consideration, such information would help physicians better plan treatment regimens and hopefully consider potentially the best treatment strategy for the individual. Now as was mentioned earlier, there are a great many genes whose variation does affect individual responses to cancer therapy, again, as part of the patients' overall response to therapy, which includes nongenetic factors. This slide focuses on colorectal cancer and the response of patients to therapy

00023

for that disease. We are not going to talk about all of these different genes or about the multiple interactions or pathways that are known to govern gene action. Instead, we have tried to choose a few hopefully well selected examples from which the MedCAC can consider the question of evidence confidence in the improvement of outcomes of patients with cancer. This slide shows five examples of the specific combinations of a gene to be tested, a cancer to be treated, and a drug being considered for therapy. Let me mention that we certainly hope the panel will use it's discretion to generalize beyond any one particular of these examples to look at more general questions or crosscutting issues that may be of importance in looking at potential future coverage decisions. Finally, as some of today's panelists

may recognize, members of a prior MedCAC panel have already looked at some of the issues involved in deciding which outcomes are the ones that really make an impact in assessing the benefit of a diagnostic test. As shown

00024

here and as is recorded for that 2009 meeting, MedCAC panelists expressed high confidence that improvements in patient-centered direct health care outcomes were most indicative of the benefits of such testing.

Among the direct outcomes that CMS favors, and I think this is true over a number of decisions over the years, better survival after diagnosis, or improvement in symptoms and function, are among the types of outcomes that CMS considers more impressive. I recognize that for many of the panelists this is already well known.

In conclusion, Lisa and I together with the rest of the CMS coverage team thank each of you for participating today, and we look forward to learning much from your discussions.

DR. GOODMAN: Thank you very much, Dr. Roche. And now we're going to move to Dr. Andrew Freedman, who is the chief of the Clinical and Translational Epidemiology Branch of the National Cancer Institute. Dr. Freedman.

DR. FREEDMAN: Thank you very much. I

00025

want to thank Dr. Roche for inviting me here, and first I want to say that I have no financial disclosures and no conflicts of interest to report. The customer has asked me to give probably an overview of this area and some of the research that's going on at NCI, some of the approaches that, scientific research that NCI is financing, and some of the ways it's translating into practice. I want to, I think this panel would be very informative to NCI to see how some of the research that we're funding is either getting into practice or not getting into practice and getting covered. So I'm just going to give an overview of the entire area and note that I'm trained as a molecular epidemiologist so I kind of see it through epidemiology eyes. So just to say, I'm at the Division of Cancer Control and Population Sciences at NCI.

So, let's start out with a quote from then Senator Barack Obama where he said, personalized medicine represents a

revolutionary and exciting change in the fundamental approach and practice of medicine and holds unparalleled promise for public
00026
health. This is of great interest in the new administration. Certainly with the appointment of Dr. Francis Collins, there is certainly a big push to look at this concept of personalized medicine or stratified medicine. I'm now going to talk about a little bit of background of genomic pharmacoepidemiology, some approaches we use to look at this, and then some of the things to consider when we try to translate this evidence from discovery to practice.
So, this is a graph that comes from the Journal of the National Cancer Institute just to show that in the last ten or 15 years, there has been an explosion of anticancer therapies, and a lot of these have been molecularly driven therapies, and certainly that's one of the reasons we're all here.
And obviously as we all know, some of these therapies can be extremely expensive, and you can see that for colon cancer, for six months of therapy it gets as high as $50,000, and of course it would be helpful to know who it's good for, when, at what point in their treatment, and how they should be used. So
00027
what we're really talking about is personalized or predictive medicine, some people call it stratified medicine, but if we have a group of patients with a certain cancer, we want to know who's going to respond to that treatment, who's not going to respond to that treatment in order to control the expense, and that's going to help inform us. We also want to know which patients will remain compliant and stay on the drug. So the various scientific research can inform us on how to determine which individuals will respond, who will have an adverse event.
This first slide shows the discipline of pharmacoepidemiology, and it is the study of benefits and risks of drug therapy outcomes among groups and subgroups of cancer patients, so we're really talking about age, diet, lifestyle factors, health status, and response to drugs.
Pharmacology is the other scientific discipline, for which I didn't put up a slide.
And then the third discipline or area is cancer pharmacogenomics, so here we're really talking how variation in the

individual's either germline or tumor genome
00028
are related to their metabolism and physiological response to drugs used in the cancer treatment, so we're talking about a number of different genomic variations. So, the way I like to think of it is that we have all these genes, immune as well as drug metabolism or drug receptors, we have the alterations in the tumor characteristics, we have the drugs or treatments, and then the clinical, environmental and lifestyle factors where a patient or a group of patients will report that the drug has an adverse event or a different outcome. And the example that I like to give is, certainly we're talking a lot about CYP2D6 today, tamoxifen is a drug that has been used for many years for breast cancer treatment. We've known for, clinicians for years have been using ER status to determine whether individuals should go on tamoxifen. Certainly we're talking about whether or not CYP2D6 genotypes inform us about responsiveness to tamoxifen, and there are some studies that indicate that antidepressants may interact with CYP2D6 and tamoxifen, and may be less
00029
effective. So all of these can each inform us as a clinician about survival, recurrence, and adverse events. So obviously, the goal is to optimize therapy so that the benefits will outweigh the risks, and this is just a slide sort of putting it all together. So, the scientific approach that has been looked at, certainly if we're talking about hereditary germline mutations, we're talking about alterations in DNA inherited from one of your parents and found in the DNA of virtually all of your cells. And then there's the acquired or somatic mutations; these are alterations in DNA that develop throughout a person's life, specifically in the tumor. So common approaches for, to look at the germline, the candidate-gene approach is one approach, and I'll just give you quick examples of each one of these. There's the candidate-pathway approach and then more recently in the last few years, a genome-wide association approach to try to identify genomic variation that can inform us of who might benefit and who might have an adverse event.
00030

So, the candidate-gene approach examines whether a particular DNA sequence variation is more frequent in patients who have a better or worse response, and we usually know something about these DNA sequence variations that can inform us, and certainly the example we're talking about today, with CYP2D6 we know that some people can be classified as extensive metabolizers, intermediate metabolizers or poor metabolizers, based on their genomic variation. In the case of tamoxifen, one of the active metabolites in tamoxifen, depending on the genotype of CYP2D6, the metabolites can vary and there can be quite a difference in the metabolism of this drug depending on your CYP2D6 genotype.

And one study that just came out in October in JAMA was very interesting because it was a very large study of over 1,300 patients, a retrospective study looking at time to recurrent events, pre-survival and disease-free survival looking at CYP2D6 and tamoxifen. And you can see there's an indication, at least from this study, that this study actually has the power to look at some of the issues that

00031

will be discussed later today, that depending upon your CYP2D6 genotype you can have definitely different outcomes.

The pathway-based approach examines biologically plausible associations between certain individuals and inconsistent clinical outcomes. It really supports the potential of looking at a range of genetic profiles to predict clinical outcomes. And I'm not going to go through this slide, but this is just to show you that 5-FU, one of the chemotherapy drugs that is used quite commonly, is involved with folate metabolism and different genes are involved in the metabolism of that drug.

And you know, here you can see, this is a study by Wu, and in the top graph you can see, just looking at one simple gene, you get some separation as far as survival outcome, but as you add additional genotypes, and you can see in the last one you have five different genotypes, you can see you have quite a wide spread. So instead of looking at one specific polymorphism on a specific gene, looking at combinations and looking through the pathway may be a more powerful approach, so that's

00032

coming this way.

The last one is the GWAS approach, and

that examines common variations for drug response by genotyping the entire genome and the SNPs across it. You don't have to look at every single SNP type polymorphism, all 10 million, but you can use tag SNPs or haplotype blocks. This is really a hypothesis-generating mechanism to identify genomic variations that we would not have identified in any other way, and this is what they usually look like. They call them Manhattan blocks, and this is really the first GWAS of cancer treatment, it was done out of St. Jude, and they looked at acute leukemia, and you can see that they had quite a few hits at the top here, but just to show you that one of these SNPs, there was a hematologic response and they actually saw some association for a specific marker. So that's coming this way as well as we try to identify more and more genes and their polymorphisms or their variations and how they relate to cancer outcomes.

So somatic alterations, we can also look at single gene alterations and we can look

00033

at protein expression. So for single gene alteration, one of the classic examples is HER2/neu trastuzumab as a receptor. This has been tested actually with the drug, and one of the major studies that came out of this was a study published in 2001 that showed that in those that were HER2 positive and had trastuzumab did much better as far as survival than those that did not. And I'll stop right there, because I know we're going to discuss that later.

The other one that we're going to discuss later is KRAS and EGFR inhibitors for metastatic colon cancer, and I will skip right to, just to show you that one of the first studies to show that KRAS, to indicate that if you were, if your tumor had a mutation in KRAS, you might not do as well for cetuximab. Actually it started with a study of just 40 patients and expanded to a hundred patients, and then we're going to see the data later today, what the evidence is to show what the relationship is between KRAS mutation and outcome in those patients treated with cetuximab and panitumumab.

00034

The other one I'd like to put up is Oncotype DX, and it doesn't quite fit with these groups, but it's a test that I believe now has been used in over 50,000 patients in

the U.S., I think I read that somewhere in the paper the other day. It's a diagnostic test that quantifies the likelihood of disease recurrence in women with early stage breast cancer, and it assesses the likely benefit from certain types of chemotherapy. And what it does is it looks at 21 genes and gives you a recurrence score, and based on that recurrence score, physicians may decide to advise chemotherapy or not to have chemotherapy. The middle range of the recurrent score clinicians are not so sure about, so actually NCI has started the TAILORx trial, I think they started enrolling in 2006, that's really looking at that middle range recurrence score, and whether or not patients would benefit from chemotherapy or not. NCI has also started other biomarker-based studies like the MARVEL trial looking at lung cancer, and EGFR biomarkers.

So, I wish I could end it there and

00035

say that a patient walks into the pharmacy and says here's my sequence and here's my drugs, but it's a lot more complicated than that obviously, and that's why we're here today. So clinical utility, analytical validity and clinical validity, are key issues that obviously are going to be discussed today.

This is a slide where I took the words right from Steve Teutsch from EGAPP, and the question we do want to answer is, does testing for genomic variations lead to an improvement in outcomes or are testing results useful in clinical decision-making? In terms of analytical validity, we want to know how good is the test in identifying the genomic variation?

Clinical validity refers to how well does the variation predict metabolism or drug efficacy. But really the bottom line is clinical utility and does the testing, this genetic testing, whatever it is, influence clinical decision-making, can it improve or worsen outcomes, does it improve clinical outcomes compared to not using the test, are the tests useful for medical, personal or

00036

public health decision-making, and really what are the harms associated with testing and subsequent management options.

And this is just a diagram to illustrate how this really works, and I adapted this from a diagram in Steve Teutsch's paper,

where you can certainly see where analytical validity is looked at, whether or not you can detect the variation. Clinical validity is in a sense the association to the prediction, and clinical utility, do the benefits outweigh the harms.

So you know, many people have talked about the levels of evidence that we need to make decisions to translate this into practice or coverage. Certainly randomized controlled double blind studies, everybody would like to see that. A lot of times we don't have that, sometimes we have to do retrospective analysis or clinical studies, sometimes cohort studies, and sometimes there's modeling and meta-analysis. We may not always have a new prospective trial to test for a new genetic variation, whether that affects outcomes, and we're going to have to think about new ways to

00037

examine the evidence and to interpret it.

So here's a couple sentences that I took from David Atkins' paper, which was a very nice paper in Medical Care in 2007 and he really talks about what a clinician would need to make a decision for an individual patient. Certainly to find out if a drug works there's a need for randomized controlled trials, but for a specific patient they may need data from trials and registries, or cohort studies, to see if this patient is really going to be able to tolerate the drug or adhere to the drug. Cohort studies and subgroup analysis will tell them if their particular patient and their particular patient's genomic profile and clinical profile would benefit them, whether the harms would outweigh the benefits or the other way around, and we get that from case control studies as well and in the end, sometimes modeling or qualitative studies of patient preferences. So the point of this is that there's a lot of data that we need to consider, or the clinician needs to consider in making those decisions. I think, you know, there's been a lot of discussion over the last

00038

few years of what evidence is needed and for who.

So I tried to put together a comparison of the pharmacogenomic markers in cancer treatment and I'm obviously still working on this and trying to get some clarification. You will see on the left-hand column, those markers that are in orange, those

are somatic alterations that are based on the tumor, and the first one, estrogen receptor has been useful for years, I'm not sure there's anything really different in the test for estrogen marker from these other markers, especially of somatic alterations. And then the ones in white are germline alterations, and you can see that some are mentioned as having an FDA label required or recommended, and some are used in practice, some are not labeled but used in practice, or the other way around. And certainly the column that's missing here is coverage by Medicare, by HMOs and so forth. So, this is another slide I stole from Steve Teutsch, thank you, that really looks at decisions, stakeholders and translational medicine, and this is just to show you that

00039

each one of these stakeholders looks for different evidence, you know, each one is focusing on a different aspect. Certainly the FDA focuses on efficacy and safety for drugs and so that's in green, the green is where they really focus on. NIH is interested in a range, they're not as interested in how much the drug costs or necessarily the clinical situation or the legal situation, or less so, but again, very interested in efficacy, safety, effectiveness and comparative effectiveness. And you know, if you skip to clinicians and patients, they really want to know if the drug is going to work in the general population or general practice, and certainly they're worried about costs in a clinical situation. So it's just to show you that each one of these stakeholders is interested in different evidence and that we need to design our studies that will certainly satisfy most of these, and where there are gaps, try to fill those gaps. So, I just wanted to take one slide to mention at NCI we are trying to address some of these gaps, and two years ago we started a Trans-NCI Pharmacogenomics and

00040

Pharmacoepidemiology Working Group, and this, you know, NCI is very partitioned, but we tried to bring together the basic scientists, the clinical trialists and the population scientists all together in one room to try to make recommendations on what, how should we fund research, what should move forward. And it was very interesting and we're finalizing our recommendations right now of where we need to go in this field to translate these great

discoveries we're making into practice. So, everybody I'm sure is aware that the IOM came out with the comparative effectiveness research report, I believe close to, in June of last year, and these are some of the cancer priorities. I want to skip right to the third one and one of our major priorities, I think they had a hundred priorities, but this was one of the major ones, was to compare genetic and biomarker testing and usual care in preventing and treating breast, colorectal, prostate, lung, and ovarian cancer, and possibly clinical outcomes. They saw this as a very important issue that needs to be addressed.

00041

And thanks to stimulus funds, we were able to recently fund seven GO grants, that's grand opportunity grants, specifically to look at genomic personalized medicine technologies in cancer. And you can see some of these sites and the information on some of the research that they've been doing is on line at the website, where you can find out exactly the types of things they're doing. And it's really the first step where NCI is really trying to push the agenda to try to compare the genomic technologies to standard of care, and what improvement.

And just to give you a little taste of some of the projects, many of these funded studies are going to look at proof of principle studies. We're also going to do some evidence synthesis and some modeling, really to find a roadmap for comparing effects and research in genomic personalized medicine.

And I think that's all I have, so I finished ahead of time so we can get back on schedule.

DR. GOODMAN: Thank you very much, Dr. Freedman. If you would just stand at the

00042

podium just for a moment, please, that's a splendid presentation. And we just of course wanted to remind all of our panelists that as we sweep through the day, including your presentation this morning, to keep in mind what our questions are that we'll have to answer, and those questions deal primarily with impact on health outcomes or, as you put it, clinical utility, as well as the accompanying evidence, the adequacy of that accompanying evidence. Do we have our next folks here? So we might proceed to that. Dr. Freedman, you're

here until, is it noon?
DR. FREEDMAN: No, I'm here all day.
DR. GOODMAN: So knowing that, we promise to get back to you today, and if you'll have a seat back down on the floor, again, thank you for a superb presentation.
And we will now move to our presentation of the technology assessment, I believe, and that will be from Dr. Tom Trikalinos. He's the assistant director of the Tufts New England Medical Center EPC, that's evidence-based practice center, EPC, and assistant professor of medicine at Tufts. As

00043

he may well tell you, the Tufts New England Medical Center EPC is one of about 13 or 14 EPCs under contract to the Agency for Healthcare Research and Quality. As part of their broader effective healthcare program, the EPCs are often tasked under contract to generate technology assessments and evidence reports in response to requests from CMS, other agencies and stakeholders, if I understand that correctly. Dr. Trikalinos, welcome.
DR. TRIKALINOS: Thank you, it's very nice being here. I don't think it's reflected here, but I recently became co-director of the EPC, and I'm going to speak about our systematic reviews on selected pharmacogenetic tests for cancer treatment, and this is work that has been led by Dr. Teruhiko Terasawa, but I will be doing the presentation on his behalf.
This is an obligatory disclosure statement. All the authors of this report have no financial, business or personal interests to disclose, and none of the investigators has any affiliation or financial involvement related to the material we are presenting here.
A bit about the background to this

00044

presentation. The coverage and analysis group at CMS has requested a systematic review from the AHRQ on pharmacogenetic tests that may be pertinent to the Medicare beneficiary population. And after discussions between AHRQ, CMS and Tufts, three pharmacogenetic tests were selected for the review. These tests were selected because they are perceived to be relevant to the Medicare population and they mostly evaluate relatively common disease conditions, as you will see.
So, we're talking about genetic tests, and I will start with a definition of a genetic test according to the relevant National

Institute. So according to NHGRI, a genetic test is defined very broadly as the analysis of human DNA, RNA, chromosomes, proteins and certain metabolites in order to detect heritable disease-related genotypes, mutations and so on for clinical purposes. So this is a very very broad definition and encompasses a lot of things.

Here we're talking about pharmacogenetic tests, so essentially we are especially interested in tests that identify a

00045

patient's differential response to specific therapies, and thereby they guide patient management. The word phenotypes here may be strange, but I think they refer to, let's say things like enzymatic expression would be the phenotype expression of some genetic alteration, that's what I presume. This is provided from part of the definition that they give.

So essentially here we discuss three pharmacogenetic tests. We have to do with CYP2D6 and tamoxifen for breast cancer, KRAS and anti-EGFR antibodies for colorectal cancer, and BCR-ABL1 mutations with the drugs being tyrosine kinase inhibitors, and chronic myeloid leukemia, CML, is the third condition. So CML was not as prevalent as the other conditions, however it was deemed of interest to the Medicare population.

This set of three systematic reviews posed common generic matters and there are variations of the matters that are particularly applicable to each one of the topics, so I'm going to describe to you the generic matters first.

00046

All three topics try to address the same four key questions, and the key questions are: Does the genetic test result predict response to therapy? What patient and disease-related factors affect the test results, their interpretation or their predictive response to therapy? How does gene testing impact on the therapeutic choice? And what are the benefits and harms or adverse effects for patients when managed with gene testing? So for each one of the three topics, you would imagine instead of talking generically about generic tests that we would substitute the specific test, and the same for the condition of the therapy.

Our systematic reviews are based on

electronic research and we searched OVID MEDLINE from its inception through the last week of August of 2009. The search parameters are a bit complex and slightly different from topic to topic, but in general they are combinations of key words that have to do with the gene, the disease and the drugs of interest. If you are interested, I can give you the exact search strategies, but they are

00047

listed in the appendix of the TA, of the technology assessment.

So, what sort of studies did we want to include? First of all, we were principally interested in both comparative and non-comparative studies that assess a test. And what do we mean by comparative? By comparative studies that would mean studies where patients were managed, where there's a direct comparison of a test and treatment strategy versus no test and conventional treatment strategy. So you could imagine this as an RCT perhaps, where people are randomized into being managed with the test versus usual care. We didn't have any comparative studies and essentially we had in reality only studies that were non-comparative when it comes to using versus not using the test. So all patients get the test, and then these studies are giving us information on the accuracy of the genetic test to predict endpoints. In the previous design we would be talking about the differential, the different clinical outcomes that we would be able to see with different management strategies.

00048

Further, another thing is that the test may be applied only in people who are treated or in people who are both treated and untreated. And we are talking about pharmacogenetic tests, and pharmacogenetic tests are tests that try to predict a differential response to treatment. So essentially amongst people who have received a drug, what studies do is that they can assess the endpoints, the clinical response perhaps, among people who have the genetic factor and people who don't have the genetic factor. Now the same thing can be contrasted in people who have not been treated, who have not received the drug. And why do we need this information or why do we want to see these kind of studies, because if I see the differential, if I see the ability of the drug or of the

genetic factor to predict the drug endpoint in people who have been treated, I should not be seeing a similar predictive ability of the endpoint in people who have not been treated. The point is that if I have information on the ability of the test to predict endpoints in both treated and untreated

00049

patients, I can see the interaction between treatment and genetic information, so I can really assert whether the predictive ability of the test is actually related to its ability to assess effects of treatment rather than to simply predict a clinical outcome irrespective of treatment. So studies both in treated-only people, only treated people, and studies both in treated and untreated people where we could assess for interactions were included.

What sort of outcomes are of interest? We're mainly interested in patient-relevant outcomes and these outcomes are mortality and disease progression. However, we soon realized that we didn't have a lot of literature on these kinds of outcomes and we decided to also study outcomes that had clinical importance, and these outcomes are generally treatment failure by imaging criteria for one of our topics, KRAS and colorectal cancer, or laboratory criteria for another topic, BCR-ABL1 and CML. We wanted studies that had at least ten patients in total, so that they have at least some patients and have a minimum sample size where one could do any calculations. And

00050

at least for key question two, we demanded interaction analyses, and for key question two we wanted to see whether there are any patient-related factors or any disease-related factors that affect the ability of the test to predict results or treatment, so for these kinds of studies we wanted interaction analysis.

From each study we extracted information and organized the result by the type of outcome. So generally we had analyses of cumulative events, for example, reference to specific time points or time-to-event analyses which gave us hazard ratios. And we had these kinds of analyses for mortality, disease progression, and treatment failure by imaging or lab exams.

Especially for the third item, treatment failure, we decided to also look at this from a test perspective, and by that I

mean that we tried to calculate the sensitivity and specificity of the genetic test to predict treatment failure. Essentially this is where we had the most data and it's not very meaningful to try to do this for the other

00051

outcomes of mortality and progression. And this is a slide that gives, that summarizes a bit about how the predictability of tests can be assessed. So this is the sensitivity versus one minus specificity plot, and each point would be a study. Studies that are perfect in terms of their predictive ability, they have perfect sensitivity and perfect specificity, so they would line up, they would all gather in the upper left corner of the graph. A study that would fall on the diagonal would be a study that has no predictive ability, and you can see it is indicated there by no better than chance. And then studies that are very specific but not very sensitive are in this region and the ones that are sensitive but not very specific are in the other region. The shaded areas have a specific meaning, but I'm not going to go into that now.

Further, we identified reports that had at least partially overlapping populations, and this is important because if you're going to do any meta-analysis and pull the studies together and pool their effects, you want to

00052

avoid duplication of information. It's very common that many of the studies out there come from the same centers, they may have the same patients in different follow-ups, or they may have collections of patients and many patients may be common. So we tried our best to identify studies that are overlapping and whenever this happened, we generally kept the larger study, the biggest study. One may argue that this way we are missing sample populations that are not reported by themselves in a single report, so we avoid duplicating information but we may be missing some information.

Meta-analyses were performed where appropriate, and we performed meta-analyses of odds ratios to measure strength of association and we also performed what is called a bivariate meta-analysis of sensitivity and specificity, an advanced process to summarize diagnostic test performance. And this was done with proper analysis and basically with random effect models.

So I'm going to go through the results now, and essentially you don't have to look at this table carefully. The only thing that you

00053

have to see is that the first topic of breast cancer has 13 studies and the other topics have 31 studies that were eligible in the end. It will also tell you that we had no studies that address key questions two, three and four, so essentially the things that we are discussing have to do with key question one which is, what is the ability of the test to predict a pharmacogenetic response?

So the response from pharmacogenetic tests are going to be, I'm going to give you a very very brief background, two words about the gene and the disease. I'm going to describe the eligible studies and their characteristics. And then I'm going to give you the evidence of key question one, give some topic-specific conclusions, and then at the end discuss some cross-cutting methodological issues.

So the first topic is CYP2D6 and tamoxifen for breast cancer. So tamoxifen is one of the popular treatments for breast cancer, and the issue is that tamoxifen itself is not the active drug. When a patient takes tamoxifen, it's metabolized in the patient's body and one of the key enzymes that do this

00054

activation, let's say, of tamoxifen is CYP2D6. Now, CYP2D6 is a gene that has many genetic variations, it has many single mutations, polymorphisms and other variations, so there are a bunch of CYP2D6 alleles, and this is just to show you that there are many many alleles that have been identified.

Now these alleles may or may not be part of the enzymatic activity of the gene. So some alleles have, you see the first line, they may have normal enzymatic activity, so a patient who has these variations would not, it would not have increasing enzymatic activity. However, there are other alleles that have decreased or even null enzymatic activity and therefore, these are the ones that theory suggests may play a role in the differential response to tamoxifen.

This is an obligatory slide. Whenever we do a systematic review we have to show you how many abstracts our search returned and how many texts, how many publications we reviewed in full text in order to select the final ones. And I'm not going to go through the numbers; in

the end we had 13 studies that were relevant
00055
here.
Now, as I said before, CYP2D6 has a lot of alleles, and these alleles can be analyzed in many ways. Now each one of us has a genotype, each person has a genotype, and imagine that there are many many alleles and there are many many more genotypes that can be analyzed in many many ways. And what I'm going to try to show you here is that the same genotypes have been coded and analyzed in different ways in the different studies, so let me try to work you through this process.
What we did is that we lined up the 13 studies, so each row that you see here represents one of the studies. And we felt a priori we should divide the different genotypes into three groups. The EM group is the leftmost and that stands for extensive metabolizers, and these would be people who have zero alleles that are slow metabolizing alleles, and we have a list of the slow metabolizing alleles from the Cancer Institute. Intermediate metabolizers would be those who have only one allele being slow and the others being okay. And slow metabolizers would be
00056
people who have genotypes who have both alleles being slow.
And essentially what we see there is that we have field cells. The field cells correspond to the alleles that relates to the genotypes that were analyzed in the individual studies, and the color is how the study itself coded the specific genotype. So green means that the study coded it as an extensive metabolizing genotype, the blue means that it coded it as an intermediate genotype, and red that it coded it as a slow metabolizing genotype.
Now the point of this graph is this: If you, if all the studies used exactly the same definition of what is slow, what is intermediate and what is fast, then you would see the same color in all columns. However, what you can see is for some of the genotypes, you may have different colors in the same column. This means that this particular genotype was treated in a different way in the different studies. Our groups of intermediate, extensive and slow is arbitrary. However, this does not impact on what is to be extracted from
00057

this figure. What is to be extracted from this figure is that different studies do not mean the same thing when they talk about slow metabolizing groups and fast or intermediate metabolizing groups. So there's a lot of genetic heterogeneity and a lot of different ways to fix it. These have been, these genotypes have been analyzed.

Description of the studies. So of the 13 studies, 11 were in women with known metastatic breast cancer. Ages of interest to Medicare, so most of the studies have mean or median ages of over 60 and five of the studies had mean or median ages over 65. There was great variability in the disease stage distribution, the proportion of lymph node involvement, the estrogen receptor status and so on, so there's a great variability in the type of patients that were included in these studies. And as you can see, there's also variability in what importance ethnic descent, if I may use this term, so European or Caucasian populations, Asian, and only one study with predominantly African-American populations.

00058

In most of the studies outcome assessment was retrospective, in five it was prospective, in eight it was retrospective.

Two of these studies were repurposed RCTs, and by that I mean that were tamoxifen versus no tamoxifen, and the investigators went back and genotyped the women who were in these RCTs, so essentially they used the RCT to get the clinical outcomes and they're using the genotype information to get information about whether the genotype would predict differential response to treatment. I think those examples with both treated and untreated people if you were to get the opportunity to, in principle at least, to assess for interactions of pharmacogenetic interactions between treatment and genetic status.

Sample sizes ranged widely, the smallest study of 21 people and the biggest study was 667. The median follow-up was between 20 and 150 months, and it was quite long in most studies.

So for the outcome of mortality, which is one of our most important outcomes, essentially five out of 13 studies had

00059

information on mortality, and none demonstrated a significant, a statistically significant

relationship between CYP2D6 defined status and the outcome. Two of the five studies, which were the repurposed RCTs I spoke of before, did not report any interaction tests, so although we had the opportunity to see whether the, whether any predictive ability of the gene would be limited to those who got the treatment, but they just did not do this test, so we missed that opportunity.

The other area was recurrence, and recurrence was assessed in 11 out of 13 studies, and most studies did not report statistically significant relationships between the CYP2D6 status and recurrence in any analysis. Four studies reported significant associations between slower and extensive metabolizer status, and increased odds for recurrence or shorter time to recurrence, so essentially the association was in the direction that perhaps would be expected by theory, if you have slow metabolizing status you have less activation, so you have increased odds for recurrence because perhaps the drug is

00060

not in its active levels. And the comparative studies did not analyze genotype-by-treatment interactions.

You know that I don't list any meta-analysis results because of the heterogeneity of the definitions of what is slow, what is intermediate and what is extensive metabolizer in these studies.

So the overall conclusions, and actually these were the only two outcomes that were assessed in the CYP2D6 studies. The overall conclusion is that there is inconsistent association between CYP2D6 status and outcomes. Studies differ in the direction and statistical significance of findings. It's unclear whether the CYP2D6 status can predict differential response to treatment in the adjuvant setting. We had a single study among the 13 that was in the metastatic setting, so 12 studies were in the adjuvant setting, one was in the metastatic.

I did not give you this information up front, but this one study was on 16 patients, it was a study from Korea, it was a very very small study and it just does not have any

00061

information because of the small number of patients, so very limited in the metastatic setting, and in the adjuvant setting it's unclear. And these conclusions are in

agreement with the ASCO 2009 practice guideline update.
So, the next story is KRAS and anti-EGFR antibodies for colorectal cancer.
DR. GOODMAN: Dr. Trikalinos, just as a time check, and you're doing fine, I think you're not quite halfway done with your slides, and I think you've got until about 10:10, and it's currently 9:35.
DR. TRIKALINOS: I may have one or two less slides at the end, so I will go on.
DR. GOODMAN: Well enough. Thank you.
DR. TRIKALINOS: So anti-EGFR antibodies for colorectal cancer, the anti-EGFR is a receptor, and this receptor controls by means of several pathways. Cellular functions, like the ones that we see at the end, proliferation, survival, angiogenesis and metastasis have been shown to have a role in colorectal cancer pathogenesis and pathophysiology. So in essence, the anti-EGFR

00062

people find the external domain of this receptor and thereby they stop its effects, and this is the way that these drugs work.
However, for example, as you can perhaps see, is one of the, causes a process that is one of the ones controlled by the EGFR receptor, and when you have mutations in this factor they may get it to be constituted on, so essentially irrespective of what is bound on the receptor, this thing activates the downstream pathway, these mutations can activate a downstream pathway, so it can essentially, these mutations can essentially abrogate the effect of the drug.
Again, these slides about the literature flow, we started from some citations and we ended up with 31 studies. And the description of the studies is that 26 of them included patients who were pretreated with cytotoxic chemotherapy, so essentially it's not in naive patients. 29 of 31 studies are in the metastatic setting and two studies are in the neoadjuvant setting, and I'm not going to give you any results about those two neoadjuvant studies because of time limitations. Mean or

00063

median age was at least 60 in 22 out of the 28, and racial composition was not reported.
However, what we can say is that most of them were conducted in Europe, six were multinational, and none were exclusively conducted in the States.

This is a brief description of all the studies, so I don't expect you to look at the table. The only thing that I'm showing you is the total of 29 which are in the metastatic setting. And most of them are on cetuximab, a few on panitumumab, and there are a few studies of both.

So, only three of the studies explicitly stated that the sample collection and KRAS testing was a prespecified aim. Five presented analyses based on RCTs and evaluated treatment by genetic status interruptions. Sample sizes varied widely and in RCT-based analyses approximately 2,000 patients were analyzed, a thousand in the anti-EGFR antibody arms and another thousand in the comparator arms. Median follow-up was ranging from one to two years approximately.

Now this is a graph that shows you

00064

overlapping studies. This is what we call a nondirected graph. You see that each publication is denoted by an ellipse, and there are edges that connect some of the ellipses. And the edges imply, the edges stand for publications where we can trace that there are common patients. So whenever publications came from the same centers, or there may be publications of multicenter results where some of the patients are included in more than one of them, and you can see how they are connected. So essentially for a meta-analysis, you want only one of the publications that are in these clusters to avoid duplication of the same information time and again.

Now this is a minimum of the overlap, and we are suspicious that there is more overlap there but we cannot really trace it, how to detect the overlap, and they are a bit more stringent than just seeing the same author or seeing the same census.

So, mortality was assessed in 19 of the 29 studies and in all of them, all 19 studies, the information would essentially comport, and the information was in the

00065

direction that one would expect, that people who have the KRAS mutation have worse response to treatment with anti-EGFR antibodies. So 18 of these studies had time-to-event analysis, had survival analysis, and as you see, KRAS positive patients had shorter median survival compared to wild-type patient. In nine of these 18 studies the results were statistically

significant themselves, so it's not that you would need to meta-analyze them to get to a significant result. The findings of the 19th study did not have time-to-event analysis but the ratio analysis was in the same direction. So essentially it was reported in all the studies in the direction that we expected. Disease progression was evaluated in 26 out of 29 studies, and all 26 studies reported progression-free survival or time-to-progression analysis. The median progression-free survival or time to progression was shorter among patients with KRAS positive tumors as compared to wild-type patients, that is patients without the mutations in their tumors, and this difference was statistically significant in 16 of these

00066

studies. Again, for the treatments for mortality and progression, for the outcomes story for mortality and progression we did not do a meta-analysis because of heterogeneity in the populations and heterogeneity in the outcome definitions, but the view is that everything is concordant and quite a few of the studies are significant on their own. However, for treatment failure by imaging, we did a meta-analysis, and essentially here we see that failure rates are higher in patients with KRAS mutations rather than wild-type patients, again the same direction, the same view. In studies of patients who had received prior chemotherapy, the response rates were typically very low, often zero, in the presence of KRAS mutations. And we have 15 known overlapping studies and we did a meta-analysis of these 15 studies. I'm showing you only the meta-analysis that is on the ROC space, and what you see here is these faint gray circles that essentially stand for the different studies, and the area of each circle is proportional to the weight that each

00067

study gets in the meta-analysis. If you have a big circle you have a bigger study generally speaking, or a study with more events, and these studies get more weight. Now this type of meta-analysis tries to synthesize two quantities at the same time, both the sensitivity and the specificity, or one-minus specificity. And the summary is given by this point, this square, the black square, and this is the summary estimate for

the diagnostic ability of all these studies together. And this dashed line is what we call a confidence region or a confidence helix, and this is actually the envelope of the 95 percent confidence in the bivariate case, for both outcomes together, for both sensitivity and specificity. So you see a sensitivity of .52 and a specificity of .93 in these 15 studies. Now, we did several subgroup analyses and we did several explorations of this heterogeneity, and I'm not going to show all the analyses to you. The only thing that I'm showing is the only thing that seems to stand out somehow, and this is that we have two studies in people who were not being treated

00068

with chemotherapy, so it was people who had not been heavily treated, and these seemed to have lower diagnostic ability compared to the other studies and they have to relate somehow to the subgroup analysis. And there is some potential explanation for this and the potential explanation is perhaps that in these people who are treatment naive, the chemotherapy itself has something to offer, so the pharmacogenomic effect is drowned, the effect of the anti-EGFR itself.

Repurposed RCTs were found, and one study assessed overall survival and treatment by KRAS mutation. And essentially we have a consistent view, we have an association, results in the expected direction, it was significant in one study, not significant in the other, but in the expected direction.

As for the other outcome of progression-free survival, we essentially have four studies and the results are statistically significant in reality in three. Although the fourth study is not, it's in the correct direction but not significant in the fourth, so again, it's a consistent view.

00069

So the conclusions for the KRAS example is that for all assessed outcomes, patients with KRAS mutations were less likely to express treatment benefit compared to wild-type patients. There is the same direction of effect in all studies and for all outcomes. And for most studies we had significant overall survival and progression-free survival. Significant treatment by KRAS mutation interactions were identified in the RCT-based analyses which are, let's say some sort of better type of studies

for several reasons. And all these results are in accordance with the guidance that was provided recently by ASCO, the FDA and the EMeA, which is the European Agency for Medicine.
Most of the studies were in the second line setting. There seems to be lower predictive ability of KRAS mutations in the first line setting, these were the studies I showed you with the two arrows, and perhaps we cannot say a lot about the first line setting or we cannot say with a lot of certainty. And there seems to be that there is similar

00070

predictability for cetuximab and panitumumab, and this is based mostly on studies of panitumumab monotherapy in pretreated patients, so we don't have a lot of evidence for, or any evidence in studies with non-pretreated patients.
And this is the third topic, BCR-ABL mutations and tyrosine kinase inhibitors for CML. CML is a leukemia, a malignancy, and so essentially the physiology of this disease is the formation of a protein, and this protein acts as a tyrosine kinase. And this is an enzyme, an enzymatic activity that leads to the pathophysiology of the disease. And there are some mutations in these genes, and there are some drugs that we call tyrosine kinase inhibitors as well as other drugs, and these drugs bind to a specific site of this enzyme and they stop it.
These are wonder drugs that have reversed essentially the clinical phenotype of CML and they have resulted in this revolution in CML treatment. There are some mutations in the binding protein that affect the affinity to bind, the binding affinity of the drug, so

00071

essentially they abrogate the effects of the drug. And there are several of them, and one that is very very well known is the T315I mutation, and this is a rare mutation that is known to completely abrogate the effect of CML when it is present.
And here we summarize the results. We have, again, the literature, 31 studies included in this topic. Overlapping studies, we had a hard time trying to distinguish nonoverlapping populations, and it seemed that especially for CML, many of the patients came from the same team time and again, many of the reports, and what we do here is we separate

first line therapy from second line therapy and third line therapy.
So first line therapy is patients who have not been exposed to any treatment before, and these are usually treated with monotherapy early in their course generally speaking.
Second line therapy is people who have already failed imatinib treatment, and there are several options; there are high dose imatinib regimens, or combinations that can be given, and you see them organized there. And the

00072

third line therapy are people who have failed both first and second line therapy, so it's some sort of very very unlucky people.
And what you see here is, again, that there's a lot of overlap in the various publications and it's not very easy to distinguish. Sometimes we have to go to the appendix and see the actual patients that are in the samples of the publications, and there it's shown that they come from a previous study, and you can sort of corroborate that there's an overlap. And what we know is that there are further overlaps that we cannot show here, and we are very very suspicious that, for example, most if not all of the patients in the third line study have been included in the previous studies as well, but it's not very easy to show. So there's a lot of overlapping information, and this is perhaps not unusual for observational studies in general.
So, all of these studies reported essentially the same area of the gene, so they have a lot of mutations and report information on the T315I mutation, which is the one that we said this is very known. And this shows you

00073

that essentially mutations are not very common, they're rare. You see from the shading that most of the studies identified these mutations in small proportions of the patients. So it's kind of a different beast compared to let's say CYP2D6, which was common genetic variations.
This is just the overview.
Essentially there's nothing, there's very few data on mortality and progression, and most of the data have to do with treatment failure by means of blood criteria, hematologic response, cytogenic response or molecular response.
There are different criteria that the studies have come up with and followed to judge the response rates, and most of them are in dasatinib for a second line treatment rather

than the other drugs, so potentially we have an evidence region that is pieced together from many different settings, so we have different lines of therapy and different drugs, and often different outcomes, and so it's a very very nonhomogeneous set of studies.

For this reason we could not do a meta-analysis. All these are in different settings, patients were in all phases of the

00074

disease, chronic phase, blastic phase, accelerated phase. Mean ages ranged between 50 and 62, so they are kind of, not in the 70s let's say, there are no studies perhaps in a bit older patients, at least to my knowledge. Intermediate follow-up between two and 61 months. Maximum sample size not that big, 670 patients. Small sample sizes, I remember. Clinical outcomes. Essentially we said there's no data on clinical outcomes, so essentially associations with overall survival or progression-free survival were reported only in one study with first line therapy. This was a study by Willis done in 2005, and they found no significant association, so any mutation with overall or disease-specific survival. Now this is a bit of a hindrance; they are basically assessing any mutation and not specific mutations, so they put everything together in a single packet and saying whether there is any mutation associated with the outcome.

And now I'm going to show you just one of the outcomes, cytogenetic response, and this is for any mutation, put all the mutations

00075

together as if they were equally devastating, let's say, and trying to see whether they predict lack of cytogenetic response. And you see that most of the data are on the second plot, and there's sparse data on the other plots, the different types of markers and the different stages of disease, but we don't really care. What you have to know is that all of the studies follow the data, so when it comes to using any mutation, there's no predictive ability essentially.

There are some studies that are very small and essentially have zero counts in the two-by-two table that we use to calculate sensitivity and specificity. And this is why for example in the imatinib-based, you see one study by Jabbour, a 2009 study, and it appears to have perfect sensitivity. However, this is

a very small study and we can't really believe that it has a sensitivity of a hundred percent. This all has to do with the variety of the mutation and the sample size being very small. So essentially for any mutation, no good predictability, we don't really have a meta-analysis and we don't want to do a

00076

meta-analysis here because we have patients who are very heterogeneous. However, we find what we expected to find for the 315 mutation, which is the one that is known to abrogate the effects of the drug. And essentially we see that all the studies are squished down to the left on the vertical axis, and what this says is this mutation has a specificity of 100 percent to identify lack of response, and this is what we know from the theory already or what we know from basic science and what is known in the field. And essentially it was very low sensitivity to identify non-responders, sensitivity, the ability to maximize true positive tests, so the ability to maximize the number of non-responders that are correctly identified by the presence of the mutation, and this number is so low because the mutation is rare. If you remember from the previous table that had the prevalences, it was in the five, seven percent range, and very often lower than that across the different studies. So the presence of any BCR-ABL1 mutation does not appear to predict

00077

differential response to treatment with TKI inhibitors. Consistent evidence that presence of the rare T315I mutation can predict TKI treatment failure that is not a hematological or cytogenetic response. I only showed you a figure for cytogenetic response, lack of cytogenetic response, but there are similar figures for hematologic response and molecular response, and the same matter exists there. So because of the complexity of this issue, it's our assessment that analysis on collaborative registries of CML patients are necessary, because there is simply no way that you can actually use the published data and disentangle all the different factors from the actual predictive effects. You cannot use published studies to predict the effects of the different mutations on treatment response. Most evidence pertains to short-term surrogate outcomes of hematologic, cytogenetic

and molecular response since we don't have evidence on clinical outcomes like mortality, like progression-free survival. Most evidence is of second line TKI treatments like dasatinib and nilotinib and from a relatively small

00078

number of referral cancer centers. As I said before, all these studies seem to originate from the same centers and therefore, there is unclear generalizability of these findings. For example, we don't know what is the actual prevalence of these rare mutations in quote-unquote garden variety CML patients throughout the world.

So this concludes the presentation about the evidence in those three topics and now I'm going to discuss some cross-cutting methodological issues, and then I will conclude.

So, one thing that was apparent throughout these systematic reviews was that treatment-by-gene interactions were often not assessed and to some extent this may be justifiable. If it's completely known that this gene essentially predicts, whatever the gene predicts when it comes to the treatment response is only treated patients, and that the gene would not predict anything in an untreated patient. However, it would be nice to show it, it would be nice to show that the ability of the gene to predict, or of the genetic

00079

variation, I'm sorry, to predict the treatment response is only among treated patients and it's not only a prognostic ability of that genetic variation. The difference between those two is that you can have a genetic variation that predicts time to death, let's say, for all patients irrespective of whether they're treated or not, and I examined the genetic variation only among those who have been treated, then I may think that this is a pharmacogenetic effect but in reality it's just a prognostic effect. So although several studies had the opportunity to do that because they had both treated and untreated patients, they just did not do it.

Sample sizes are small and when you have small sample sizes, you have diminished power to detect small effects. We know that we don't actually know how big the pharmacogenetic effects are expected to be, we don't have a lot of empirical data on how big pharmacogenetic effects are in general. What we know, though,

is that prognostic effects of genes for associations with common diseases are small, so we're not talking about huge effects. And if

00080

one thinks that the pharmacogenetic interactions are likely in terms of magnitude and strength, it would mean that you would need a lot of people and big sample sizes to detect subtle and small pharmacogenetic interactions. The other thing is that, and this is a general observation, is that repurposed RCTs, what we call repurposed RCTs, which is the ability to take an RCT and genotype the participants and retrospectively go in and see whether there is a pharmacogenetic effect. This seems to be a very neat way to perform studies with pharmacogenetic effects, because you are essentially using the whole RCT machinery and you have a very good adjudication of outcomes and you have a very good follow-up of the patients, and for many cases the samples have been collected from the patients while the RCT is going on. The actual analytic validity of tests to detect genetic variations in these samples are very good, so they are essentially as good as if they were done prospectively and during the time of the RCT conduct. Now this is important for variations that are somatic. There are two types of

00081

variations, there's somatic variations and germline variations. Germline variations, genetic variations are heritable, and these are variations that are stable and do not change throughout our lives. And this is, for example, variations from CYP2D6, so the CYP2D6 case was a germline case and these can be assessed at any time, even after the RCT has been conducted and concluded, before or after, it doesn't matter, they will never change. So, a problem might be with the somatic ones like the KRAS mutations and the BCR case, which are more volatile, and they may evolve during the course of the treatment, and that's why it's important to have samples during the RCT conduct. However, if all stars align, then it's not, you cannot detect a huge bias that is introduced by using this type of repurposed RCT. So repurposed RCTs cannot measure the effects of testing on patient outcomes or the effects of testing on treatment decisions, because they are essentially retrospective exercises. And this is, the second bullet says

what I said before, that genetic analysis from
00082
archived but prospectively collected samples is generally accurate. The catch is that there are a lot of pharmacogenetic tests that can be examined and more will pop up, so there's an opportunity for that. When you do a lot of tests in the same population and you examine the same thing in the same population, you run into a case of multiple comparisons, and this multiplicity of comparisons has to be taken into account in your statistical analysis. Otherwise you will find spurious results. And in general when a result is found in a genetic study, or in a pharmacogenetic association study is better, it should better be evaluated in an independent population and in an independent study to control for the danger and the rate of false positive findings. As we said, there is heterogeneity in genetic exposures, this was particularly evident in the KRAS studies and the BCR case, and again, when you have a lot of alleles, you have a lot of opportunities to group them together and to analyze them in the way that you would like, and you can actually get a statistically significant result from analysis
00083
if you play enough with it. I dare say that we have an example in mind. There's a particular study that essentially analyzed a lot of genes and it has a lot of SNPs in the CYP2D6 case, and it can be looked at in ways that are not immediately obvious. We cannot see a logical pattern behind the proof of these variables, so in their main analysis they may find nothing, but then they go on with this exercise of looking at them differentially and they find a margin that is a statistically significant result. This is in my mind a demonstration of data dredging. So the heterogeneity with genetic exposures cannot be really tackled with meta-analysis of root data, so it's perhaps important to go on and have meta-analysis from individual patient data, and this is actually something that has been done in other cases or other genetic tests, like in the warfarin example. There are several statistical issues. Adjustments for multiple comparisons were not documented in the included studies, so we have
00084

a large number of possible hypotheses, and again, this entire issue of multiplicity of comparisons, and statistical significance findings are not even at the five percent level, they are actually much worse than that. The other thing that is particularly pertinent to germline mutations and germline variations, that is variations that we get from our parents and which don't change through our lifetime, is that adjustments for potential confounding factors are too confusing, or are at least debatable. So let's see what we could do, or why, what some cases are where you should not be adjusting for confounding. And you would not be adjusting for confounding if you have a factor that is in the causal path, and this is because if I have the genetic exposure and I have a confounder or a third variable that is in the path, it's influenced by the genetic exposure that might affect the outcome, I should not be doing naive adjustments or, in that case, because it results in essentially conditioning complex ways, and masking the actual effect of the exposure on the outcome. I could do more

00085

complex things, there are structural equations or other approaches that are proper, but simple adjustments are just not given to the story. You would adjust if you had confounders, and what are confounders? Confounders are causes of the outcome that are also associated with the exposure but are not affected by the exposure. This is a mind-boggling thing to provide an understanding of, what confounders are. However, as you can see in the causal diagram, confounders would be affecting the exposure and the outcome, and they may induce an association, they may make an association appear that would disappear if you took the levels of the confounder into account.

Now, the thing is that when you're assessing germline mutations or germline variations, you cannot have this previous relationship where a different confounder was affecting the exposure. And this is because the exposure, our genotypes are protected by what is known as Mendelian randomization. And genotypes are essentially randomized to a meiosis, to information during the formation of

00086

the human being. So essentially they cannot be confounded by something else, they cannot, for

example, smoking cannot affect which genetic variations you have because you can only, there's temporal comparisons here. You can stop smoking when you're already an adult, but your genetic variations have already been laid out during meiosis.

This is my last slide. So essentially for germline variation, adjustments are probably not warranted, for two reasons. The third factor is, if a confounding factor is affected by your genetic makeup, it's in the path so it should not be adjusted for, at least not in naive ways. And if it's not in the path, it's not in the causal path, it cannot be a confounder because of Mendelian randomization.

And the final slide is that multiple studies on each topic frequently originated from a limited number of specialized centers and identifying non-overlapping populations becomes, or can become a problematic issue. And also, this poses a threat to the generalizability of findings, this is something

00087

that we get from the CML example.

So this is where I conclude.

DR. GOODMAN: Thank you very much, Dr. Trikalinos. If I'm not mistaken, you and your team will be here for the balance of the day; is that correct?

DR. TRIKALINOS: Correct.

DR. GOODMAN: Thank you. Panel, it's time for our break and I think we all need it. We've noticed that you've been taking a lot of notes and we have a lot of questions. During the break, Dr. Satya-Murti, I just want to check with our panel and see whether or not you want to stick with the agenda as is, which would have us, following the break, go directly to our scheduled public comments and then to our open public comments, or whether you want to shift the agenda just a bit in case you have some immediate questions for our morning presenters, which might help us focus. So let's just talk briefly about that while we break and we will do your bidding as such, and we will confer with the CMS staff about that.

I want to thank very much, Drs.

00088

Freedman and Trikalinos and team members, for superb presentations. I can't promise we got all that, but we certainly appreciate you answering some questions and we will certainly

have some for you. Whatever your watch says, add 15 minutes to it and we will start again. Thank you.
(Recess.)
DR. GOODMAN: We're going to reconvene right now. If our panelists would have a seat, we'll reconvene. The panel has many questions already for our first two speakers. However, we're going to try to impose some self discipline and push through the agenda as is with our scheduled speakers, scheduled public comments, open public comments, which will force our panel, including me, to set some priorities and ask them in an organized way. We have nine scheduled public comments, each of which has five minutes, and not five minutes and one second, but five minutes for your total presentation. I will give you a one or two-minute warning if it looks like you might need that.
First up is Dr. Diane
00089
Allingham-Hawkins. Dr. Allingham-Hawkins. And as Dr. Allingham-Hawkins makes her way to the podium, just a little reminder to our panel yet again that we're seeing a lot of material today. Please do keep in mind what the questions are that we need to answer this afternoon, and as we discussed at the break, those questions really deal with health care outcomes and adequacy of accompanying evidence, health care outcomes and adequacy of accompanying evidence.
Dr. Allingham-Hawkins.
DR. ALLINGHAM-HAWKINS: Good morning. My name is Diane Allingham-Hawkins and I am a molecular geneticist and a cytogeneticist. I am here representing Hayes, Inc., which is an independent health care research and consulting company located in Lansdale, Pennsylvania. Hayes does not, nor do I personally, have any financial involvement with the manufacturers of any products being discussed, and my travel to this meeting was funded entirely by Hayes. For more than 20 years, Hayes has been an industry leader in providing health technology assessment on a wide variety of new,
00090
emerging and controversial health technologies to our worldwide clients, which include hospitals and health care systems, managed care organizations, government agencies and employers.
As we have heard a number of times

now, pharmacogenetics is a study of how an individual's genetic makeup influences their response to a drug, and pharmacogenetics is a cornerstone of personalized medicine, which is a form of medicine that uses information from a patient's genetic makeup together with information about environmental exposures to tailor their care in order to prevent, diagnose and treat disease.

So what evidence is necessary to evaluate pharmacogenetic tests? Evidence must address the analytical validity of this test which is the ability to accurately detect the change of interest, clinical validity which is the ability of the test to detect your clinical outcome of interest, and clinical utility which is the impact of the genetic test on patient care. Ethical, legal and social implications, which are safeguards and impediments of the

00091

test, must also be considered in the context of the other elements.

Other considerations, although not as critical as those from the previous slide, include the cost of the test, does it make any financial sense? How does the test impact current clinical practice, does the use of the test make a difference in how a particular clinical situation is approached? Quality of life and patient preferences, in some cases it's the use of a given test that may be preferable over another. And the future of the technology, is this technology, while perhaps not yet viable, likely to make an impact in the future?

What kind of studies are we looking for? Ideally we would like to see large prospective randomized controlled trials that clearly show the clinical utility of the test, but the reality is that such studies are few and far between. Having said that, most case studies are retrospective in nature and relatively small. But having said that, the existence of a number of smaller studies with consistent outcomes may be sufficient to make a

00092

determination on a given test.

To demonstrate the uses and outcomes of pharmacogenetic testing, I would like to talk about the example of KRAS. Sequence variants in the KRAS gene have been linked to treatment response in a number of cancers including metastatic colorectal cancer and non-small cell lung cancer.

In evaluating the evidence related to variances in response we've seen to treatments with monoclonal antibodies in metastatic colorectal cancer, we found that there are no large prospective trials, and as we heard examples earlier, there is sufficient consistent evidence from smaller studies that KRAS status impacts response to therapy. For non-small cell lung cancer, however, the current evidence is less compelling that KRAS status does impact treatment response to tyrosine kinase inhibitors. It is clear, therefore, that pharmacogenetic tests, even those involving the same gene, must be evaluated on an individual basis, to insure sufficient evidence exists to

00093

support the use of the test for that application. Hayes has evaluated the evidence associated with 20 different pharmacogenetic tests to date and we found sufficient evidence to support the use of just five of these tests. So the remaining tests, which includes two of the five under review at this meeting, while promising in some but not all cases, are not yet proven to improve patient care.

DR. GOODMAN: About one minute.

DR. ALLINGHAM-HAWKINS: The conclusion, then, while pharmacogenetics has the potential to revolutionize drug therapy by ensuring that the right patient receives the right drug at the right dose at the right time, evidence is currently lacking for the majority of pharmacogenetic tests currently available. Manufacturers must be encouraged to perform sufficiently powered prospective studies that unequivocally demonstrate the benefits and risks of these tests, and results of the studies must be evaluated by independent entities. Ongoing evaluation of the evidence is essential to the development of meaningful coverage policy.

00094

With that I will conclude my comments and thank you.

DR. GOODMAN: Thank you very much, Dr. Allingham-Hawkins, very helpful. As Dr. Mitchell Burken is making his way to the podium, just a kind suggestion to all of our speakers today. If you are thinking about talking about something you think the panel has already heard, you don't need to repeat it. If it's a main point or some issues that we

haven't heard yet that are directly germane, those are the ones we'd like to hear, so that might help you make your short presentation even more efficient. Dr. Mitchell Burken, medical director of IntegriGuard. Sir.

DR. BURKEN: Just a correction for the record. When these slides were sent to CMS I was an employee of IntegriGuard. At this point I will just be representing myself and not the company, so this disclaimer statement is really not pertinent.

What are the relevant questions?

Well, the general topic of BCR-ABL and imatinib may be construed to include two separate issues in their corresponding sets of questions. One

00095

of those issues would be the issue of BCR-ABL monitoring during therapy. The other issue was one that was touched upon very nicely previous to the break on mutations.

I'm going to discuss mutations in a slightly different context than Dr. Trikalinos and I'm going to use it to make a greater, more global point about how we as payers tend to look at new technologies and how we tend not to look at new technologies, and this diagram here shows that there's a pyramid starting out with test validation leading up to clinical utility.

The important point here is that, and the reason I drew the pyramid this way is I wanted to make it look like an iceberg, because there's a large component called test validation that we as payers really don't see very often, which really involves the mathematics and the biomathematics of internal validity. And I'm not going to go into detail on these internal validation techniques, but again, it's something that we tend to lose sight of when we're thinking as payers today, when we're making these types of coverage decisions.

00096

So what I did is posed the question, well, if you wanted to design a panel of mutation markers to test, in the last panel, again, Dr. Trikalinos talked a little bit about T315I and its role in the mutations, but, you know, how might we construct a panel. And again, I just did a quick PubMed search, started with 33, caught over 3,300 references, but basically boiled it down to only four abstractions that warranted a full article retrieval. I did also a supplemental Google search.

Again, this is just a capsule summary of the four PubMed studies. You will note in the Branford study is where it's fewer than ten common mutants account for 60 to 85 percent of all mutations. So the question becomes, and if we go back to our pyramid, you know, are these articles helpful in test validation, clinical validity and clinical utility, the answer is no.

But let me actually go back to this slide and point out that this paradigm is something I adapted from the Center for Medical Technology Policy Effectiveness guidance

00097

document that's listed here, and I'm just going to leave this up while I exit the podium, because I think that's a very very interesting and compelling document that really helps to organize the thinking on all the phases of validation and utility. So thank you.

DR. GOODMAN: Thank you very much, Dr. Burken. Next is Jeff Voigt, principal, Medical Device Consultants of Ridgewood. Mr. Voigt.

MR. VOIGT: Thank you. My name is Jeff Voigt, I'm an independent research or reimbursement consultant. Due to the five-minute limitation I'm not going to talk about suggested solutions to the issue being presented. However, the handout that's been provided which is entitled Examining the Evidence For Clinical Utility and Testing does. I'm going to talk a bit today about the 800-pound gorilla in the room, which is the definition of clinical utility, which is rather troubling to me and some of my clients. I have no financial ties to making this presentation, I'm here on my own.

I and companies I work with have

00098

recently experienced frustration in dealing with CMS policy developed from inputs provided at the February 2009 MedCAC meeting. I believe the policy is flawed and will ultimately hurt the development of and access to clinically useful genomic tests. This policy was developed based on a simple query administered to the MedCAC group at the February 2009 meeting. The question asked that the panel relate it to the best type of evidence required to support a finding of improved patient-centered health outcomes based on the results of a diagnostic genetic test. In this question the answer was provided; the best type of

evidence needed to infer that the diagnostic test improved health outcomes is, surprise, improved health outcomes. There are numerous issues that were not addressed in that query, nor appear to have been considered in the CMS policy as it was developed, including the practicality, cost, timing, ethicality, and patient access to important medical advances. Many of these same issues were brought up in public comments and were reflected in the February 2009 meeting

00099

transcript. MedCAC in its 2006 recommendation made by its operations and methodology committee in establishing guidelines for evaluating diagnostic tests stated the following: The recommended approach for evaluating diagnostic tests when direct evidence is not available is to determine the extent to which there are changes in patient management, particularly when the management strategy has been demonstrated to be effective, such as improvements with established associations. In this case, intermediate health outcomes may also be considered. Why CMS and MedCAC have not considered pharmacogenetic tests is troubling, especially since it's been used by CMS for evaluating the clinical utility of other diagnostic tests. The MedCAC's survey back in February 2009 also appeared not to include the inputs from important constituencies such as the companies that actually develop these tests, statisticians who understand the nuances and issues surrounding the evidence gathering, or patients who may have actual benefit from these

00100

tests, all with vested interests in seeing that these tests and technologies are accessible and clinically useful, and likely have some unique experience and insights into the practicality of proving out clinical utility. What are the issues surrounding CMS's definition of clinical utility equating to patient-centered outcomes? First, the definition will undoubtedly be picked up by private payers and used as their definition for clinical utility. Second, being able to establish the direct effect of clinical test results on health outcomes is extremely challenging, sometimes unfeasible. The impact of a diagnostic genetic test on health outcomes is

very often confounded by the variable effects of such things as physician behavior and decision-making, treatments or interventions employed, patient adherence to treatment regimens or other patient behaviors which occur following the diagnostic test. In other words, it takes a leap of faith to conclude that the results of a diagnostic test had an, or any effect on the outcome.

00101

Third, the financial ramifications of having to establish clinical outcomes for payer coverage can be enormous, costing tens of millions of dollars. These essentially become drug-like trials.

Fourth --

DR. GOODMAN: One minute, sir.

MR. VOIGT: Okay. Fourth, establishing a direct effect of genetics test result on a health outcome presents enormous problems for IRB approval. In order to demonstrate the clinical benefit of a new diagnostic test over an existing one seldom can be randomized to a treatment, therapy or intervention that matches the gene expression in the new test, and there would be some randomized treatments known to be ineffective based on results of the inferior test results. If clinical outcomes as defined above by CMS becomes a requirement for establishing a positive coverage determination, it will reduce investment in new genetic tests and the market introduction of these tests, and ultimately their use. This in turn will have an adverse effect on the quality, access, and potentially

00102

the overall cost for care. If there are others in the audience who have similar concerns, it is respectfully requested that they also voice their opinion and please read the entitled Examining the Evidence For Clinical Utility and Testing.

DR. GOODMAN: Thank you very much, Mr. Voigt. Sorry, but your time is up. We appreciate your input. You may leave the podium now.

I just remind the panel as I think all of you know, MedCAC is not here about policy, we're not a policy-making body, we're an advisory body. Policy in some instances is made by CMS, not by us, contrary to what you might have heard. Thank you, sir. Our next speaker is J. Russell Teagarden, clinical practices and therapeutics, Medco Health

Solutions, Inc. Sir.
MR. TEAGARDEN: Thank you. I am from Medco, and I don't have anything to disclose other than I'm from Medco and we have some commercial programming around testing and so forth embedded in it. I have a more robust set of slides here than I will be able to get to, I

00103

have a beginning, middle and end, and I will stick with the middle for the most part.
Here, Medco is a PBM, and it's big. One in five Americans, their pharmacy benefit is managed by us in some way. And just to give you a sense, why we're interested in this question is because we advise payers of pharmacy benefits on their plan designs, we implement various utilization programming for them, and we, at the size we are, whether we like it or not, and we do like it, we're in the public health system and so we have an interest more broadly in the safe and effective use of drugs, therefore our interest in anything that can make us more effective and gives us better precision to do that.
I want to focus mostly here on what's going on in the private sector that addresses the question the committee is being asked, mainly about what level of confidence should you have in evidence and how you should assess it and so forth. And I'm here to tell you that there is already some of those assessments going on in the private market in our domain. And for example, there are, several of these

00104

tests that you are looking at are already embedded in coverage policies for drugs. In other words, coverage for drugs are contingent on some of these tests already, that's quite common in the private sector.
Furthermore, there are many plan sponsors signing up for some commercial program around it, and I want to give you an example of what I mean by commercial programming. I'm going to zero in on the tamoxifen program that we make available for our clients now. You've heard about the issue with tamoxifen and from our own data we do collect information on testing, and we see that something like 20 percent of our patient population are maybe at risk for less than effective outcomes with tamoxifen.
So what we do is, we know who's on tamoxifen in our universe, and with those payers that are interested in doing this, we

will contact the patients who are on tamoxifen, we'll -- I'm sorry -- we contact their physician first, describe the situation, and ask if they're interested in ordering that test. If so we talk to the patients, the

00105

patients go forth, and we facilitate the testing with our partner labs, they get a couple swabs, they swab themselves, they send the samples to the lab, the lab reports the rules to their doctor and to Medco. And then if we see something at Medco in our therapeutic resource centers that indicates that further elaboration is needed, such as a poor metabolizer, or make sure the doctor knows this and what alternative is available. Or is he an extensive metabolizer and we know that the patient is on a CYP2D6 inhibitor, we can further elaborate on that and help them get to a better therapy regimen with that in mind. Currently we have over 200 clients in these programs, they represent over seven million covered lives, and from what our account management people tell me, the uptake on these kinds of programs has been faster and more expensive than anything we've ever done at Medco, so there is a lot of interest in this in the commercial market. So what might look like a particular case, you see a prescription, and we get some lab results for metabolizers and we will follow

00106

up, and then we can see drug therapy changes as appropriate. This might be what a typical case would kind of look like. Now what we do with that information too, we can leverage it, we get this phenotype in, it may be perfect for other drugs. So just like if a patient has an allergy, we're able to notify a pharmacist, physician. So in that case when a drug comes in, here's the same thing, we tie this to other drugs where 2D6 phenotypes are relevant and we're able to tell people, pharmacists, physicians, when we see a prescription come in for another drug in which this phenotype is relevant. This is some of our early findings on uptake with physician patients. This actually goes across both our tamoxifen and warfarin programs, I don't think they're broken out. But you can see that we get hold of our physicians, about two-thirds of them say yes, let's do that, and then 82 percent or so of the patients are good for it.

I'm going to end up here by showing you some results of a survey we did with AMA where we got 10,000 surveys back from docs to
00107
give us some sort of sense of what's driving adoption or not. And we see that many physicians are of a mind that genetics, will drugs do, or we will know the drug effects, but that they don't feel particularly well prepared for it, but expect to have to be doing it sooner.
We have some other comments that we provided about evidence gaps and getting to what we need to establish some external validity on the data and the research because our role is external and we need external validity, and I will leave it at that.
DR. GOODMAN: Thank you very much, Mr. Teagarden. Next we have August Salvado, from Novartis. Again, I encourage all, if you could please focus on information that will help us address our questions, that would be great. Sir.
DR. SALVADO: I am the vice president -- I'm a hematologist-oncologist and I am the vice president for clinical development and medical affairs at Novartis Pharmacology, and I'm responsible for the hematology side. What I would like to do is I
00108
would like to thank the panel, first of all, for allowing me to make a few comments regarding achieving better outcomes for CML patients through molecular response monitoring.
And I need to make a clarifying statement because what I'm addressing in terms of what is before you is a very different question than what was brought up by both Dr. Trikalinos and Dr. Freedman, which has to do with mutational testing. We're looking at, or supporting genetic testing and molecular monitoring of each transcript to follow the course of the disease and to allow physicians to make better therapeutic decisions going forward.
I'm not going to go into this slide very much except to say it was already brought up by Dr. Trikalinos that the disease results from a translocation of a portion of the chromosome nine on the long arm to chromosome 22, and that results in a fusion protein. And that's the core of what we're supporting here, that fusion protein is both necessary and sufficient to produce a phenotype of the

disease, and when the disease is adequately
00109
treated, that fusion protein disappears. And when resistance occurs the disease is reactivated, either through mutational mechanisms or potentially through non-mutational mechanisms, the levels of this fusion protein, again, rise, and therefore are useful in potentially following the development of resistance in patients and helping physicians make therapeutic choices.
In 2001 the FDA approved the first TKI inhibitor, imatinib, for the treatment of this disease. Since that time I would like to point out that imatinib has two other generations, two second generation models of tyrosine kinase inhibitors, one from Novartis and one from another company, have also been approved for patients who are failing first line therapy.
These TKIs can reduce progressively the disease burden to a level that is below that that can be standardly detected and useful by standard cytogenetic tests, so a more sensitive test is really needed to monitor patients going forward.
Monitoring patients, of course, are important not only in terms of assessing their
00110
response to initial therapy but it also, as it turns out there, the kinetics of that response and the depth of that response, the durability of it, and the risk for future progression of the disease. So when patients are being followed, early identification of unsatisfactory treatment response through identifying molecular transcripts are actually very important in terms of being able to make a therapeutic decision for that patient.
Molecular monitoring is done by real time quantitative PCR, and when you apply that test it is routinely at least three times more sensitive than standard cytogenic testing on bone marrow samples. Additionally, molecular monitoring is performed on peripheral blood which is more convenient and less invasive, and the levels of those transcripts as they rise can very early detect when patients are beginning to fail treatment with standard therapy.
I'm going to skip that slide.
DR. GOODMAN: About one minute.
DR. SALVADO: And I am going to basically go through to looking at some data to
00111

show you that -- let me go back -- data from the initial study of Gleevec that was done now over eight years ago. These are results that were shown recently, that were published at that seven-year time point, showing that patients who developed a deep molecular response, notice on the left-hand column here, to a level of what is called a major molecular response, were less than or equal to one percent of their initial value by an international standard, versus those who don't develop that depth of response, have differences both in event-free survival and in transformation to blast crisis and to accelerated stage disease.
This validates that with later data. This simply says that those patients who achieved molecular responses were progressive patients who developed accelerated stage of disease.
DR. GOODMAN: Thank you very much, Mr. Salvado, we have to move on. Next is Dr. Michael Dugan, who is with Genzyme. And to all, we do appreciate your understanding of the need for us to go through these promptly. We

00112

will do our best to get to many of these issues during the Q&A period, but we appreciate your patience with us.
DR. DUGAN: My name is Michael Dugan, I am the vice president of pathology services for Genzyme Genetics, representing ACLA, American Clinical Laboratory Association. The association represents national, local, regional, commercial and hospital-based laboratories. I have been in this capacity for several years with Genzyme, we have performed several of these tests, almost all of them. And I was previously the medical director of Specialty Laboratories, which performs about 2,500 tests for hospitals across the country.
I want to just briefly speak to a few points or several of the key points were already covered. One is that the laboratory role traditionally has not been one to establish the clinical outcomes comparison of the particular tests prior to providing those tests. We as laboratory directors are primarily charged with assuring that we can develop a test which identifies a particular analyte with a high degree of accuracy and

00113

precision, that's our charge. The clinical utility determination often varies over time

and they are the subject of clinical trials often funded by NIH and other bodies and organizations that establish the clinical utility. Hence, it's not directly within our primary purpose to do that, with rare exceptions.

I think it's already been mentioned, the difference between inherited tests, I'm sorry, inherited genetic alterations and those acquired. Historically speaking, pharmacogenomic tests were just supposedly for things that were inherited genetic variations in metabolism enzymes in the liver. Thus, the CYP2D6 and a different related pathway, UGT1A1, those are traditional pharmacogenomic tests. The others relate to molecular alterations specific to the tumor, and as elaborated, there are many tests with different methodologies that are used to identify those molecules for purposes of diagnosis, prognosis, prediction of response to drug, and also, as our last speaker just spoke to, the monitoring of the response to therapy.

00114

So it's sort of like having a speedometer on a car telling you how fast you're going, but it doesn't really tell you whether or not you're going to get there. That depends on whether or not you get a flat tire along the way. There are various complexities in testing that have been largely skipped in these discussions of tests such as KRAS or BCR-ABL.

We've provided some of that documentation from other papers provided to you that are very important. BCR-ABL, for example, FISH for diagnosis, RCT-CR for monitoring, stable time for mutation, detection of the T315I1, very different methodologies, very different applications.

And finally, just speaking to the difficulty in using outcomes to establish the clinical utility of these tests, I would remind you of one example of really a pioneer in pharmacogenomic tests that not being discussed today, and that is HIV genotyping for drug resistance in retroviral patients. To measure the effectiveness of the genotyping, you go to another molecular test, the HIV viral load, to

00115

measure whether or not the patient has rising or falling viral load. But the test doesn't tell you what the ultimate outcome of the patient will be, it doesn't predict whether or

not the patient is going to get a lymphoma and die of that or not. So it's akin to, some of these tests are used to sort of measure the size of your parachute as you're falling to the ground, but they don't necessarily tell you when or if you're going to hit the ground. Thank you.

DR. GOODMAN: Thank you very much, Dr. Dugan. Thank you for those comments. Next is Dr. Bruce Quinn, from Foley Hoag.

DR. QUINN: Thanks. Bruce Quinn, Foley Hoag. I have no direct financial conflicts with this meeting. Like Dr. Goodman, my firm works with hundreds of healthcare clients but no one supported me to be here today.

At the MedCAC today we've talked about these five genetic tests with three questions, sufficient evidence, net health outcomes, and relevance to the Medicare population. I would like to talk about ways of viewing the data

00116

before answering those questions. We think of a pipeline from basic research to clinical trials to meta-analyses to practice, usually for stuff, for devices or for drugs. But there's also a similar pipeline for process. In the thought process of evidence-based medicine, we've had plenty of testimony for years about the rules for evidence review, as we saw this morning. As Dr. Trikalinos said, there are strict generic rules for the reviews, but a review is not a policy decision. There is thought capital that's tremendously interesting coming out in the last couple of years, I've got the citations here and I would be happy to e-mail anyone my talk, by Michael Rawlins, Lawrence Green and others, about how to use the matrices after the trial is done.

So focusing on diagnostics, I'm going to lead up to talking about KRAS. Diagnostics are about reducing uncertainty. What's your cholesterol? I don't know. It's 185. You've asked the question and gotten an answer. William Osler, here in Baltimore, said ask the patient, the patient will tell you his disease. One of the things about HER2/neu and

00117

KRAS, of course, the patient can't tell you, so we ask a lab test. We can say what's your blood type? The lab test says A negative. But something's missing, there's no clinical utility, there's no context. In real life you've been bleeding, your hemoglobin is eight,

you need a transfusion. What's your blood type?
For HIV, what's your T-cell count?
The answer is four. Is your current medication working? The answer's no. Do we need to change your prescription? The answer's yes.
So you get the utility by moving to an upstream question. But look what happens. We've changed on the left a question that we can't answer into the T count, or T-cell count, a question that we can answer. We've changed a question we can't answer into a question we can answer, and that means we need to know what do we need to know to bridge between the question on the left that we can't answer and the question we can answer, which is a lab test.
Let's move to KRAS. We asked the lab test, is the tumor's KRAS wild type or mutated?
The answer is mutated. The critical question

00118

is, being EGFR positive, will Vectibix help you? Now the answer is no. Now, what do you need to bridge between those two questions? We don't have time to present the full analysis, but the two key things are accuracy of the test in the lab and the population epidemiology to the response to chemotherapy. Those are the two key things to know. You could dream up other questions, you know, what about a one in a thousand mutation, but they're much more minor.
Given those two points shown in blue at the top, neither one of them is addressed by an RCT. You may need to address them, you do need to address them, but it doesn't mean that a prospective RCT would address those questions. And in fact with mostly retrospective data, good retrospective RCT type data, all over the world people decided that KRAS is a good thing clinically. I think this is a way to think about why that decision was made the way it is.
DR. GOODMAN: One minute, Dr. Quinn.
DR. QUINN: Thanks. Now there are tests where the bridge between the clinical

00119

question and the lab test does require a prospective RCT, and the same approach will help show why. What's your CYP and VKOR genotype for warfarin metabolism? What should your warfarin dose be? Now here you cannot go straight backwards from that to the question. There are all those other blue boxes that would need to be filled in, and a randomized

prospective trial is the perfect way to do that, because you take one variable, knowing the genotype, you randomize everything else away, and you get the result or the impact of that one variable. And in fact, CMS said that for warfarin safe testing, a prospective randomized trial was the right answer. CMS did not say that for KRAS, and this is just a graphic way of presenting the difference. So I think by framing the questions forwards and backwards in this manner, I think helps focus the decision, so people know what they're talking about, and if they know what you're talking about, people can agree or disagree, and move the process forward. Thank you.

DR. GOODMAN: Thank you very much,

00120

Dr. Quinn. Next is Dr. Jan Nowak from Molecular Diagnostics Laboratory at Evanston Hospital, representing the Association for Molecular Pathology and the College of American Pathologists. Dr. Nowak.

DR. NOWAK: Thank you. I'm here on behalf of the College of American Pathologists and the Association for Molecular Pathology. CAP has a membership of nearly 17,000 pathologists, board certified pathologists and pathologists in training. More than 6,000 laboratories are accredited by the CAP, and approximately 23,000 laboratories are enrolled in the college's proficiency testing program. Many of our members serve as medical directors of clinical laboratories and have had something to say as part of their keener repositories about appropriate test usage. So in fact whether a laboratory offers a test or not does go through the medical director of the laboratory.

AMP has nearly 1,800 physicians and doctoral scientists who perform molecular diagnostic testing, and most of the molecular diagnostic laboratories in this country are

00121

directed by AMP members. So I'm not going to give any clinical data here, but I am going to give you some usage data on these five tests. CAP offers proficiency testing for each of these five analytes, and you can see the various surveys, you can see the enrollments in these proficiency tests, so 1,200 laboratories participate in some kind of HER2 proficiency tests. BCR-ABL, and now this is BCR-ABL quantitation, and it has been

pointed out, that this is a test that's needed to determine major molecular response, this isn't the mutation test, which is very esoteric, and we don't even have a proficiency test for that mutation, so I'm not sure what the issue is there.

And you can see the enrollments in these other tests. CYP2D6, as was pointed out, how complicated a test that is, so there are relatively fewer laboratories doing that, that does not surprise me. Likewise, the BCR-ABL is relatively low because it is not an easy test to give, it requires some expertise.

In preparation for this meeting we performed an impromptu survey of AMP members.

00122

There were 75 respondents and you can see the breakdown of reference laboratories and non-reference laboratories. You can see the number of beds served by the non-reference laboratories, it's all over the place from small hospitals to large major medical centers.

The question, do you perform these tests in-house, and you can see for HER2, BCR-ABL and KRAS, the vast majority do perform the test in-house, and I think that's a reflection of the medical directors' assessments of clinical utility. The numbers are a little bit less for in-house performance of CYP2D6 and UGT1A1, where the test is provided through reference laboratories.

Of these non-reference laboratories, here's an assessment of the volume of testing, and you can see that it's very high for HER2, BCR-ABL and KRAS, and it's somewhat less for CYP2D6 and UGT1A1, possibly reflecting the more limited clinical situations in which those tests are performed.

We asked that same question of the reference laboratories and the numbers are high all across the board. There's a lot of this

00123

testing going on.

So on this survey I took the opportunity to ask the very same questions that the panel is going to address this afternoon about their confidence, whether there's sufficient evidence to determine whether testing affects health outcomes. And so in response to that, you can see that for HER2, BCR-ABL and KRAS, there is huge confidence that there is sufficient evidence to answer these questions. There's somewhat less confidence about CYP2D6 and UGT1A1, but then fewer people

actually perform these tests. I point out the black bars, where people simply said they didn't know, and that points out that there's an educational component in understanding what these tests are and how they're used, and that's important to remember. Clinicians are not aware, or pathologists, we're just simply not aware of how these tests can be used, and that doesn't really reflect on our lack of clinical utility.

DR. GOODMAN: One minute, Dr. Nowak.

DR. NOWAK: In answer to this question whether, their confidence level regarding

00124

improved health outcomes, you can see again for HER2, BCR-ABL and KRAS, the 70 responders to this question were overwhelmingly confident. They were somewhat more guarded in their confidence about CYP2D6 and UGT1A1. And again, I'll point out the green bars, people who simply said that they did not know, means that they're simply not educated about this, they simply aren't in a position to make a decision. So in summary, I think these five tests represent a spectrum of tests, they vary in their clinical applications, their clinical impact and their clinical usage. I think one needs to evaluate each one of these tests on their own in their own clinical situation, as the evidence will not be uniform across the board. In the judgment of molecular diagnostics laboratory directors, the confidence for affecting outcomes is strong to very strong for all five of these tests.

DR. GOODMAN: Thank you, Dr. Nowak, we appreciate your comments. Next is Dr. Steve Brotman, from AdvaMed. Dr. Brotman.

DR. BROTMAN: Thank you. My name is Steve Brotman, I'm a pathologist by training,

00125

and I'm here on behalf of AdvaMed, the Advanced Medical Technology Association. AdvaMed's member companies produce diagnostic products that are transforming health care by enabling earlier disease detection and improved patient management. Our tests are used in clinical laboratories, physicians' offices and homes throughout the world, and our members range from the largest to the smallest in vitro diagnostic technology innovators and companies. Thank you for holding this MedCAC meeting to consider and make recommendations on the evidence that supports the use of specific pharmacogenomic tests in the diagnosis and

treatment of several particular cancers. This issue is an especially important one for Medicare's 44 million beneficiaries. Today the panel will have to evaluate the level of evidence in each of five pharmacogenomic tests and their uses as companion diagnostics, providing information critical for appropriate use of highly potent anti-cancer drugs that must be deployed carefully.

The questions posed to the panel focus on the use of these tests in guiding the use of

00126

specific therapies for particular cancers. These tests offer the hope of using genetic information to speed cancer detection and treatment, to monitor more effectively cancer tumor development, to identify those patients most likely to respond to available anti-cancer regimens, to head off adverse events, and to reduce costs. These tests give us the ability to personalize how medicine is practiced by tailoring care to individual patient needs. We are pleased to see the research efforts of our members bear fruit in the laboratory cancer tests that offer patients the possibility of earlier detection, more effective treatment, better case management and improved patient outcomes.

I would like to leave you with three points today as you consider the evidence bearing on a number of pharmacogenomic tests for cancer. First, AdvaMed members support evidence-based decision-making. The needs of patients, including Medicare beneficiaries, are paramount. And better evidence will result in improved patient outcomes and enhanced beneficiary access to high quality care.

00127

However, we should all be aware that generating evidence on diagnostic tests and other new technologies and procedures is challenging. Tests vary significantly in number and purpose, and the pace in innovation and product development for diagnostics is much quicker than in, for example, the pharmaceutical area. For many diagnostic tests, isolating the impact of the test on health outcomes can be particularly difficult because the patient outcomes typically depend on many factors that go well beyond the information that the diagnostic test provides. Evaluators should be careful not to conclude that the absence of direct evidence means lack of effectiveness.

Secondly, you have to acclimate the research community as a whole to recognize the diversity of the test and the application. Diagnostic tests can be used to detect diseases before symptoms appear, enabling earlier and improved treatments and cures, and when used rationally, can be used to improve patient outcomes and reduce cost of care by determining which patients do or do not require more costly

00128

interventions, and evaluating which physicians are practicing in accordance with evidence-based best practices. They can manage patient care, they can reduce the management of patient care in hospitals where clinical lab tests can be used to determine whether a patient should be admitted and what treatment options should be used, or whether a patient should be discharged. They can also be used to measure or assess quality of care provided to patients with specific conditions. Additionally, they can be used to predict benefits or harms of taking specific medications, moving drug treatment away from a one size fits all approach to the right drug for the right patient or the right dose for the right patient approach. They can also be used to provide patients and physicians with increased control over chronic conditions through personalized realtime treatment and disease management regimens, yielding rapid results tailored to a patient's unique circumstances. They can also be used to allow

00129

providers to conduct a wider variety of tests at a patient's bedside, including pointed care testing and rapid and accurate response to drugs that will improve health outcomes.

DR. GOODMAN: One minute, Dr. Brotman.

DR. BROTMAN: They can provide critical public information on individual population models by identifying appropriate interventions, enabling physicians and patients to make decisions regarding critical biomarkers, or identifies statistically significant populations for continued research. These different types and uses of diagnostic tests demonstrate the multiple applications of these tests and the importance of the assessment of these technologies in light of these varied applications. Third, the development of

pharmacogenomic tests is a rapidly moving area with enormous potential. Pharmacogenomic tests may be able to stratify patient populations based on the risk of suffering a disease, targeting these potent and expensive treatments for those at greatest risk, and minimizing adverse patient events from ineffective

00130

therapy.

DR. GOODMAN: Dr. Brotman, thank you very much for your comments. I'm sorry.

DR. BROTMAN: I have really only 30 more seconds.

DR. GOODMAN: That would be a minute too long. We appreciate your comments very much, sir, and our court reporter appreciates your testing his limits.

(Laughter.)

He may have a word with you about that later, I don't know. He's a former Marine, I might add.

Thank you all to our nine speakers.

We have two people who have, I believe, two minutes each to offer what we're calling open public comments, and these are two minutes each. And since these speakers do not have slides prepared, I'll ask them to make their way to the podium that's in the center of the room for their two-minute presentations, which will also allow our court reporter to change his angle to see you and hear you better.

And the first name that is on the list is Scotti Hutton, from the Colon Cancer

00131

Alliance, I believe I said that right. Is Scotti Hutton present? Keep it to two minutes.

MS. HUTTON: Thank you. Good morning. My name is Scotti Hutton and I am with the Colon Cancer Alliance. I thank the committee for allowing us to speak today. The Colon Cancer Alliance is the oldest and largest national patient advocacy organization in America. It's dedicated to colorectal cancer, which is the second leading cause of cancer death in the U.S. Colorectal cancer takes 50,000 lives each year, with 150,000 being diagnosed. One in 19 Americans will be diagnosed with colorectal cancer, with someone being diagnosed with colorectal cancer every four minutes. 1.2 million Americans are currently battling colorectal cancer and because it is primarily an elderly disease, as the current population ages, those numbers will rise.

Personalized medicine is already having an impact on the colorectal cancer patients' treatment. Molecular testing is being used right now to identify those colon cancer patients likely to benefit from new

00132

treatments, and newly diagnosed patients with early stage colon cancer can now be tested for the likelihood of recurrence. Some day soon we will know which therapies to give to which patients.

The evidence is already there. As we have seen with Erbitux and Vectibix in colorectal cancer patients with a mutant KRAS gene. Thanks to new technology, we are now spared those who will not benefit from unnecessary therapies or ineffective therapies. Such an idea seemed an unattainable dream only a few years ago.

Personalized medicine promises many medical innovations and has the potential to change the way treatments are discovered. It's already clear that personalized medicine promises three key benefits: Better diagnoses and early intervention, more effective drug development, more effective therapies. We all have one goal, an integrated policy framework that balances the interest and health of the patient, protects industry and investment, and scientific interest, without hindering advancement of this tremendously important

00133

sector.

DR. GOODMAN: Thank you very much, Ms. Hutton. We appreciate your comments. Next is Volker Wagner, from AmGen. Mr. Wagner.

DR. WAGNER: Thanks for the opportunity to speak. My name is Volker Wagner, I am a medical oncologist and hematologist, and medical director at AmGen's clinical development oncology.

We would like to make the panel aware of data in previously untreated metastatic colorectal cancer patients that was presented after the cutoff date of Dr. Trikalinos' analysis, data from a randomized study in those patients, that was presented at the European Cancer Conference and also at the ASCO GI in Orlando a few days ago.

In the so-called prime study, more than a thousand patients with previously untreated colorectal cancer were randomized to either a standard chemo or a standard chemo in combination with panitumumab. The trial was

designed to prospectively analyze the treatment effect by KRAS, and in this study in patients with KRAS wild-type tumors, in two months

00134

significantly improved medium progression-free survival in those previously untreated colorectal cancer patients, and so the trial confirmed the predatory nature of KRAS in this setting, and we would be happy to provide further details if needed. Thank you.

DR. GOODMAN: Thank you, Mr. Wagner. Mr. Wagner, has that study made it to the, has it been accepted yet in a peer reviewed publication?

DR. WAGNER: The data from this study has been submitted for publication to the Journal of Clinical Oncology.

DR. GOODMAN: Submitted. Thank you, sir, very much.

Maria, I believe those are our nonregistered speakers; is that correct?

MS. ELLIS: Yes.

DR. GOODMAN: All right then. At this point it would be helpful to the panel if Dr. Freedman and Dr. Trikalinos and team could come to the front so that we can shine the bright light of enlightenment upon you.

Okay, MedCAC. We very much appreciate your ability to drink from a fire hose thus far

00135

today, and we will see what sort of feedback we can give. The time now is for questions to our presenters, and our presenters were Drs. Freedman, Trikalinos and team. Dr. Trikalinos, you have at least one team member with you, I understand?

DR. TRIKALINOS: Yes.

DR. GOODMAN: Thank you. I know all of this, at least appearing from your jottings this morning, a lot of you, a lot of us have questions we would like to bring to bear in here, and what we will try to do if at all possible in a concise way as much as we can, let's anticipate the need for our having to address some questions about outcomes and the adequacy of the available evidence. With that in mind, I want, we will start taking questions. Please be concise with them, please let us know to whom if at all possible they are to be directed, and keep in mind what we are trying to do here. First question, Dr. Satya-Murti.

DR. SATYA-MURTI: Dr. Trikalinos, your tamoxifen metabolites paper, your TA concluded

before the Schroth paper in October JAMA, that
00136
seemed to have further evidence from archival tissue too. Would you have changed any of your conclusions based on that? That paper hasn't been included in the material given to us.
DR. TRIKALINOS: I have not reviewed that paper in detail so I cannot tell you how the conclusions would change if I had reviewed that paper. My conclusions are based on the totality of the evidence with this set of studies, so you would have to integrate this study yourself into the context of the papers that I described.
DR. GOODMAN: Thank you, Dr. Trikalinos. Dr. Trikalinos, I do want to clarify one thing. This panel is addressing the evidence accompanying five tests and your, I believe I noticed this earlier when reading the materials, that your technology assessment did not assess all five tests; is that correct?
DR. TRIKALINOS: No. The technology assessment addresses only the three tests that are described in the title and these were the tests that were set in the beginning.
DR. GOODMAN: Thank you for that clarification. Next is Dr. Mansfield, and Dr.
00137
Kaul. Dr. Mansfield.
DR. MANSFIELD: Dr. Trikalinos, I was curious. At the end of your slides you listed some statistical issues in which you pointed out the possibility of errors due to multiple comparisons as well as that confounding factors are not a problem for germline mutations. However, I was not able to determine to what you were referring when you were discussing those problems. Were there particular studies that had those problems and was it widespread or was it something that we should concern ourselves with?
DR. TRIKALINOS: The general comments about multiplicity of comparisons and assessment of association with outcomes in treatment response, I think that these are perfectly general in all genetic studies and in the totality of this body of evidence. And my personal opinion is that this is also pertinent to pharmacogenetic tests and genetic associations beyond the three ones that we reviewed.
In particular, though, we were motivated to bring this up, especially from
00138

studies that we evaluated in the CYP2D6 example, and as I briefly alluded to in my presentation, it was an opportunity to slice and dice this piece of evidence the way that one sees fit, so one can essentially identify associations. There is no problem in trying to address many many statistical hypotheses, but one has the obligation to properly account for them in the list of comparisons.

When it comes to the germline mutations and the fact that methodologically there is not a need to perform adjustments for germline mutations, this is pertinent only to the CYP2D6 example, and this is a theory, this is epidemiologic principles that dictate this.

DR. MANSFIELD: One more question. Should we assume that the studies that you discussed are flawed in this way?

DR. TRIKALINOS: What the effect of overadjusting is on the actual treatment effects that are described in these studies is not easy to pinpoint. There are methodological papers that show us that overadjusting, especially in the presence of rare outcomes or rare events, may result in associations that

00139

are perhaps even in the wrong direction than one would expect.

DR. GOODMAN: Thank you, Dr. Trikalinos. Next is Dr. Kaul, followed by Dr. Matuszewski.

DR. KAUL: This is for Dr. Trikalinos again, and being a pathologist, I can't help but think about the assay type issues that go into performing these tests. So when you're looking at the KRAS, for example, do you consider, when you're comparing your different studies, how tissue is selected and what assays, the issues of assay analytic performance that might make these results quite disparate amongst the various studies, do you ever get into that level of detail?

DR. TRIKALINOS: I did not get into the level of detail. However, for extractions where information on how the samples were collected and which methods were used to obtain the genetic information. However, we did not find anything in our sample for sensitivity analysis that suggests that there is a difference according, at least with respect to the items that we have extracted.

00140

Is there something that -- Dr. Dahabreh is my colleague, a physician by

training, and he's also part of the team.
DR. GOODMAN: Do you have a specific answer to this?
DR. DAHABREH: The vast majority of studies or data was collected in the direction of sequencing, so we cannot for sure say that if we can't find the difference between the methods because we don't have enough data or there is data out there, but as far as we can actually tell, there is no difference.
DR. GOODMAN: Dr. Kaul, did you have a point to make in light to the response you just heard?
DR. KAUL: I think this is an area that we need to be cautious about, some of these analytic issues, because if one is selecting an entire section of tumor or tissue from a tumor and using a less sensitive analytic method, one might miss KRAS mutations that could be there, so there are some technical issues worth considering.
DR. GOODMAN: Thank you, Dr. Kaul. Dr. Matuszewski is next.

00141

DR. MATUSZEWSKI: This is a question for Dr. Trikalinos. This is a question as it relates to performing the health technology assessment. When you're doing the assessment and categorizing the studies, do you get a sense of what the velocity is of the evidence generation? When you first start a topic, obviously you go back as far as possible, but in some cases you get a sense, is the evidence still evolving, is it plateauing, and are you looking forward to more evidence to review? I think as part of an AHRQ you're looking at creating a report every three years. Are you satisfied that that's an appropriate time interval before you take the next look at all the evidence that could be generated between when you stopped and in three years is adequate, that your conclusions wouldn't change?
DR. TRIKALINOS: That's a very tough question to answer. It is our impression that there's a lot of activity in the CYP2D6 case and there are studies that might be coming out, perhaps in association, more related to association with outcomes, I'm not sure whether

00142

this would be the case. We are not planning to do an update search for this particular report. My answer is a long answer. It's that I do not have a good sense of whether this current body

of evidence is premature, let's say, and a lot of things are coming out.
DR. GOODMAN: Thank you, Dr. Trikalinos. I would add, Dr. Matuszewski, that when we have our discussion period at the end of the day, if you have a thought about any evidence gaps current or anticipated, it might be a good time to maybe make note of that.
Next are Dr. Pao and Dr. Fischer.
DR. PAO: Yeah, I had two questions. First I think may be for Dr. Goodman. In terms of BCR-ABL testing, which tests are we actually talking about? Are we talking about mutation testing or the RT-PCR testing, because they appear to be very different for the questions we're going to be voting on.
DR. JACQUES: We in the question did not differentiate among the various platforms that one might use to identify these particular genetic variations. I think that would lead well beyond the scope of a one-day meeting if

00143

we were to get into that detail. In terms of your questions, think of it along the lines that if there is a test that accurately measures what it purports to measure, does that additional information then affect health outcomes?
DR. PAO: Thank you. And my second question was for Dr. Trikalinos, which is, basically there was a lot of data on the genetic heterogeneity of the CYP2D6. Is it possible to go back and reanalyze those and recategorize those just in the major categories, and then see a better outcome with those categories?
DR. TRIKALINOS: The short answer would be that this would be challenging from the published literature. This is why one of the recommended solutions was that for people who were born with the mutation, that we go on and analyze individual patient data with exact information on a patient level. I presume someone could be a bit liberal and be willing to put together genotypes that would be like; however, this would be open to criticism and moreover it would be introducing some noise

00144

that would perhaps mask associations.
DR. GOODMAN: Thank you. Dr. Fischer is next, followed by Drs. Teutsch, Juhn and Eng. Dr. Fischer.
DR. FISCHER: Thank you, Dr. Goodman. First of all, I think I speak for the panel by

saying that we appreciate the hard work that went into those presentations. As you, Dr. Trikalinos, as you very nicely said at the beginning of your presentation, that you were not going to discuss questions two, three and four, and I think I understand that. However, this panel, I suppose, and discussions in the future will be concerned with the relevance of genetic testing outcomes in patients.

Now, what is interesting to me is that whatever data we have as far as the outcome of the KRAS gene comes from retrospective studies in which outcomes are present. Is there some specific reason why in these types of studies such as you recounted for us, there is seemingly no concern, or perhaps an attempt to, as to find out what happened to these patients and what their outcome was. In other words, was there earlier occurrence, later recurrence,

00145

did they survive, did they not survive? I mean, that is the kind of thing that analysts like this, and I'm sure there will be others, decide whether or not there is a net health benefit to what, the area that you work in. So I mean, is there a particular reason, is it because genetics as far as its evolution has concerned itself more with figuring out what the genetic makeup is and not heretofore concerned with what outcomes as far as human health is? I think I speak for many members of the panel in asking that question.

DR. GOODMAN: Thank you, Dr. Fischer. Just for point of clarification, if the technical person who's managing the slides could go to those questions. I think it's at about slide seven of Dr. Trikalinos' presentation. Dr. Fischer, I think these are the questions to which you are referring.

DR. FISCHER: These are the questions I was referring to, these are the questions that Dr. Trikalinos pointed out.

DR. GOODMAN: So what Dr. Fischer pointed out and he wants you to confirm, you were charged with answering four questions, and

00146

it's not that you didn't seek -- what happens is that you sought the evidence for these and you found nothing dealing with questions two, three or four; is that correct?

DR. TRIKALINOS: So, I would love to summarize questions two, three or four, and we searched for this information. However, this information is not available in the published

studies.
To the second part of your question, you are indeed correct that ultimately the value of the pharmacogenetic tests also is going to be judged by the impact that they have on patient-relevant outcomes. And therefore, this is why we assessed mortality, disease progression, and also this softer outcome, the intermediate outcome, and as you can see, we had limited evidence for the current outcomes.
DR. GOODMAN: Dr. Trikalinos, just to make sure I understand, you did get evidence for question one, which deals with response to therapy.
DR. TRIKALINOS: Yes.
DR. GOODMAN: Is that, for example, progression-free survival?

00147

DR. TRIKALINOS: This would be mortality, progression-free survival, or response according to the intermediate outcome of disease treatment failure. The intermediate outcome was, if you recall, cytogenetic response for CML, and it was treatment failure by radiologic imaging for the KRAS.
DR. GOODMAN: So for the three of our five tests that you examined, for the three of five tests, there was evidence for question one, and some of that evidence included mortality, morbidity?
DR. TRIKALINOS: Mortality, progression-free survival, which means disease progression.
DR. GOODMAN: Mortality is not the same as progression-free survival, correct?
DR. TRIKALINOS: Correct. So progression would be either death or worsening of the disease generally speaking.
DR. GOODMAN: Progression-free survival does not mean longer survival?
DR. TRIKALINOS: It doesn't mean longer overall survival necessarily. So mortality -- let me recast it. Overall

00148

survival, which means live or die. Progression-free survival, which means time to either death or progression of the disease. So you can see that progression-free survival is a composite of it, it consists of either of two events, whichever happens first.
DR. GOODMAN: Okay. And you've got nothing on question four, which looks at benefits and harms. Were those benefits and harms or adverse effects dealing with the tests

themselves immediately?
DR. TRIKALINOS: So, these are benefits and harms that are incurred by the process of testing for these particular tests.
DR. GOODMAN: For the process of testing.
DR. TRIKALINOS: So for example, this could be increased anxiety, or quality of life, or --
DR. GOODMAN: Thank you. I wanted to get -- by the way, I skipped Dr. Scheuner, who was next, followed by Dr. Teutsch.
DR. SCHEUNER: I guess I have a couple questions of Dr. Trikalinos, and maybe Dr. Freedman too. So the paper by Schroth,

00149

et al. in 2009 in JAMA that we had access to, you did not include in your assessment, but Dr. Freedman did I believe allude to the article in his presentation. And I'm just wondering if he might give us what his impressions were of that article and does it maybe give us, you mentioned some large sample size, and it did have some statistically significant data in, I believe it was disease-free survival and maybe even overall survival?
DR. FREEDMAN: Again, I would really not comment about the strength of the evidence of that study. My point in putting up that study, although it wasn't in the technical evaluation, was to show that larger cohort studies are needed to demonstrate some of these effects that we might not see in all retrospective analyses or clinical trials. I think it's a very interesting study, I think it's something that needs to be looked at closely, but I haven't reviewed the study close enough to form an opinion or advice on that.
DR. GOODMAN: Thank you. Dr. Scheuner, did you have a follow-up question?

00150

DR. SCHEUNER: I have a different question, yes.
DR. GOODMAN: Why don't you just take that question and we'll move on. Go ahead.
DR. SCHEUNER: It has to do with the BCR-ABL issue again, so I guess Dr. Trikalinos, you could be the one who might answer. We read in the materials regarding diagnosis of typical CML versus atypical CML. We read about monitoring of the disease, but it appears that the review was specific to looking at different mutations in the tyrosine kinase domain and how

those affect response to therapy. So those are three separate indications for a test that it sounds like we're calling it all-in-one test, but it's actually for three different things. So I think for the panel's benefit we need to be very specific about what you want us to vote on, because I would vote differently for diagnostic purposes, monitoring purposes, and then making a decision about therapy.

DR. GOODMAN: Thank you.

Dr. Jacques -- I would remind you, though, that the questions say health outcomes, you get there one way or another with health outcomes.

00151

Dr. Jacques.

DR. JACQUES: Yes. Essentially when one looks at the regulatory framework with which the Medicare program deals with tests, there is a section in the Code of Federal Regulations that says essentially, a diagnostic test, at least as a minimum, must be ordered by the physician treating the patient and must be used by that physician to essentially guide the management of that patient.

So when we get to any particular question, if the panel feels like their response would be nuanced based on how that interpretation would be, what usually happens is if the panel all agrees that they, or in general or by consensus, and the chair agrees that the question needs to be addressed only in a particular context, then the panel will vote with a common context. On occasions, sometimes there is not necessarily consensus about that, and what will happen will be, essentially the panel is asked to vote on it sort of as it stands and then in follow-up discussion individual panel members may say, you know, I voted this way because of this; if the question

00152

were asked in a different way, I might do it a little bit differently.

DR. GOODMAN: Dr. Scheuner, does that help?

DR. SCHEUNER: No, it doesn't help. I think you're not understanding what I'm asking, that when you have a patient with CML, we look at the 9;22 translocation, the Philadelphia chromosome, and that tells us if it's typical CML, which is like 90, 95 percent, versus atypical. And then we can also look and see with molecular and cytogenetic techniques to monitor response to therapy. And then lastly, there's this issue of looking at specific

mutations that might affect response to therapy. So there are three different things. And I think, in my understanding of what was presented to us from the AHRQ review, it's the last thing and only the last thing that was assessed in the technology assessment, and he is nodding his head, but could he answer, am I correct?

DR. TRIKALINOS: Yes, that is correct. Our review does not assess tumor load, these were considered different tests. Our test is

00153

mutations.

DR. GOODMAN: Dr. Scheuner, are you okay with that?

DR. SCHEUNER: Yes, he answered my question.

DR. GOODMAN: We're going to go to Dr. Teutsch next.

DR. TEUTSCH: Dr. Trikalinos, we got into this a little bit a moment ago when you said you did not really find the information about harms associated with testing, presumably false positives and false negatives. Can you talk a little bit about how that relates to what we really care about, what's the incremental value of these tests, particularly vis-a-vis alternative therapies that might happen if the testing were not done. Was that looked at, was there evidence of that so that we can truly assess the harms and benefits or the alternatives?

DR. TRIKALINOS: We did not find evidence on benefits and harms in the reviewed literature. There might be evidence on benefits and harms on genetic tests in general, and I'm just making this clarification first.

00154

Now what you're asking is essentially you're putting me on the spot to try to give you my assessment of the interplay or the likelihood of having the disease, having the test, the downstream effects of, whatnot, and this is something that ideally would be done in the context of a decision analysis or a singular analysis, or a prospective trial.

DR. TEUTSCH: So what I guess I wanted clarification on, we really don't have information about that evidence that would inform therapeutic choice, so it would be based on inference, secondary kind of information, correct?

DR. TRIKALINOS: Our report did not find this kind of information in the actual

studies.
DR. TEUTSCH: Do we know that it exists and we simply didn't find it or do we know --
DR. TRIKALINOS: If it were --
DR. TEUTSCH: -- that it does not exist?
DR. TRIKALINOS: So, my interpretation is that it's not there. But the decision
00155
analysis or whatever like that, should be done by you.
DR. GOODMAN: Thank you, Dr. Trikalinos. Dr. Teutsch, you got your answer; is that correct?
DR. TEUTSCH: Yes.
DR. GOODMAN: Okay, thank you. Dr. Juhn is next.
DR. JUHN: My question has to do really more on pathology and less to do with the specific questions of the specific items that we're looking at today, and the methodologic question really has to do with some of your general comments about heterogeneity, especially with the mutations because of the classification or categorization issues. So my question really has to do with applying so-called standard technology assessment approaches, perhaps extracted from the way that we look at more traditional diagnostic tests, how would those approaches have to change to take account for this heterogeneity?
DR. TRIKALINOS: That additional evidence in this case --
00156
DR. JUHN: And maybe the prior question to that is should they change, should we try to use the same framework that we have for, let's say A1c testing, which is pretty linear in terms of the different categories, and use that same type of standard methodology for reviewing the literature, and then apply it to this area where you have really a host of different test characteristics?
DR. TRIKALINOS: My quick answer is that this particular challenge is the limitations of the evidence itself, the fact that there's a lot of heterogeneity. There are methods in the array of methods that we have in evidence-based synthesis that can account for heterogeneity. However, these do not really give you the answer that you're ultimately interested in. These methods can only tell you

that there is a distribution of three or four effects. And these, with random effects, distribution has a given meaning and a given heterogeneity given by ability, but this is not informative. So this is a major limitation of the data, and my gut feeling is that there's no methodologic advance than can go around it.

00157

The only thing, I think, would be to actually get the individual patient data.

DR. JUHN: And so many times, these technology assessments serve as a guideline or set of guidelines for future investigators. So I guess my question really has to do with a very practical question, which is if someone is trying to design a study looking at these various questions, they look at your technology assessment and they look at some of the flaws that you've seen in the current studies, what is the specific advice that you give them to say the next time you do this assessment, you know, this TA three years from now, and this person's paper has met the various criteria --

DR. GOODMAN: Allow me to interject, Dr. Juhn. That's a fascinating question. At this point in our discussion, I think it's probably not the best way to spend our time. It may be a great thing to discuss at the end of the day once we've taken a more careful look at the evidence, if you don't mind. But it is a superb question and we appreciate it. Dr. Eng is next, followed by Dr. Hayes and Dr. Satya-Murti. Dr. Eng.

00158

DR. ENG: My question is for Dr. Trikalinos. I am referring to the slide, the KRAS slide on mortality. I don't know what the number of the slide is, but my copy says 19 out of 29.

DR. GOODMAN: So that is probably the 37th or 38th slide.

DR. ENG: The second bullet says, in nine of the 18 studies, analyses were statistically significant. So my question as I'm trying to see the relevance and the importance to the Medicare population is that on a previous slide you said that when you looked at the KRAS studies, 22 of the 28 studies had a mean or median age greater than 60. So my question is, how many of these nine that were statistically significant in favor, in the direction of favoring this test to look at the effect, of those nine, what was the median age?

DR. TRIKALINOS: I cannot give you this answer off the top of my head, so I would have to go back to the studies and see which ones were there. Short answer, though, is that all the studies seemed to, their point

00159

estimates are in the same direction, and the fact that some of them are statistically significant where others are not may be a factor of their size. I can understand your question. I would have to go back and see how many there are.

(Discussion off the record.)

The other comment is that the median age of the other studies doesn't mean that it's much much younger than that.

DR. ENG: Yes, I understand that, but this is a really critical point for me. This was the only discussion in which you actually mention there were a number of studies that were statistically significant in terms of the three tests that you were looking at. And also, this one was also the one that had the higher median average age.

DR. TRIKALINOS: So, I cannot give you this answer right now, but I could give you this answer after some more calculations.

DR. GOODMAN: Is this something you could get during lunch today or at some later time?

DR. TRIKALINOS: My colleagues say

00160

that we will get you this information.

DR. GOODMAN: So the answer is yes. Thank you, Dr. Eng. Dr. Hayes is next, followed by Dr. Satya-Murti. And panel, please keep our answers focused, let's look at the evidence and stay on point. Dr. Hayes.

DR. HAYES: In my opinion a lot of your discussion was based on prognosis and prediction as you went through, and at the end it was clear that you separated things. But the question remains with BCR-ABL, the same question that you should have answered, I'm not sure that you answered it. It seems like there are three uses. One is diagnosis, is this atypical CML? Second is prediction, do we expect that any of these three drugs is likely to work to select patients. And the third is monitoring, can we take that patient and monitor their progress after we start the treatment.

And it seems like you mixed all three of those in your final summary, and while Dr.

Jacques gave us sort of an analysis, it's hard for me to score until I know which of those three things you analyzed, rather than just

00161

giving a yes or no.

DR. TRIKALINOS: So, let me clarify. If you can imagine it like a table of six things, so we have two types of tests, the mutations and the non-mutations from our transcript levels and other studies, and also the three topics that you mention now, which is differentiating between typical and atypical, prognosis, which actually I would break down into a prognosis for first line treatment, second line and third line treatment, and the third thing is monitoring. And it was pointed out this morning that the context of tumor load and transcript load is something that's done and it's something that's, as I perceived, mainstream.

We are focusing only on mutation testing, so our technology assessment distinguished between these types of studies. We did not assess differences between typical and atypical CML, so this is something that we did not review, in a sense. However, we distinguished between prognosis in the three settings, first, second and third line therapy, and monitoring studies.

00162

I did not present you any results of the monitoring studies but they are in the report, and if I may briefly summarize, mutation testing for monitoring studies, there were a relatively small number of studies, I don't remember, so what I mean is people who started on a treatment, they were started automatically and then, for example, the patient samples are tested every month, three months at the beginning, and then six months, so there is a lot of variability in the testing intervals, there's a lot of variability in the outcomes assessed, and we could not actually pin down any information that was very very useful from these types of studies. They were so heterogeneous, first in the interval of testing and the frequency of the mutation testing, and secondly the outcomes that they actually described. It's not -- my perhaps way of putting it is that perhaps there are studies that were describing more about the pathophysiology of the disease rather than informing us on the frequency or prediction of the final outcome.

DR. GOODMAN: Thank you, Dr.
00163
Trikalinos. We're going to take one more question before our scheduled lunch break, from Dr. Satya-Murti. Doctor.
DR. SATYA-MURTI: Two points here. The Schroth paper seems to have impressed more than one of us, and of course you didn't look at it, I understand that. But as I interpret the paper, it still answers your key question one in that a PM, poor metabolizer, had a poorer prognosis, but it does not answer the rest of the questions also, so it very much tallies with what you said, that there is no overall difference in survival, and they admit to a limitation, they acknowledge that it was done from archival tissues variation, so I had interpreted it, and Jeff Roche had also looked at it.
My question then in terms of overall survival, and this will go to Dr. Freedman too, is in cancer epidemiology, what is the survival benefit, is there a general consensus as to months or years? Because in our outcomes we're talking about survival benefit in addition to PFS, so are there any metrics there that either of you can tell us?
00164
DR. TRIKALINOS: So your question is perhaps, what would be a minimally clinically important survival difference?
DR. SATYA-MURTI: Yes.
DR. TRIKALINOS: My short answer is that this would depend on the disease. Faster killers would have a different minimal clinical important difference. I don't have a number for you for the three disease conditions that I described.
DR. GOODMAN: Dr. Freedman, on that point?
DR. FREEDMAN: Just real quickly, it really depends on the question and the evaluator. Some clinical trials say a few weeks is important, a few months, a few years, so it really depends on the question and who is evaluating the evidence.
DR. GOODMAN: Okay, thank you. Panel, with that we're going to proceed to lunch. During your repast we're going to think about what it takes for us to get to answers for our questions. We've had some clarification on the roles of these tests. You're free to talk, if you want to talk about that amongst yourselves
00165

at lunch, that's fine. We're going to focus in on trying to fill in those information gaps and then move to our voting questions ultimately. We will reconvene at one o'clock promptly. See you then.
(Recess.)
DR. GOODMAN: As we reconvene, I ask that our presenters, which are Dr. Freedman, Dr. Trikalinos and team, come to the front of the room. And then please do be available, those of you who were our scheduled public commenters, because we consider you experts as well, and we hope that you will be available. While some of our questions will be directed to our first presenters, some questions may be directed to some of our scheduled public commenters as well, the idea being that this is a little bit broader take than sometimes we do, we want to be able to draw from the expertise of the room, and it goes beyond just our first couple of presenters.
And of course there's going to be some discussion among our panel. Here's what we're going to do. We know that we've got five tests about which we are going to be answering

00166

questions, and for each of those we have to look at the adequacy of the evidence upon which to make some finding, we're going to ask about what that finding is.
We recognize a couple things. First of all, there are five tests and we heard only about three of them from the technology assessment, we know that there's not as much information as we might like, that's one important consideration. And another important consideration, that among these five tests, I would suppose it's probably BCR-ABL mostly, for which there are multiple applications of the test. And you heard our panel ask and deliberate a little bit this morning about which application are you talking about.
So we're going to go through these one by one, and basically this panel is in search of evidence for outcomes. We all recognize that the evidence in some cases is kind of patchy, there seems to be strong evidence in a few places for some applications, weak evidence in others, nonexistent evidence in others. So this panel is looking for, give us some evidence, will you, on the impact of these

00167

tests on healthcare outcomes, and we will try to work it that way. And again, we recognize

that there's sort of new doubts on what we've heard so far vis-a-vis the TA and so forth. And for each one of these in our discussion, we're going to start out with is there enough evidence to go on, and then we'll move to what might that evidence say, in sort of a lineup that, well, we hope to get to our voting session. Is that okay, panel, as an approach, an imperfect but perhaps practical trial? Let's do this. Test (a), CYP2D6 for breast cancer patients who are candidates for tamoxifen, this was one of the three that we heard about from the folks at Tufts and we're going to ask, and the panel can chime in, we're going to focus now on, is there enough evidence for this test upon which to make some decision about its impact on health outcomes, okay? Not surrogate measures, health outcomes. And I would ask, start with you, Dr. Trikalinos, you've worn your path well in the carpet from that chair to the microphone. Considering CYP2D6 for breast cancer patients who are candidates for tamoxifen, starting with

00168

your technology assessment, did you find, just summary for us, if you would, whether you found sufficient evidence upon which to make some judgment or observation or finding about the impact of the test on healthcare outcomes.

DR. TRIKALINOS: So in summary, our review of these studies suggests that there is inconsistent evidence on whether genetic variations in CYP2D6 can predict response to treatment when it comes to survival, progression-free survival.

DR. GOODMAN: So you did find evidence, and you found it to be inconsistent with regard to impacting patient outcomes?

DR. TRIKALINOS: There are studies that give us some information on these outcomes and the other clinical outcomes. However, these studies are, first of all, heterogeneous as I described, and they point to different directions.

DR. GOODMAN: So they're heterogeneous and the results point in different directions with regard to healthcare outcomes.

DR. TRIKALINOS: With regard to the healthcare outcomes.

00169

DR. GOODMAN: Thank you. Now, panel, does any panelist have a question on this issue or the sufficiency of evidence for this test? You can direct the question to any of our

initial presenters or perhaps any of the other nine presenters who came after. Dr. Mansfield, is it?

DR. MANSFIELD: Yes. I don't know who might be able to answer this, but is there any reason to believe that the recent Schroth paper would change the overall trend of evidence in one way or another, and does it have similar methodological flaws to the other studies?

DR. GOODMAN: Remind us. Was that study to which you referred included or not in the technology assessment?

DR. MANSFIELD: It was not. It was published after they did their assessment, I believe.

DR. GOODMAN: And it's in peer reviewed literature, correct?

DR. MANSFIELD: It's in our packets.

DR. GOODMAN: Just wanted to make sure we stated that. Dr. Hayes, on this point?

DR. HAYES: And I will be specific to

00170

this point. I'm actually part of the group that generated the initial pharmacokinetic information suggesting that endoxifen was not produced in patients who are variant variant for CYP2D6, so at least we got the ball rolling. I'm not an author on any of the outcomes papers that have come since. I actually agree very much with your assessment. The Schroth paper is curious in that a third of those patients came from the Mayo Clinic study which had been previously published not once, but twice before. And if you take those out, then you're really left with yet one more study that's retrospective from Germany in which the samples were collected from other resources, so I think it has many of the very same flaws that you pointed out. It is large but there aren't that many events; in fact, there are only about seven events when you take out the Mayo Clinic study. So our group, which we called the Consortium on Breast Cancer Pharmacogenomics, also known as COBRA, has been very cautious about making recommendations on this. There is yet another report, not

00171

published in the peer reviewed literature, but presented at the San Antonio Breast Cancer Symposium, in which a lot of people got together and called themselves the tamoxifen pharmacogenomics group, very similar to the warfarin group, and those data were presented

by Dr. Goetz, and curiously, the outcomes for patients on tam, on tamoxifen who were either poor metabolizers or rapid metabolizers were absolutely overlapping in terms of Kaplan-Meier curves. He then went on to explain why he thought that the process they had gone through was flawed and that he didn't believe the data he was showing, and that led a lot of us to wonder what was going on. So I agree with you, I think the data are still quite mixed.

DR. GOODMAN: So Dr. Hayes, in response to Dr. Mansfield's very well phrased question, it sounds like you're saying that the additional information of the Schroth paper does not move us in one direction or another very strongly.

DR. HAYES: From my standpoint, I do not believe it does.

DR. GOODMAN: That's very helpful.

00172

Dr. Scheuner, did you have a point on that? I saw you nodding your head.

DR. SCHEUNER: No. I just appreciate that comment very much.

DR. GOODMAN: So do we. Any other questions for our presenters or the speakers on the matter of CYP2D6 for breast cancer patients as candidates for tamoxifen, with regard to the sufficiency of the evidence? Dr. Satya-Murti.

DR. SATYA-MURTI: The sufficiency part I don't know, but one of the presenters today, Dr. Teagarden on the PBM, you mentioned that you have your patients undergo the CYP2D6 testing, if I interpreted you correctly, and then you go on to pharmacy management. Who pays for these tests when you have the patients go through that?

SPEAKER: I think that Dr. Teagarden left.

DR. GOODMAN: Is Dr. Teagarden here? He is not here. Any other points or questions about the sufficiency of evidence for this test? Dr. Teutsch.

DR. TEUTSCH: Could I just get clarification? Are we talking about

00173

sufficiency of evidence in the sense that based on the test as is currently being practiced, or as it might be sometime in the future?

DR. GOODMAN: Well, we're talking about the sufficiency of available evidence, so if there's not evidence of how it might be used in the future, I would guess that --

DR. TEUTSCH: What we heard is that

the test is being, because there's classifications all over the map, that's part of the heterogeneity of these studies. And we could say well, if somebody actually -- in that sense the information is probably insufficient where it's probably being done. On the other hand, if you say that's as good as the test is ever going to get, you can say the evidence is insufficient and you shouldn't do it.

DR. GOODMAN: That's a point well made, and that might enter into on how you vote on that question, Dr. Teutsch. Dr. Juhn, on this point?

DR. JUHN: I wanted to address the Schroth paper that we have been addressing from October 2009, and again, I'm not familiar with that literature. But the paper itself, and if

00174

you look on face value, the number of patients they had, there it wasn't a classification issue because they were very consistent about their classifications. I found it actually to be quite convincing that there really was a separation of the curves between especially the two extreme groups, the poor metabolizers and the rapid metabolizers.

DR. GOODMAN: So Dr. Juhn, you're submitting that the Schroth paper does provide some additional useful evidence.

DR. JUHN: I think so, and this I think is a challenge with a technology assessment, that it's a snapshot in time and it's not an ongoing snapshot, and I think this JAMA paper in October 2009 should be considered as a way of interpreting the pretty equivocal findings of the TA.

DR. GOODMAN: Other comments on that point? And again I would remind us, we can only deal with the evidence that's available. Was there a follow-up point to Dr. Juhn's point on that paper, the Schroth paper? Okay. Do any of our presenters have a point to make about the availability, sufficiency of

00175

the evidence on this test? Seeing none, all right.

Let's force the issue now a little bit. We talked a little bit about what evidence is available, we heard Dr., the folks from Tufts say that it's inconsistent, heterogeneous, seemed to point in different directions. We had some discussion about this last paper. Does anybody have any questions or statements they want to make about what this

evidence says with regard to impact on health outcomes? Because we heard some discussion about the tests, about the evidence, and we characterized what the evidence looks like. Now let's look at what might the evidence be telling us about impact on outcomes. Dr. Scheuner, is that you?

DR. SCHEUNER: I guess I have a question about the heterogeneity and the different classification of the subjects, and I was just curious if population stratification or, you know, with the different parts of the world or different ethnic groups would explain the different classification of the same genotype.

00176

DR. GOODMAN: Dr. Trikalinos, that's a great question. Heterogeneity might just mean that you were looking at different subpopulations. What does it mean?

DR. TRIKALINOS: So within each study, I don't think that we have a major issue for population certification. Population certification is sort of a confounder, quote-unquote, that would be a problem in association status, but might not be confounding in its epidemiological sense. And what it says is that you may be mixing different people from different descents or from different genetic background, and the mere fact that allele frequencies differ across different population strata may induce a foreign association when you actually do the analysis. Now all of these studies, I don't think that population stratification was a problem in these studies. Population stratification would be a problem if you had mixing of population within each study. We don't really see this as a problem if we have different studies conducted in different

00177

regions or different places. If this is not an issue within each study, then a study in people of one descent and people of a different descent would give you an unbiased measure of the genetic effect.

DR. GOODMAN: So Dr. Trikalinos, you're saying then -- please say yes or no.

DR. TRIKALINOS: In my opinion I don't think that population stratification is a major threat or a major explanatory force for the heterogeneity in the actual results.

DR. GOODMAN: That's helpful to note,

thank you. Dr. Scheuner, did that help answer your question?
DR. SCHEUNER: Yes, I suppose so. So that in terms of a particular genotype, the association is always going in the same direction in every population studied; is that correct?
DR. TRIKALINOS: No. Actually what we are seeing is that for particular genotypes the associations are not in the same direction. That's what I commented on as heterogeneity. They are pointing to different directions and some of them are significant in one direction,

00178

some of them are not.
DR. GOODMAN: Good, thank you very much. Dr. Fischer I think was next. I'll get to everybody.
DR. FISCHER: You know, about the comments that I'm hearing about the paper in JAMA, I think Maren, you were on the TEC when we accepted the single study from the Society for Cardiac Surgery on cardiac revascularization, and then had to eat our words a year later when four studies came out and said the TEC was totally wrong. So I would urge us, you know, it may be a very good paper, I think the fact that you have a population that's previously been published twice and you wonder why it's included, and when the gentleman sitting to my right who is quite knowledgeable says that, it really actually leads me to say we should disregard that paper.
DR. GOODMAN: Dr. Fischer is saying disregard. Okay. It sounds like we have a mixed opinion here, that happens in real life, on that particular paper. Dr. Satya-Murti is next.
DR. SATYA-MURTI: We have clinical

00179

oncologists in practice, John and Dan and others, presumably. So let's say in an ideal world circumstance, you know the status going on forward with the treatment. So if you already knew it prospectively going forward, would an oncologist so alter his or her therapy in practical terms at bedside, would any of you change your mode of therapy based on the results, given that the result is already available, and not only for this test, but would it apply to others?
DR. GOODMAN: Let's start with Dr. Hayes on that point, but try to be concise about this.

DR. HAYES: I'll try to be. I think the answer is, would and should would be two different answers. For CYP2D6, especially in the Medicare population, the stakes aren't that high because we have another class of drugs that are quite effective. So you can argue even if these data are wrong, we're treating many of these women with an AI already, but some women can't tolerate an AI and tamoxifen is a perfectly good alternative. And if we have said that if their CYP2D6 is variant

00180

variant and if that's wrong, it means now they can't take an AI, and we've just told them that tamoxifen won't work. So again, our concern is whether these data are real or not yet, and there are many biological reasons why they might not be.

It's a great idea. So for the Medicare population there is some concern; for the younger women there's a huge concern because then the concern is you have to turn off or take out their ovaries and put them on complete estrogen depletion, and that may have real health effects in terms of long-term survival. We believe this is a very important question that needs to be answered.

DR. GOODMAN: Great, thanks, Dr. Hayes. Dr. Pao.

DR. PAO: Yes. I know Dr. Teagarden is not available, but he did present data on Medco actually doing the testing and basically (inaudible) tamoxifen. So I was wondering if there was any outcomes data from that study or from that practice showing that there was a benefit to the patients who actually switched over, if anyone is aware of that?

00181

DR. GOODMAN: It hasn't been presented today, correct?

DR. PAO: I didn't see it.

DR. GOODMAN: I don't see anyone leaping to the microphone with a response to that. Let's point up the question a little bit more even now and I will ask our presenters, whether Dr. Freedman or Dr. Trikalinos and team or perhaps others. We talked about the availability of evidence, we've talked a little bit now about where the evidence might point. What is the most persuasive rigorous evidence for impact of this test on healthcare outcomes? What's the best thing we can point to for evidence of impact of the test on outcomes for this indication? Dr. Mansfield.

DR. MANSFIELD: Can I ask for a clarification on what the desired outcome is?
DR. GOODMAN: Well, it's defined here as health outcomes, and we talked about that in terms of morbidity, mortality and health-related quality of life. Dr. Trikalinos, do you want to give us some guidance on that, health outcomes?
DR. TRIKALINOS: The outcomes that

00182

were assessed were essentially mortality and progression-free survival or progression-free disease. That's what the evidence reported for CYP2D6. There may be other outcomes that the panel may need to debate on others.
DR. GOODMAN: Dr. Mansfield, did you want to respond, or is that sufficient.
DR. MANSFIELD: That's sufficient, thank you.
DR. GOODMAN: Thank you. Dr. Scheuner.
DR. SCHEUNER: I just have kind of a question/comment. Back in February 2009 when MedCAC met, and so this is in the slide presentation, the very first slide, where it looks like we defined patient-centered health outcome. This idea of change in patient management by physician was considered in the context of health outcomes. So I just want to clarify that the TEC assessment, you didn't search for literature that addressed change in behavior by a physician, you only looked for the indirect and direct healthcare outcomes, which were items B and C on this slide.
DR. TRIKALINOS: I should remind that

00183

key question three assesses impact on diagnostic thinking, which would fall under this area. We found no studies that could answer key question three.
DR. SCHEUNER: Okay. So even with the BCR-ABL, I'm going to just -- I'm sorry, I can't go to that.
DR. GOODMAN: Not yet. I promise, we will get to BCR-ABL. So in response to that question, they looked for it, they didn't find it, as I understand it.
On this question, Dr. Hayes.
DR. HAYES: So, I'll speak up again and I want to say, we hope that the answer to this is right, CYP2D6 is predictive for tamoxifen. There's a single paper that was published in the Journal of the National Cancer Institute by Tuglia, et al., in which they

modeled that if you take patients with tamoxifen and take out the patients who are variant variant, and then compare the expected outcomes of patients who are wild-type wild-type, and they're rapid metabolizers, two women who got aromatase inhibitor without any sort of selection, the patients who were

00184

wild-type wild-type with tamoxifen should actually have a better long-term outcome than similar patients who got an aromatase inhibitor. But that was based on only the Goetz study.

And our concern is not that, we're not saying that we don't think it's right, we're just saying that we don't think we have the data to tell if it's right or not yet. The data are mixed, as we've heard, and there are several other studies coming down the pike.

DR. GOODMAN: So thus far, I don't think we've heard anything in addition to the following, that the evidence is inconsistent, the results point in heterogeneous directions. Have we heard anything that is a conclusive concise bit of evidence showing that this test has an impact on health outcomes as defined previously? I don't see anything on this point, okay?

All right. Let's move on. The next test, (b) in our list, is one for, about which we did not hear from the technology assessment, which by the way was a superb technology assessment, thank you. We understand that you

00185

weren't asked to look at all these tests. And this one that's up now is UGT1A1 for colon cancer patients who are candidates for irinotecan.

Now, have we heard any presentation or evidence on this matter this morning from anyone? I don't believe we've heard anything. And was there anything in your packet that addressed this? Okay. Which panelist might want to speak to that issue? Steve Teutsch is being thrust forward by Dr. Scheuner, and Dr. Teutsch, if you will just help us out here, we want to kind of do this in two pieces. The first piece deals with the sufficiency or adequacy of evidence and then we'll get to, if there is what we consider to be sufficient or adequate evidence, then we'll get to what it says, if you could sort of bifurcate your response that way.

DR. TEUTSCH: I wish I could be real

-- this is based on recollection from about a year ago, but the EGAPP group reviewed UGT1A1 --
DR. GOODMAN: I have to interrupt you. Please tell us again what EGAPP is; not

00186

everyone knows.
DR. TEUTSCH: Evaluation of Genomics Applications in Practice and Prevention. It's a CDC-based panel that's an independent federal panel, that was constituted to review, to develop methods and case studies for evidence-based practice in genomics.
DR. GOODMAN: Superb, thank you. Proceed.
DR. TEUTSCH: And one of the topics was UGT1A1 for this indication, and you have the recommendation from that group and I was on the group, so I already disclosed that conflict of interest. I don't recall all the evidence, but basically the primary question that we were originally asked was does it reduce harmful side effects, particularly diarrhea and neutropenia. But in the course of that investigation that was expanded to what are the tradeoffs, and what became apparent is that while there was a good bit of evidence about the fact that you could change the dose and reduce the dose, coming from mostly retrospective kinds of studies, that there looked like there was a suggestion, but not

00187

based on sufficient evidence, that there were tradeoffs in the benefits in terms of the impact on the effectiveness of the drug.
DR. GOODMAN: Did you say sufficient or insufficient evidence?
DR. TEUTSCH: There was insufficient but suggestive kind of evidence which led the EGAPP panel to make the recommendation that information was currently insufficient to assess the tradeoffs between the, if you will, improvements in reducing harms, and the potential decrease in benefits from a dose adjustment of the therapy.
DR. GOODMAN: So insufficient but suggestive evidence --
DR. TEUTSCH: I think the way I interpret the bottom line is, if you decrease, if you have an effective lowering of the blood levels, you decrease the harms and you decrease the effectiveness of therapy, and concomitantly if you increase them, you increase the harms and you increase the benefits potentially. And

that was why, because the studies were mostly about harms, you'd get one sense, but if you looked at it overall, we couldn't assess what

00188

those actual tradeoffs were.
DR. GOODMAN: You could not. And the quality, Dr. Teutsch, of the chain of evidence from the test to our health outcomes, you would characterize as how, strong, weak?
DR. TEUTSCH: Fair, because they were all retrospective for the harms and the benefits were really not clear. That is, they went back and looked at them in patients who had gotten the drugs.
DR. GOODMAN: So the evidence for adverse effects was fair, and the evidence for benefits --
DR. TEUTSCH: Insufficient but suggestive. That led to an overall conclusion that we could not make a recommendation as to whether or not it was helpful to do the testing.
DR. GOODMAN: So from the standpoint of EGAPP, the evidence was not sufficient to make that finding.
DR. TEUTSCH: I believe so.
DR. GOODMAN: Thank you, Dr. Teutsch. Any other panelists want to comment on the sufficiency of the evidence with regard to

00189

UGT1A1 for colon cancer, the sufficiency of the evidence? Yes, Dr. Fischer.
DR. FISCHER: I can just read the article entitled An Evidence-Based Review, in which they tried to review all the evidence.
DR. GOODMAN: You won't read the whole article to us, will you?
DR. FISCHER: No. Given the large number of colorectal cancer cases diagnosed each year, a randomized control trial on the effects of irinotecan, those modifications in patients with colorectal cancer. So they called for further studies and a promising trial.
DR. GOODMAN: They called for further studies. Dr. Cox, on this point.
DR. COX: I obviously practice clinical oncology and treat these patients and I must say that the tests, I can speak a bit for the evidence, though I'm not claiming to have reviewed this entirety of evidence. But as it relates to outcomes in patient care, which you reminded us here, these questions all lead up to, I don't think the sufficiency of

the evidence is there for this to be used in prime time to be making decisions about patients.

DR. GOODMAN: Insufficient evidence for use in prime time. Okay. Any other comments about the sufficiency of evidence and that indication? Yes, Dr. Scheuner.

DR. SCHEUNER: So in looking at the EGAPP review with respect to clinical utility and options for modifying patient care, we've heard about modifying the irinotecan regimen or the dosing, and that by lowering the dose we don't know if we're going to also lower the effectiveness of treatment. But there were two other options that were suggested, and that is using another drug instead of irinotecan, and to treat patients with colony-stimulating factor before the first cycle of chemotherapy to prevent the occurrence of febrile neutropenia.

So again, I think in reading this review, and maybe Dr. Teutsch could confirm that again, it was limited to no evidence on those alternative strategies as well.

DR. GOODMAN: Thank you for raising that point because if there were, that would still be useful. Dr. Teutsch, did you want to add, or did Dr. Scheuner cover it?

DR. TEUTSCH: That's fair. I mean, it was part of the implication.

DR. GOODMAN: Thank you. Dr. Janjan.

DR. JANJAN: I think one thing we're missing here with regard to the evidence of outcomes is patient-reported outcomes. You can talk about hospitalizations and whatever, but the impact on the patient is not reported, and those clinical outcomes are really the bottom line of what we're trying to assess today. And so across the board we're lacking patient-reported outcomes in any of these data that we're assessing today.

DR. GOODMAN: Dr. Janjan is emphasizing that we talk about health outcomes, and this is a message that we like to repeat to the Agency because we know that they're open to serious reports, it's not just the so-called traditional hard endpoint outcomes but patient-reported outcomes, and we know the Agency is aware of this.

Dr. Trikalinos, please remind us, at least for this instance, did you seek or did

you find information about the so-called patient-reported outcomes? I just want to make sure we've got our boundaries correct here.
DR. TRIKALINOS: So, outcomes beyond the three that I mentioned were not reviewed, so the answer is no, did not seek.
DR. GOODMAN: You did not seek those, okay. Possible that they may be there, you did not seek them and so did not turn them up.
DR. TRIKALINOS: Perfect.
DR. GOODMAN: Thank you for that point, Dr. Janjan. Is there anyone else among our presenters today that can tell us about evidence pertaining to the impact of this test on healthcare outcomes, i.e. the availability or sufficiency of it, or the actual evidence for impact itself, directly or indirectly, on this test? Okay.
Panel, if I see no further comments, if you don't mind, I will move to the next test.
DR. FISCHER: Did you want to vote on these as we go along?
DR. GOODMAN: No. We've got a chunk of time to vote in a little bit. What I'm

00193

suggesting is that it may be more helpful for us to kind of go through the body of evidence, it might help us to calibrate our votes later on if you kind of looked across here, if you don't mind. Is that okay with the panel? We will return to those questions, I promise.
Dr. Hayes?
DR. HAYES: May I ask Dr. Cox to maybe speak up again? Because it seems to me this assay doesn't really predict diarrhea, it predicts neutropenia. And so would you as a clinician, or do you as a clinician feel that the data are sufficient to use this to either use a different drug, or to just go ahead and use growth factors, and you would not give growth factors to a patient who's wild-type, for example.
DR. COX: I don't think the data is sufficient, nor in practice has the test held to be a utility that allows you to do that. And I think, again, for all the frustrations that I think many of the panel members, or at least I feel when we're talking about laboratory testing and trying to identify the impact that a test has in predicting a

00194

population of patients with all the other heterogeneity, and say because of the result of

that test you're going to behave differently and have a huge impact on a patient population, I think for this given test that evidence and the practice just isn't there.

DR. GOODMAN: Okay, thank you. So again before we leave, any other comments about evidence for the impact itself, okay? We talked about sufficiency and availability of evidence. Does anybody want to make any further comments about what the available evidence does say about the impact of the test and outcomes, any further comments on that? Okay. If you don't mind, we will move on, if that's okay.

The third test is HER2/neu for breast cancer patients who are candidates for trastuzumab. Let's do this again. So the first question is, let's consider the availability and sufficiency of the evidence. Now correct me if I'm wrong, this was the other test that was not covered in the technology assessment. Dr. Trikalinos is nodding his head yes, so this was not covered there. And where,

00195

was this evidence covered in the materials given to us as a panel? Yes, it was. Would any panelist care to take the first go at what the available evidence, about the sufficiency of the available evidence upon which to make some kind of finding later? Dr. Satya-Murti first and then Dr. Hayes. Dr. Satya-Murti.

DR. SATYA-MURTI: As I understand, I need education here. There are HER2/neu tests and other tests with HER, so my doubt has been, in reading up on all of this, it depends on what kind of HER2 test you're utilizing, and someone might be able to comment on that, because this, of the five we're looking at, this one is explicitly dependent upon the quality of the HER2 test and the expression.

DR. GOODMAN: We will address it in a moment. I just want to point out that the questions posed by CMS don't per se talk about individual tests, of course, and this was something we were thinking about this morning, what are they talking about here, about what test it is, and it may matter what test, especially for laboratory result tests.

00196

Dr. Mansfield is first.

DR. MANSFIELD: So, I would like to personally address that point. FDA first cleared the drug with the immunohistochemistry

test through, I believe there are two companies who have an approved one, and there are many laboratory-developed tests that have not been approved by FDA. This was followed on by approval of a FISH test for amplification of the HER2 locus. Subsequent practice has varied on whether it's believed that the IHC should be run first followed by the FISH, or whether FISH should be used as a decision point for patients who score in the intermediate range with the IHC tests. Some people are now using the FISH test only apparently, also at least one approved test, maybe two, and many laboratory-developed tests. I just wanted to put that in there for people's information.

DR. GOODMAN: Thank you, Dr. Mansfield. Other comments? Dr. Hayes.

DR. HAYES: Let me begin to establish some credentials on my part. About three years ago the American Society of Clinical Oncology and the College of American Pathologists joined

00197

together to form a panel to specifically address this issue. It is Wolf, et al., published simultaneously in the Journal Clinical Oncology and the American Journal of, whatever the College of American Pathologists' journal is. And I won't go through that in great detail or refer you to it, but we felt there was very strong evidence that patients who are positive, using classic definitions of positive, using either immunohistochemistry or FISH for amplification, are very likely to benefit from trastuzumab or lapatinib, which is a tyrosine kinase inhibitor, and those are both very active drugs, both in the metastatic and adjuvant setting.

There are modest data to suggest that patients who are less positive or negative won't benefit, and we believe the data for patients whose cancers are completely negative are fairly strong. It's the in between in which there is some concern, but right now at least the ASCO/CAP guidelines suggest only giving these two drugs to patients who are either three plus through immunohistochemistry, or a clearly amplified ratio greater than two

00198

by FISH. And so, that's just sort of a summary of what we might have heard this morning.

DR. GOODMAN: Other comments on this test? Yes, Dr. Kaul.

DR. KAUL: I think this is an area where we're going to continue to see some

evolution in what goes on in the laboratory. We have tests that are FDA-approved, we have some very reasonable tests out there, but we're going to continue to tweak what we do to make sure that the results are a bit more accurate. There is a gray zone and we're trying to get rid of that, and so there will be some evolution. That shouldn't be confused with the underlying target not being worth looking for and using clinically, I think.

DR. GOODMAN: But with regard to the available evidence currently, what is the panel seeing as far as the sufficiency of that evidence to make some finding about the impact of the tests on health outcomes? Dr. Kaul.

DR. KAUL: I think there's sufficient evidence to make a decision.

DR. GOODMAN: Okay. Any other comments about the sufficiency of the available

00199

evidence? Yes, Ms. Atkinson.

MS. ATKINSON: Is that evidence for the Medicare population?

DR. GOODMAN: And we are going to address that.

DR. HAYES: Yeah, I can address that. Part of the problem, of course, is that it comes mostly from clinical trials and most clinical trials address younger women, just for cultural reasons, not for any medical reasons. Older women just tend not to go in our trials. However, there's no reason to believe that either of these two agents, trastuzumab or lapatinib, are less effective in HER2 positive older women, and the reason not to give it to them might be because trastuzumab can cause heart failure so lapatinib might be substituted for that, or because you might feel that this patient is in a nursing home and is not worth treating for social reasons. But in terms of science and medicine, we believe it would apply to older women.

DR. GOODMAN: Any comments about the next part of the question, which is what does this available evidence tell us? It sounds

00200

like many of you concur that it is sufficient. What does the available evidence tell us about the impact of the test on health outcomes, would anyone like to summarize that for us? Dr. Hayes' hand is up first.

DR. HAYES: I'm sorry to keep bumping up, but this is something that I know about. We believe it has had a huge impact already.

In regards to, A, the amazing efficacy of these two drugs, both in the metastatic and particularly in the adjuvant setting over the last ten years. B, in terms of not giving these very expensive and potentially toxic drugs to patients for whom it appears there is very little benefit. So we believe that the health impact for this test has been quite large.

DR. GOODMAN: Thank you. Any other comments from the panel about what the evidence is telling us about the impact of the tests on health outcomes, anything to add to what we've heard so far? Any of our presenters want to add to anything that we've heard so far? I know that this is one that was not addressed in the technology assessment, so I want to make

00201

sure we've heard what we need to hear. Okay. I don't see further hands being raised. Good. If the panel doesn't mind, I want to hold off on (d), which is the BCR-ABL, until after discussion of KRAS if that's okay, because I know we've got more of a differentiation in the application of the tests for that one, though I do want to return to it. So let's move to KRAS. This is KRAS for metastatic colorectal cancer patients who are candidates for either cetuximab or panitumumab, and I know this is one of the three that was addressed in the technology assessment and we heard some other presenters comment on it as well. Let's address first about the availability and sufficiency of the evidence with regard to what the test might do for health outcomes. Anyone want to summarize what we heard, and/or ask our presenters what we heard about the sufficiency of the evidence? Dr. Teutsch.

DR. TEUTSCH: I heard we have really pretty nice evidence that shows that you could identify these patients who didn't respond, but I also heard we didn't get any information

00202

about potential harms or how it fit into alternative forms of therapy and the harms and benefits of those alternatives.

DR. GOODMAN: So those are some areas in which the evidence --

DR. TEUTSCH: At least we did not hear it.

DR. GOODMAN: We did not hear it. Dr. Trikalinos, if you don't, mind, can we prevail on you yet again, and it won't be the last

time, where was the evidence on KRAS? It sounded like there's some pretty good evidence for a certain aspect of this, but where was the evidence and where was it not?
DR. TRIKALINOS: So, as Dr. Teutsch summarized, there was consistent evidence pointed towards the same direction for mortality, progression-free survival and treatment failure.
DR. GOODMAN: Consistent, did you say?
DR. TRIKALINOS: Consistent. However, Dr. Teutsch correctly pointed out that there was no evidence that was presented that was weighing the tradeoffs between benefits and harms. This is because we did not find mention

00203

of harms of testing in the studies we reviewed, but I also made clear that we did not seek or evaluate any economic evaluations or decisional analysis for papers that would usually contain this type of information.
DR. GOODMAN: Thank you. Any of our panelists? Dr. Pao, on this point, the availability of evidence for KRAS?
DR. PAO: I just want to clarify. The harms of testing, you don't mean just doing the mutation testing? I will just remark that patients who receive cetuximab and panitumumab have considerable side effects, the majority have a rash, diarrhea, about 20 percent of the patients actually have a hypersensitivity reaction. And so you could, although the evidence is not presented, you would save patients who would not benefit from the drug from a lot of side effects, in addition to cost.
DR. GOODMAN: Okay. Dr. Hayes.
DR. HAYES: Sorry to speak up, but again, the American Society of Clinical Oncology has issued a statement on this, so I want to offer on that. We felt that the

00204

available evidence were quite strong and it's not just that it's a decreased likelihood of benefitting if you have mutated RAS. There is not a single study that I'm aware of that actually shows any benefit of having mutated RAS. And so this is a quite powerful predictive factor as far as we can see in terms of predicting no benefit. And I absolutely agree that these drugs are not benign, they do have side effects, and they're very expensive, believe me. So we felt that there was quite sufficient evidence to suggest that patients

who have mutated KRAS should not receive either of the two available antibodies. The data for a positive benefit if you're wild type are more mixed, I think, but they're sufficiently positive to suggest that there probably is a benefit in terms of the outcomes that you laid out in patients who are wild type. And so the ASCO recommendations are that everyone be tested. May I ask you a question, since I just spoke? Dr. Nowak spoke to assays, and that is a concern I think, and perhaps we can ask him if he might wish to discuss a little bit about

00205

the available assays.
DR. GOODMAN: That's fine as long as it's in the context of how good the available evidence is, if there's a differentiation between the assays that would bear upon that question, it would be relevant.
DR. HAYES: That's my point, whether or not there is variability within those assays that might have led to some of the heterogeneity in the data that we saw for patients who are considered wild type.
DR. GOODMAN: Dr. Nowak, would you care to respond in a concise way to that question?
DR. NOWAK: Testing for KRAS is not particularly difficult in a molecular lab, and there are a number of approaches to doing that. As with any of these assays, there are concerns about tumor heterogeneity or the sensitivity of a particular assay, and there's different ways to approach that. If you have a very sensitive assay then you don't have to be very critical about what your specimen is and what the tumor proportion is. On the other hand, if your assay is somewhat less sensitive but you do

00206

take account for how much specimen, how much tumor you're putting into the test, then you've essentially addressed that issue. And the standards for appropriate sensitivity are still in development but they're very much being discussed and I think being addressed. And just as ASCO/CAP put together guidelines for HER2 testing, and there are new guidelines coming out on testing for ER, there will very likely emerge guidelines for KRAS testing that will address those. As I pointed out, there is a proficiency test available now from CAP specifically for KRAS. It is a new proficiency test that has been

evolved. I think it's going to follow very much in the manner that HER2 proficiency tests evolved.
DR. GOODMAN: Any reason to believe that the quality of available evidence has been affected by any variation in the way that testing has been conducted?
DR. NOWAK: I suspect not. I think that the evidence, those papers that address the utility of KRAS testing have used adequate testing methods. I don't think the methods are

00207

in question there.
DR. GOODMAN: Dr. Hayes.
DR. HAYES: The reason I asked this was for a specific reason. We felt, from what Dr. Nowak was just telling us, that if you have positive mutation, that's almost certainly real, the odds of a false positive are very small, and there is no evidence that either of these drugs works in those patients from these retrospectively done studies on prospective trials. It is the negatives that are concerning, that is, it is more likely there are false negatives, and perhaps that's where proficiency would help. And that may reflect why some of the data are not, some of the trials are not positive even in the wild-type patients, and others are.
But we felt there were enough trials there that said that these drugs do work in wild-type patients to overcome the heterogeneity of the assays, so that was my point.
DR. GOODMAN: So in the wild type the drugs work sometimes?
DR. HAYES: I would say more than

00208

sometimes, but not always.
DR. GOODMAN: Dr. Nowak.
DR. NOWAK: There are other reasons why the drugs would not work in those patients with that wild-type KRAS. Probably the biggest cause is the mutation in another gene called BRAF, which probably accounts for another 10 percent of patients who will not respond to monoclonal anti-EGFR therapy.
DR. GOODMAN: Sounds like an area for further research.
DR. NOWAK: And that is being researched. You may be talking about BRAF next year.
DR. GOODMAN: Perhaps. Thank you, sir. Dr. Mansfield.

DR. MANSFIELD: Yeah. I was going to mention the lack of -- the very good positive predictive value of the test is likely due to mutations or differences downstream from KRAS, which are now starting to be studied. There is no currently FDA-approved KRAS test, although we understand that there is interest, and I do agree with Dr. Nowak that it's likely that the analytical validity of the tests that are used

00209

is not out of whack in a way that would mislead you.

DR. GOODMAN: Good, thank you, and thank you, Dr. Nowak.

On the matter of sufficiency of the evidence before you, KRAS sounds like a place where the panel considers there's pretty strong evidence in general, and the evidence for KRAS's impact on healthcare outcomes is derived largely from retrospective analyses of several RCTs. So not prospective experimental studies in RCTs, but we've used data from available RCTs and done retrospective subanalyses to make this distinction between the KRAS positives and the wild type.

Since this is an area with relatively stronger evidence, I wonder if any of the panelists would care to comment on the extent to which you would need more or seek more evidence in the form of prospective studies, prospective trials, RCTs or other, to confirm this, or are you largely satisfied that this is what you need to know? Dr. Janjan.

DR. JANJAN: I was just going to say that AHRQ has indicated that repurposing some

00210

of these studies is an acceptable way of evaluating because of the cost and time involved with prospective trials, that it's too inefficient to have to do everything in a prospective manner. And given the personal costs and the societal cost of not being able to determine what patients will respond, I mean, look at the range that we have here with the first question, the group with the question, if we could determine what patients should not get AI in therapy and instead get tamoxifen, look at the cost savings to society if we did that, versus being able to determine who should get HER2/neu therapy based on HER2/neu positivity.

So the range that we're seeing today I think shows that to wait another five or ten years to get this data out would not be

prudent, and I personally think retrospective analysis would be acceptable, and agree with the AHRQ.

DR. GOODMAN: Dr. Mansfield and then Dr. Teutsch. Dr. Mansfield.

DR. MANSFIELD: My understanding is that KRAS testing is already so firmly

00211

entrenched in clinical practice now that it would be virtually impossible to run a prospective trial and some people might suggest that it would be unethical, although I haven't examined that myself.

DR. GOODMAN: Thank you. Dr. Teutsch.

DR. TEUTSCH: I agree with the prior comments to this point, but I would think that rather than -- we're not making advice on coverage decisions, but this sounds like a time and place where while it would be appropriate to, you know, sanction and continue its use, you would like to get this prospective evidence on its clinical use and the impacts. I don't think that necessarily has to be an RCT, it could be a registry, it could be other things. Whether this falls under the, we talked about this earlier, coverage of evidence development or something like that, it might be something that's worth considering because it might confirm what we have from the retrospective information.

DR. GOODMAN: Thank you, Dr. Teutsch. Dr. Cox.

DR. COX: I'm going to struggle with

00212

this a little bit, and being more of a trialist to begin with, I think when you look at, again, I'm struggling with we're looking at basically whether this test helps us to find basically one of the drugs in our armamentarium and provides guidance on how to best take all the other tools and create a treatment plan with the best outcomes. So when you talk about the utility, and I as a clinician am still going to be looking for prospective evidence that will help me understand how to take the tools I have, including these EGFR antibodies, to figure out how to stick them into a treatment regimen. I mean, metastatic colon cancer has gone from six months median survivorship, eight months, to now over two years, largely because of a cascade of therapies. So this one test does affect clinical outcomes of patients, but I'm reacting on the thought of being able to mine prior studies to really help me. I'm sure

there are some answers that can be had, but I think you're still going to have to do prospective studies to figure out how to use it.

DR. GOODMAN: Well, it sounds like you

00213

like the strength of the evidence, especially relative to other types of tests, but you may want some prospective data collection, not necessarily an RCT, that might strengthen your observations and how you might use the test itself. Is that correct, Dr. Cox?

DR. COX: Again, I struggle when talking about a test that has an impact on the choice of therapy. If that test is going to help describe the choice of therapy you still need to have prospective data on the therapies you're choosing.

DR. GOODMAN: On the KRAS test then, we've talked about the sufficiency of evidence. Would anyone care to comment on or summarize a finding with regard to what the evidence does tell us about the impact on patient outcomes? Some of you have already suggested that, but would any panelist like to iterate that, please? Yes, Dr. Fischer.

DR. FISCHER: It looks like you're going to make about 20 percent of the patient population miserable with absolutely no benefit, from the data that I reviewed before coming here, and I think that's important.

00214

DR. GOODMAN: If you --

DR. FISCHER: If you give some of the agents which are not innocuous to a group of people who are KRAS mutated, I don't think I saw any evidence at all that there was any benefit at all, and yet the side effects are significant.

DR. GOODMAN: Okay. Any other comments on what the evidence can tell us about the impact on patient outcome? Any other comments about that? Our speakers, is there anything we missed that's important about the impact of this test on health outcomes? Okay. Dr. Pao, and then we'll move on.

DR. PAO: I would just second the thoughts of Dr. Mansfield. It would be very difficult to do a randomized controlled trial now in a KRAS-positive patient population. Randomizing is tough enough, and I don't know how many patients would actually sign up or accept the test randomizing.

DR. GOODMAN: Yeah, I think that point

was well made, but it did sound like some other kind of prospective data collection might be okay, nonrandomized. Yes, Dr. Hayes?

00215

DR. HAYES: A very quick clarification for our colleagues who are not medical oncologists. To my knowledge, KRAS has no impact on whether the chemotherapies that we routinely use work or don't work, and so to my knowledge KRAS mutations are very specific to the efficacy of these two antibodies directed against EGFR. So it's not that we're saying I'm sorry, Mr. Jones, but you can't have any therapy if you're mutated. It's I'm sorry, but these two antibodies won't work for you, but the other therapies are equally likely to work on you as anyone else. So there are many options for patients with colon cancer and increasingly more so, it's just that these two antibodies against EGFR seem not to work in patients who are mutated.

DR. GOODMAN: Well summarized, thank you.

Let's return, then, having done tests (a), (b), (c) and (e), let's return to test (d), which is the BCR-ABL. And I know we discussed a bit this morning about how the interest to the evidence questions depends a little bit on how the test is being used. And

00216

this question is, with regard to the sufficiency of the available evidence, on BCR-ABL for chronic myelogenous leukemia patients who are candidates for imatinib. So we're going to talk about the sufficiency of available evidence. And as we discuss this, I want to go back to Dr. Kaul, if that's okay with her. Dr. Kaul, can you remind us about what differentiation we need to consider with regard to how the test is used?

DR. KAUL: There are, I see it as two separate tests, you can even pigeonhole them further if you want to get more detailed, but the two tests that I think we could break it down more simply are identifying and quantifying the fusion transcript that defines the BCR-ABL translocation.

DR. GOODMAN: Say that one more time.

DR. KAUL: It's identifying and quantifying the fusion transcript that is associated with the BCR-ABL translocation, and that is tantamount to making a diagnosis of CML and it's also the target, and has been for probably a decade, the cornerstone of

monitoring patients on chemotherapy. You
00217
monitor the levels of this transcript and see how they're responding to treatment.
DR. GOODMAN: Dr. Kaul, for shorthand purposes, so you're --
DR. KAUL: Following response to treatment by looking at the tumor burden going up and down.
DR. GOODMAN: Do we want to call that for our group's sake the fusion transcript or the diagnostic application, what's our short term for that?
DR. KAUL: It's diagnosis and monitoring.
DR. GOODMAN: Diagnosis and monitoring, so that's one.
DR. KAUL: And then the second area I see partly is based on the technology or assays, we need to study this, but we heard about this morning very nicely, and that's something that's quite new and I think in its infancy, and that is identifying point mutations that occur in this transcript that are associated with a failure of response to treatment, and that's being investigated in newly diagnosed patients to see if they should
00218
even go on one of these treatments in the first place.
It's also being used in some centers where they're having their standard consensus-driven, every-three-month transcript levels, they're piggybacking onto that another test to see if they can detect any of these point mutations that might herald impending treatment failure. But this is a totally different test, different purpose, different technology, and I think is much more immature than the not yet mature, but the very well established transcript level monitoring that we've been doing for many years.
DR. GOODMAN: Dr. Kaul, I'll stick with you if you don't mind. With regard to diagnosis and monitoring versus point mutation of the test, do you -- and we'll start this discussion with some others -- do you think that we as a panel should address both of those uses of the test insofar as they might affect health outcomes or just one?
DR. KAUL: No, I think that the diagnosis and monitoring part is mainstream medical practice now, so I think a discussion
00219

can be had a little bit after the fact because this is used widely, even more so than RAS, the previous example, and I don't think that these patients can be managed without that, so the impact there I think is quite clear and well established.

I think we can have a more interesting and immature discussion about the point mutations if we so choose.

DR. GOODMAN: Does anyone on the panel want to vouch for doing just one of these, or are you satisfied that we might address both of them? Dr. Satya-Murti.

DR. SATYA-MURTI: If there was that much difference, I think we ought to take them up separately. You summarized this so well, and then we heard also this morning, and then there were several presenters who were talking about the load of the transcript rather than the mutation. So I would find it's easier probably to split the two.

DR. GOODMAN: I see no objection to splitting them. Let's take diagnosis and monitoring first then. Let's talk about the adequacy of available evidence regarding the

00220

impact of the use of the test for diagnosis and monitoring on patient outcomes, how good is the available evidence, and then we will move to what does that evidence say. Dr. Kaul, you're up.

DR. KAUL: We have dozens of peer reviewed articles, we've got many multi-centered clinical trials, and we've got consensus guidelines from a couple of professional organizations at least outlining the use of these quantitative assays in monitoring patients and also leading to diagnosis, so I think there's very clear evidence.

DR. GOODMAN: So you alluded to a lot of evidence, but then part of it said multi-centered trials, I think I heard you say that, which sounds like better evidence.

DR. KAUL: And it's been published in peer reviewed literature. It's already out there.

DR. GOODMAN: Okay. So you're saying that the body of evidence is rather robust?

DR. KAUL: Yes.

DR. GOODMAN: Okay. Dr. Eng and then

00221

Dr. Fischer.

DR. ENG: May I ask, is the evidence

so robust that, say for an older patient that doesn't want bone marrow biopsy or treatment for remission, is the evidence, this monitoring so robust, so good, that you can say that's okay, if your blood test is less, you know, we don't have to go to the bone marrow? Because one of the papers said that it's still sort of like the gold standard, to get that bone marrow.

DR. KAUL: I'm not a medical oncologist taking care of leukemia patients, but the bone marrow will allow you to get a full karyotype. As patients evolve, they will develop other chromosomal abnormalities that you can't necessarily get without doing the full karyotype, and so that may still be of use. You can at least detect and measure the transcript level in the peripheral blood, so that can avoid use of bone marrow in some settings.

DR. GOODMAN: Dr. Fischer.

DR. FISCHER: I am under the impression that we were splitting, that we were

00222

not using the diagnosis.

DR. GOODMAN: No, we're going to get two looks, one is diagnosis and monitoring and the other is point mutation.

DR. FISCHER: I understand, but that's what got lumped just recently. In other words, I understand that people use this for diagnosis and I understand that people use this for recurrence, or what I would call recurrence but other people might call monitoring. Are we going to take those together, or are we going to take those separately?

DR. GOODMAN: I suggest we take those two together, diagnosis and monitoring as a bolus, and the other would be point mutation. Dr. Pao and then Dr. Hayes. Dr. Pao.

DR. PAO: Going back to Dr. Eng's question, I think that after the first bone marrow the molecular test is much much more sensitive and it gives you a much better idea of the tumor and allows you not to repeat the marrow.

DR. GOODMAN: Dr. Eng.

DR. ENG: My question was really the strength of the molecular test to practice, and

00223

actually that is a reflection. If it is that strong, then we wouldn't have to have any more bone marrows. We have five.

DR. GOODMAN: Dr. Pao.

DR. PAO: I do believe if you had changes in your quantitative transcripts in your blood vessels that the doctors would make some treatment decisions based upon that. I don't know if you would actually repeat the bone marrow. That's what I'm saying, that the bone marrow would not be as sensitive as the molecular assay?

DR. GOODMAN: Dr. Eng, is that sufficient at least for now? Dr. Hayes.

DR. HAYES: So, I also do not treat leukemia patients, but my understanding is that this doesn't completely abrogate bone marrows but it decreases the number of bone marrows that are done because if a patient is negative as she's being monitored, they don't do serial bone marrows, they wait until it comes back up. I actually would like to perhaps ask Dr. Salvado from Novartis, one of the questions I asked him during the break was, I think the clinical validity of these assays both for

00224

diagnosis, is this a classic or an atypical CML, and for monitoring, are quite strong. In other words, this assay tells us that is the case, that the patient does have classic CML and that a patient you felt was doing well is starting to progress. The issue is, does it help you make, is it of clinical utility, does it help you make a decision that helps the patient by using those data, and the question really is, does the next therapy work, so it's worthwhile deciding if they are progressing. The answer he gave me is yes, that when imatinib quits working by virtue of the rising transcript level, that the next generation of drugs do work, and so there is clinical value in identifying that. And while it's hard to prove it, it is probably that they are more effective when it is at a low tumor burden than if one waits until you have circulating CML cells, although again, that's not been proven to my knowledge. And then finally, when those quit working, although this may not be relevant to the Medicare population, bone marrow transplant has been shown to result in prolonged

00225

disease-free survival and even cure rates in younger patients with this disease, and it is much more likely to be effective before a patient goes into blast phase.

DR. GOODMAN: Dr. Hayes, let me just interrupt. Do return to this point, though.

DR. HAYES: Yes.
DR. GOODMAN: For diagnosis and monitoring, how strong is the available evidence with regard to impact on outcomes?
DR. HAYES: That's where I'm going. I think that the data suggests that it is pretty strong.
DR. GOODMAN: Thank you for that.
Dr. Fischer.
DR. FISCHER: Again, I don't treat patients with CML but I happen unfortunately to have a number of friends who have CML, who've had it for a long time, and many of them, when the diagnosis is originally made, choose not to be treated, and they go on for quite a long period of time. One of them is 25 years since the original diagnosis and he still goes skiing and doing well, at 85. So my question is, how accurate is the test that we are proposing to

00226

use as diagnosis? I think Dan, I think you referred to this somewhat, that after the diagnosis is made and the patient is being followed, maybe had a bone marrow and the patient is being followed. Is this something that one can use to follow a patient who chooses not to be treated and say it's time for you to be treated, which I think in this particular disease is probably an important part of diagnosis.
DR. GOODMAN: I'm still looking for strength of available evidence here, so make sure it goes back there. Any response on part of the panel to Dr. Fischer's point? Dr. Hayes.
DR. HAYES: To my knowledge it would not be of value because those patients are already going to be screened as positive for the transcript level because they've got circulating cells you can see, so I don't think it would help.
DR. GOODMAN: Dr. Hayes, as long as you've got the mic already in your hand, let's move to, summarize for us, then, what the available evidence does tell us about the

00227

impact of diagnosis and monitoring with this test on health outcomes. What does it say?
DR. HAYES: So, I believe the data are pretty strong to suggest that, A, initial diagnosis, but probably more importantly, monitoring these people will have a huge impact in regards to staying with the therapy begun or switching to another therapy that's likely to

be effective.
DR. GOODMAN: Now, hold on. When you're talking about staying with the therapy or not, are you implying that there's a strong link between that and health outcomes?
DR. HAYES: Yes.
DR. GOODMAN: Okay. Does anybody have any comments about that, anything they want to contest about that comment or conclusion? Dr. Satya-Murti.
DR. SATYA-MURTI: Not so much to contest this, but we didn't come prepared to hear this this morning, but it seems after listening to you both, it appears to me that this is a perfect surrogate model, even better than HIV or the hepatitis C virus model. This transcript model not only diagnoses, it tells
00228
you what's out around the corner, so this is a very good test. As I say, it keeps occurring to me as a non-oncologist, it's a perfect surrogate for this disease.
DR. GOODMAN: Thank you. Let's look now at the point mutation application of this test, again, BCR-ABL for chronic myelogenous leukemia. For point mutations, how good is the available evidence, what's the sufficiency of the available evidence upon which you might draw some conclusion or finding about its impact? Dr. Kaul, if you don't mind, I'll ask you to take the microphone.
DR. KAUL: Well, I actually was surprised when the technology assessment addressed this in such detail. It's an area that people are investigating actively, and I think that's wise, but I don't think we have enough evidence to support using this clinically routinely at this point. Evidence is still lacking.
DR. GOODMAN: The evidence is lacking. Would anybody have an alternate view on that? Dr. Pao.
DR. PAO: Well, I can't give any data
00229
on health outcomes, but as Dr. Trikalinos said, the one mutation that does make a difference is T315I, which as we've heard (inaudible). Therefore, testing for that could make you eligible for a subsequent new trial if they're trying to target T315I.
DR. GOODMAN: Okay.
DR. PAO: But that might be outside of the range of this discussion.
DR. GOODMAN: Thank you, but did you

care to comment on the sufficiency of the evidence of that report?
DR. PAO: As I say, I don't have specific evidence for that at this time.
DR. GOODMAN: Dr. Kaul.
DR. KAUL: I'll just follow by agreeing. I think what we're learning about is the biology of the disease and how resistance occurs and that's going to be evolving, so in another year or two it may be a very different story, but we're still pretty early in that process.
DR. GOODMAN: Any of our speakers care to comment on the sufficiency of available evidence with regard to drawing a conclusion

00230

about its impacts on outcomes? Is this Dr. Burken approaching the microphone?
DR. BURKEN: This is Dr. Burken. As my slides indicated this morning, my own review that was teed up just for this meeting today indicated that point mutation testing, if the way you're approaching it is to look for a panel of mutations that might be an optimal panel of mutations, just that particular question, I found that it was not ready for prime time, and that would seem to be in agreement with what Dr. Kaul has told us.
DR. GOODMAN: Thank you. Anyone on the panel, or Dr. Burken, before you leave, want to say anything now about what the available evidence might say about impact on outcomes? It sounds as though there's agreement that the evidence is pretty scarce right now. Yes, Dr. Scheuner.
DR. SCHEUNER: So, I'm going to go back to the T315I mutation and somewhere to the KRAS discussion. It seems to me that that particular point mutation and all the studies that were in the TEC assessment, there were

00231

multiple, showed that that particular mutation, you would not respond to therapy so you wouldn't want to give that drug. So, I don't know if we want to hear more about that from any of the speakers.
DR. GOODMAN: So you're saying that there is an instance where point mutation does provide definitive evidence, is that what you're saying?
DR. SCHEUNER: Yes. We saw a slide where he looked at a set of genetic response and the specificity was a hundred percent, and

so maybe we could hear a little bit more about that, as opposed to any mutation, then yeah, it's right along the diagonal of the ROC curve, so I believe that one mutation actually is very predictive of response to therapy. You want response.
DR. GOODMAN: Thank you, Dr. Scheuner. Dr. Trikalinos.
DR. TRIKALINOS: So, this observation is correct. All studies are consistent, essentially, that when this mutation is present, then there is no response to the drugs that were assessed.

00232

DR. GOODMAN: And what mutation is that? We're calling it what?
DR. TRIKALINOS: T315I.
DR. GOODMAN: That's the T315I.
DR. TRIKALINOS: And this is something that is not new and this is something that is known to the community of researchers who are treating this disease. And the key point here is that this is a mutation that is rare, it's a small percentage of people who have not responded to the first line treatment that have this mutation.
DR. GOODMAN: Okay. Dr. Scheuner, you're satisfied with that?
DR. SCHEUNER: I just had one follow-up to that. So, is it rare enough that there are concerns about the analytic validity of the assay to identify the mutation and claim perhaps a false positive? I mean the concern, you say it's rare, but if it costs five cents to do the assay, then maybe it would benefit that handful of people for whom the drug is just simply not going to work.
DR. TRIKALINOS: I have no knowledge of whether there's an issue with analytical

00233

validity for this particular mutation so I cannot answer, I cannot give you an answer. However, these are not the only considerations, the cost.
DR. GOODMAN: Thank you. Was there a comment, was it Dr. Hayes briefly, and then Dr. Pao.
DR. HAYES: So, again, I am not a hematologist, and I'm reading now from the American Society of Hematology comments which they submitted, representing hematologists, and I won't read you the whole thing.
DR. GOODMAN: Okay. What comments are they? Are they evidence-based comments that

address the evidence?
DR. HAYES: So, what they say is that the T315I mutation is not sensitive in vitro to any of the available agents. Patients with the T315I mutation indeed have no response to any available TKI. These patients should be offered a stem cell transplant when eligible. So I actually am neutral on this, except to say that I think it's fair to hear what the American Society of Hematology has suggested to us, but they don't provide a

00234

jumble, they provide a single reference to that comment.
DR. GOODMAN: They do or don't provide a single reference?
DR. HAYES: They provide a single reference. There may be more, but they provide Leukemia, Updated Concepts of Management, that's a review in the Journal of Clinical Oncology this year.
DR. GOODMAN: Okay. That covers it for BCR-ABL.
So we've looked at all five of these tests and we've asked about the sufficiency of available evidence, and then we've talked about where applicable, where that available evidence might lead us with regard to impact on outcomes. Let me ask the pleasure of the panel here, and I'll just kind of offer you a choice. At the rate we're going, I'm confident that we will be done by four o'clock, so that will help assure you. How would you feel about moving directly into the voting for questions one and two now, or would you like to take a 7.5-minute break, and then return to voting? Vote? Okay. So what we're going to do now

00235

is we're going to vote on questions one and two, having been very nicely informed with the discussions that you've given us over the last nearly hour and a half. Then we'll look separately at question three, which is going to address the matter of the generalizability, so we will address that separately after we address these first two questions.
And I know that at this juncture we have to ensure that our CMS staff colleagues are ready to do our voting and I will just give them a moment. And while we're doing that, just as a reminder, when we do get to the BCR-ABL, we are going to split it into those two sections, one of which was diagnosis and monitoring, the other one of which was point

mutations. I just want to remind the panel and our audience, and with confirmation from Dr. Jacques and Maria Ellis, could you remind us about the nonvoting members and how that works?
MS. ELLIS: Okay. Everyone will vote, so we will have two separate scores, one with all the members, all the panel, and then the other score will be just the voting members, so

00236

that's how that goes, there will be two different scores, two different sets of scores.
DR. GOODMAN: But you'll only record the scores one time, and when you record them you'll differentiate internally between the voting and nonvoting members.
MS. ELLIS: Yes. And also, there are voting sheets in your packet on the left-hand side for you to also record your votes just in case, so I can double check with what you have and what I'm putting in the system right now. They should be on the left side of your green folder, on the left side in the back.
DR. GOODMAN: Okay. Does everyone have the current voting sheet?
MS. ELLIS: Does everyone see it, it's in the green folder.
DR. GOODMAN: Dr. Pao.
DR. PAO: How would you like us to address the split in the BCR-ABL question?
DR. GOODMAN: When we get there I'll break it into two parts.
MS. ELLIS: And if you could just write at the bottom or somewhere beside it, you can just put Dx and your score, and then point

00237

and your score. Does everyone have the score sheet?
DR. GOODMAN: Does anyone not have the score sheet in front of them?
MS. ELLIS: They should be, again, in the green folder on the left-hand side. Thank you.
DR. GOODMAN: I presume that no one has any further very important comments before we get into the voting. Dr. Matuszewski.
DR. MATUSZEWSKI: Are decimal places allowed in terms of voting?
DR. GOODMAN: Well, actually not for voting purposes, Dr. Matuszewski, but when we determine between questions one and two, a mean score of 2.5 or greater qualifies a question one test for question two, a score below 2.5 disqualifies discussion of a question one test

for question two.
DR. MATUSZEWSKI: That's why you need a .5, and I have one made up here.
(Laughter.)
DR. GOODMAN: Dr. Matuszewski, thank you for that. We needed some of that levity, that's great, we will take it where we can get
00238
it. Okay.
Let's start with question one. Just a reminder, question one is about sufficiency of evidence, not about what the evidence says. Question two is about what the evidence says in the cases for tests that have sufficient evidence.
So question one asks, how confident are you that there is sufficient evidence to determine whether pharmacogenomic testing affects health outcomes, including benefits and harms, for patients with cancer whose anticancer treatment strategy is guided by the results of testing as described below:
And the first test is CYP2D6 for breast cancer patients who are candidates for tamoxifen. A scale of one to five where one is the weakest, five is the strongest, sufficiency of evidence.
(The panel voted and votes were recorded by staff.)
DR. GOODMAN: Thank you. The same question regarding sufficiency of evidence for tests impact on health outcomes, this time for UGT1A1 for colon cancer patients who are
00239
candidates for irinotecan, sufficiency of evidence.
(The panel voted and votes were recorded by staff.)
DR. GOODMAN: Next is HER2/neu for breast cancer patients who are candidates for trastuzumab, HER2/neu breast cancer patients, sufficiency of evidence with regard to impact on health outcomes, one least confident, five most confident.
(The panel voted and votes were recorded by staff.)
DR. GOODMAN: Next is one we're going to break into two parts, and this is for BCR-ABL for CML, patients who are candidates for imatinib, and the first one regards the sufficiency of evidence for tests impacting health outcomes where the use of the test is for diagnosis and monitoring as we had described that earlier this afternoon. So for

diagnosis and monitoring, that application of this test, BCR-ABL for CML, do you have low confidence or high confidence, along a scale of one to five?
(The panel voted and votes were

00240

recorded by staff.)
DR. GOODMAN: And now the next, again for BCR-ABL, is with regard to the point mutations, and I will just say that if there's at least one point mutation for which you think that there is sufficient evidence, you can use that, and if we need to have further discussion about that later, that's fine. But the point mutation application of the test, one is low confidence, through five, high confidence.
(The panel voted and votes were recorded by staff.)
DR. GOODMAN: And the last of the five, again, is KRAS, KRAS for metastatic colorectal cancer patients who are candidates for cetuximab and/or panitumumab. How confident are you with regard to the sufficiency of the evidence for impact on health outcomes where one is low and five is high confidence?
(The panel voted and votes were recorded by staff.)
DR. GOODMAN: And now Ms. Ellis is going to tell us which of those five, actually six, because (d) has two parts, for which of

00241

those was the mean score greater than or equal to 2.5, and for those we will address question two.
MS. ELLIS: We have (c), and we have (d)(1), and (e).
DR. GOODMAN: So among the six total, including two for (d), the only ones that achieved a score of 2.5 or greater were (c), HER2/neu, (d), BCR-ABL for the use of diagnosis and monitoring, and (e), KRAS; is that correct?
MS. ELLIS: The only ones that had more than 2.5 were (c), (d)(1), and (e).
DR. GOODMAN: Okay. We confirmed that, okay.
MS. ELLIS: And that is for just the voting members.
DR. GOODMAN: Correct, for the voting members. Although we're recording votes, all the votes will kick in and were recorded separately.
MS. ELLIS: Correct.
DR. GOODMAN: Okay. So the first

question we'll look at here for number two is going to have to do with the impact itself, and we're going to start with HER2/neu, but I'll

00242

read you the question.
For those items where the answer to question one was at least in the intermediate range, which is a mean score of 2.5 or greater, how confident are you that pharmacogenetic testing improves health outcomes for patients with cancer whose anticancer treatment strategy is guided by the results of testing as described below:
And for the first it's HER2/neu for breast cancer patients who are candidates for trastuzumab. How confident are you that the test improves health outcomes?
(The panel voted and votes were recorded by staff.)
DR. GOODMAN: We will move to BCR-ABL now, and do recall that this is about the first of those two uses, and this is going to be for diagnosis and monitoring. So how confident are you that this test, BCR-ABL with the diagnosis and monitoring application, improves health outcomes for patients with CML who are candidates for imatinib?
(The panel voted and votes were recorded by staff.)

00243

DR. GOODMAN: We're going to move to KRAS now, KRAS on a scale of one to five. How confident are you that this test, the KRAS test improves health outcomes, and this is KRAS for metastatic colorectal cancer for patients who are candidates for cetuximab and/or panitumumab? Impact of the test on improving health outcomes.
(The panel voted and votes were recorded by staff.)
DR. GOODMAN: So Ms. Ellis, you've got answers to questions one and two completely now, I believe; is that correct?
MS. ELLIS: Yes.
DR. GOODMAN: Okay, very good. Let's move to question three, and I'm going to turn to Dr. Jacques just for a moment for clarification. Question three asks about the confidence of the panel regarding whether these conclusions are generalizable to, A, community-based settings, and B, the Medicare beneficiary population. Dr. Jacques, CMS did not break this out by test. We would be glad to break it out by test if you would like, or what is your

preference?
00244
DR. JACQUES: Our preference is actually that you not have to break this all out by test unless you want to be here well beyond your flights, I think it might take that long. What our sense is of that question, and it's a recurring question in every MedCAC, unless you believe that there is some reason why you can't make a somewhat general statement about the applicability of the evidence to essentially an older population, we would like you to sort of address it in toto.
If you believe for some reason that one of these tests, for some reason there's a red flag going up saying this one should be treated differently, you have the option to do that if you would like.
DR. GOODMAN: Okay, thanks.
Dr. Atkinson, did you want to make a comment?
MS. ATKINSON: Dr. Eng had asked about the mean age of those studies; can we have the answer to that question first?
DR. GOODMAN: Thank you. Dr. Trikalinos, please do.
DR. TRIKALINOS: As a reminder, we were asking about mortality and KRAS and as you
00245
remember, there were nine studies that were significant in terms of hazard ratios or ratio for mortality, and the answer is the median age was above, bigger or equal to 60 in seven of them. In two of them we don't have the reporting of the median age. And also, in none of these, it was above 65. So mean ages for these nine studies, two are not reported and the remaining are between 60 and 65, or 64, let's say.
DR. GOODMAN: Dr. Eng, is that helpful?
DR. ENG: Yes.
DR. GOODMAN: Dr. Satya-Murti.
DR. SATYA-MURTI: When we say community-based settings, do we mean those who are without a university affiliation? Because many of these tests are referenced out anyway, and do any oncologists want to comment on that?
DR. COX: I think my take of 80 percent of the patients who are treated with cancer in this country are treated in community-based centers, it could be institutional-based but not academic centers. So we may want to talk about this, but I would
00246

say that the conclusions are how these tests are used in hospitals in a community-based practice.
DR. HAYES: This is a very real concern for the joint ASCO/CAP guidelines committee, and the first thing we found was a modestly scandalous heterogeneity of how HER2 was done, and the CAP has taken that under their wing and built in proficiency testing. As you saw, there are well over a thousand centers participating in that, which I think is very encouraging. So in my opinion, it's applicable in a community setting as long as the people in the community, whether a university or a private hospital, pay attention to details and do the assay correctly. I don't think there's anything specific to a university versus a non-university setting for providing the assay.
DR. SATYA-MURTI: You are going to be referring out most of these tests, Dr. Cox, isn't that correct? In your own practice, if you need an ABL monitoring, BCR-ABL, would you be doing it in your own laboratory or would you be sending it out?

00247

DR. COX: Maybe I would redirect that question to Dr. Nowak, who just presented data from the CAP survey. In our institution, a community-based institution, we send out all of these studies.
DR. GOODMAN: Dr. Nowak, did you want to comment on that, address this particular question on community?
DR. NOWAK: It depends on the assay. BCR-ABL is technically complex to do it well and that's why there are relatively fewer laboratories that do that. One needs to, if you're going to do it properly, one should have a sufficient volume to establish a laboratory-determined baseline for your patients, and in the community setting it's unlikely that a laboratory would ever establish that baseline in any reasonable time, so it's better that that test be sent out to a center that has sufficient volume and proficiency in doing it. Tests like HER2 that are done by multiple methods that are accessible to many laboratories, immunohistochemistry more so than FISH, those are done, should be done, are done in community settings.

00248

What's the third assay? KRAS. KRAS is a molecular test so it's less likely to be

done in a small community setting, but there are many large community hospitals now that have a molecular testing capability, and KRAS testing will be within their capability.
DR. GOODMAN: Thank you. Let me just go to Dr. Scheuner first.
DR. SCHEUNER: So, this is kind of related to the discussion about, as I understand and I actually have some funding from CDC to look at this, that most errors in genetic testing occur in the pre-analytic and the post-analytic phases of testing, not so much with what's happening in the lab. But for example, once a report is received by a clinician, do they truly understand what that report means. So that's some of the funding that I have at RAND, is trying to develop a model genetic test report that clinicians understand. So I do have concerns about the implementation of this in the community setting where the recipients of a lab report may not really understand what that lab report is saying, and I don't know that we have a lot of

00249

evidence about that.
DR. NOWAK: Certainly pre-analytical and post-analytical elements -- I mean, most areas related to laboratory results are clerical errors, they occur before or after the actual testing, but the analysis is usually pretty good. Certainly CAP and the ANP are very concerned about test interpretation and the importance of having qualified individuals interpret those tests for the clinician, so your concerns are appropriate and shared by others.
DR. GOODMAN: Dr. Janjan?
DR. JANJAN: Well, this goes to level of experience of the community physicians. I mean certainly with regard to CML, that's been around for a long time and that's part of training, and anybody getting out of medical oncology residency would know how to apply that within clinical practice. Some of these more recent things, they may or may not input that into their clinical decision-making.
So, you know, I don't want to, I'm not suggesting we break these out according to the different tests, but on the other hand, I think

00250

there will be some heterogeneity and maybe John could talk to us more about when the fulcrum of a test comes out, there's more data, and when do you say well, that's enough data, that I'm

going to now incorporate it into my clinical decision-making.
DR. GOODMAN: Okay. Do you want to comment quickly, Dr. Cox?
DR. COX: It is, I mean, a translation of what we learn in our science, the practice is one of the things that bedevils all of our professions, and certainly when it comes to diagnostic studies.
The only comment I would make is it's often, I would say it's pretty dependent on the strength of the evidence. Was it about four years ago that ASCO commented about HER2 in an adjuvant setting? I would say nearly all oncologists adopted that in June after it was presented in May, because of the overwhelming strength of the evidence.
Whereas you look at two of the tests we talked about, the CYP2D6 and the UGT1A1, that evidence has just languished, and I think you're right. If you were to go to a community

00251

oncologist and ask him how he utilized this in practice, you would get a wide variety of not knowing what to do with this data. But to me that brings some truth about the strength of evidence.
DR. GOODMAN: Thank you, Dr. Cox.
Dr. Nowak.
DR. NOWAK: I think you have to distinguish between clinical utility and if the evidence is there that this is useful and should be done, that applies to oncologists at big medical centers as well as oncologists at smaller places, they should be offering those opportunities for testing and for treatment to their patients. You need to distinguish that from the quality of testing, and that's a real issue, but that's a different issue, and that's being addressed. But it shouldn't influence the strength of the evidence as to whether it is appropriate to test and is appropriate to be used in a certain kind of therapy, that should be uniform for everybody.
DR. GOODMAN: Fair enough. We do care, however, about the extent to which observations about impact of the test on

00252

outcomes do apply outside the ideal settings, accounting for many of those intervening factors. The community setting is typically different than the ideal setting, and we want to know what the applicability of the evidence is, the extent to which you can cross that

border, so it can be relevant.
So, let's do this. Oh, excuse me, Dr. Fischer.
DR. FISCHER: This question is addressed to Dr. Jacques. First of all, as somebody who has practiced in an academic medical center all my life, but has some relationships with nonacademic medical centers as a chair who sends residents out to nonacademic medical centers, I would venture to say that the variability in academic medical centers is larger than we would say at the beginning, and that sometimes the quality that we see in a big private hospital where we send a lot of residents is better, point number one.
Point number two, I know I'll probably get struck by lightning, but that's all right, point number two, from a practical point of view, Dr. Jacques, could we say this is great

00253

stuff, but you can do it, you can't do that, is that appropriate?
DR. GOODMAN: Dr. Jacques.
DR. JACQUES: If the panel believes that the evidence supports a recommendation from the panel that this particular technology is either nuanced enough in its science, complex enough in its implementation, or whatever, that the evidence of benefit is restricted to certain types of situations, whether that's the person doing it, the health system within which it is being done or something else like that, the committee is certainly free to make that recommendation to us.
As I mentioned a bit earlier, we don't have an open national coverage determination on this, so it's not like you're going to say well, gee, only these people should do it, and then tomorrow there's suddenly going to be some Medicare policy that says only certain people can do this. So the question really is about the evidence, so if you believe that based on what you know from the evidence about how these things are done and how practice happens in the

00254

community, at issue here, as Dr. Goodman pointed out, is not simply the test being performed in a referral center versus a community, but the whole chain, including the actuation of meaningful data, is also part of the community.
So whereas maybe by history one might have assumed that in an academic medical

center, and I also worked in one, that gee, if the oncologist and the molecular geneticist or someone else want to have a conversation if there's a nuance, gee, they'll run into each other in the hallway or whatever, and they will have that conversation, versus the busy community practitioner who may have to take time out, may have to get a phone call, may or may not get a response, et cetera, gee, it may never happen. Certainly those assumptions I think are common among physicians; whether they are true or not is arguable.

DR. GOODMAN: Okay. Briefly, Dr. Hayes, and then I want to kind of move to a vote on this.

DR. HAYES: Only to say that rather than distinguish academic versus community,

00255

would be accredited versus nonaccredited, and again, CAP has a really lovely accreditation system now. One would hope that perhaps funding agencies might say if you're not accredited to do this test, we're not going to pay you to do it. In that case the market will take care of itself. The places that don't want to become accredited because it takes too much time to do with relatively infrequent tests won't do it anymore, and those that do will. It won't matter if it's a large community hospital or a large academic hospital.

And the same thing's true for treatment in my opinion, you know, small hospitals won't do bone marrow transplants because it's too much trouble putting them together. Large hospitals, academic or not, will, because they see enough patients to do it, so I think that's really the filter.

DR. GOODMAN: Thanks, Dr. Hayes. Dr. Satya-Murti.

DR. SATYA-MURTI: A very brief question. None of the three tests we're looking at now and voting on is a CLIA-based

00256

test, is it not?

DR. GOODMAN: These are laboratory developed tests subject to CLIA, correct, Dr. Nowak?

DR. NOWAK: That is correct, and so they may not be among the small number of tests that are specifically mentioned in CLIA as being reportable. CAP's laboratory accreditation program basically extends that to all laboratory tests, and all laboratory tests

in a CAP-certified, CAP-accredited laboratory have to have proficiency testing.
DR. GOODMAN: Thank you. Proficiency testing which by the way isn't outcomes testing, it's proficiency testing. And Dr. Mansfield, just to ensure that we're somewhat on track here, none of these are FDA test kits per se.
DR. MANSFIELD: HER2/neu has at least two approved IHC test kits, and I believe two approved FISH test kits. There is an approved BCR-ABL test kit, although I don't believe it's on the market anymore, if it ever was, and there is no approved KRAS.
DR. GOODMAN: Good, thank you. So in

00257

at least those two cases there are some test kits available which makes them regulated by the Food and Drug Administration. Thank you for that.
Let me pose the question this way. For the five tests that we've discussed, I'm going to ask you about each one and we will try to just move through this quickly, I'm going to ask you for some quick discussion and then vote. If there's anything in particular about that test that bears upon its generalizability to the community or to Medicare beneficiaries, and if someone has got a comment on that, great. If not, we'll just move on.
So for example, starting with CYP2D6 for breast cancer patients who are candidates for tamoxifen, is there anything special or particular about what we know about that test, the availability of evidence or what the evidence says about its impact on outcomes that would differ or is otherwise remarkable for community-based settings or in the Medicare beneficiary population, that stand out from those criteria in any way? Dr. Eng and then Dr. Teutsch.

00258

DR. ENG: My comments are really about the Medicare population. Studies are not really done on Medicare populations, and yet breast cancer as well as colon cancer are, I shouldn't say common, but in the Medicare population beyond 65 we have factors such as medication, you know. Most of the elderly not just have cancer, but have heart disease, diabetes, other chronic illnesses, and they're all on medications. So the problem that I have, or the concern that I have is that these studies really haven't looked at medication

interactions with the targeted treatments. I mean, they're all fine tests, their treatments are all point to point, they're effective, so that's really my concern.
DR. GOODMAN: Was your concern, I'm sorry, applying to breast cancer tests and the colon cancer tests?
DR. ENG: Yes. Well, the CML and the BCR-ABL as well, but we don't see as many patients in the Medicare population with CML.
DR. GOODMAN: Thank you. Dr. Teutsch.
DR. TEUTSCH: Perhaps Dr. Mansfield can speak more clearly to some of this, but I

00259

worry that particularly for the UGT1A1 test, the FDA label actually talks primarily about reduction of harm to consider it, things like that, which would suggest that people who have reasonable interest and familiarity might be aware of that, but not necessarily the fact that they don't balance harms and benefits that aren't very clear. So I think that unless you're in, I'm not saying that necessarily an academic environment will get it right either, but you've got a particular problem if this information doesn't get out efficiently to a provider, so I do have concerns about that sort of use.
DR. GOODMAN: Is the concern about the generalizability to the community, and/or the Medicare population?
DR. TEUTSCH: It's primarily to the community.
DR. GOODMAN: Thank you. Anything else about CYP2D6? I think that was nicely addressed.
Now UGT1A1 for colon cancer, anything particular or remarkable about the generalizability to the community or to

00260

Medicare beneficiaries?
DR. TEUTSCH: I was referring to that too.
DR. GOODMAN: That's what Dr. Teutsch was referring to just now, good. I just wanted to make sure there weren't any other comments about that.
What about HER2/neu, the applicability of the evidence to the community setting and/or to the Medicare beneficiary population, any comments about that, the generalizability or as we sometimes say, the external validity of what we've got for evidence to those settings, community care and Medicare? Any further

comments about that? Okay. BCR-ABL, the evidence for that test insofar as it might apply to the community, anything special or remarkable we need to know about that as it might cross over from sort of a reference lab to a community setting? Seeing none, KRAS. We talked about the sufficiency of the evidence, what the evidence says. Is there anything we heard today that is remarkable with regard to, or might be remarkable with regard to a community

00261

setting application of that evidence, or particular to the Medicare beneficiary population? Any comments about that? Okay. I'm wondering if you want to take these as a group or one on one. Any preferences? I'm glad to take them one by one or in a group. Any comments or preferences by the panel? I don't want to lose information here but I don't want to belabor it either.

DR. JANJAN: I think we should just give CMS what they want, as a group.

DR. GOODMAN: As a group then, fair enough. I see concurrence both among federal employees and nonfederal employees. So this will amount to two votes then, one is going to be for community-based settings and one's going to be in the Medicare beneficiary population. And so, three says, how confident are you that these conclusions, and the conclusions we drew, remember, were about two things, the sufficiency of the available evidence as well as what the evidence said about impacts on outcomes, so we're asking you to kind of put those together and consolidate those.

00262

DR. MATUSZEWSKI: Is that just for the three that we found some evidence on?

DR. GOODMAN: Thank you for clarifying that. It's about all five tests. For the five tests, including the two applications of BCR-ABL, how confident are you that the conclusions you drew today about availability of evidence and its impact on outcomes are generalizable to community-based settings, where one is you have very low confidence and five you have high confidence that it's generalizable.

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

DR. GOODMAN: Now the same question about confidence with regard to its

generalizability for all the tests to the Medicare beneficiary population. So again, if you have concerns about the evidence we heard about today that is especially nonapplicable to the Medicare beneficiary population, you would want to note that, and one is low confidence, high confidence is five, generalizability to the Medicare population.

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

DR. GOODMAN: Okay. So those are questions one through three pretty comprehensively. Now, in a pretty focused way here, panel, and I will say to Maria Ellis, Maria, I think it's a good bet that we'll be done by 3:45, I think that's a conservative estimate, so if you need a little bit of a time check, I've got 3:03 now.

MS. ELLIS: The shuttle, they said they'll try to get here at 3:30, but they know they should be here by 3:45.

DR. GOODMAN: So we're at the right place, then, with regard to time. It's a prediction now, it remains to be seen how our panel does.

So here's our discussion now, and this is not a voting question. And let me preface this as follows, and we said this at the last several MedCAC meetings. Certainly one of the important uses of MedCAC meetings is to get some kind of reading on the evidence with regard to particular kinds of technology, and that's good. One of the other very useful aspects of these meetings is to try to provide some signals, if you will, to the market, and by the market I'm saying the innovators, manufacturers, doctors, patients, patient advocates and so forth, so to signal your ideas about where there are some evidence gaps in the particular cases of these technologies, and more broadly.

So let's talk now with some nicely focused comments, having gone through the day now quite intensively, about any important evidence gaps that you've seen with regard to these technologies today, these five tests, and any recommendations that might accompany those observations about how they should be addressed. And again, no Power Point allowed, no long dissertations. Please zero in if you can on important evidence gaps and what we might do about them. And let's start with

Dr. Mansfield.
DR. MANSFIELD: With reference to 2D6 testing for tamoxifen, one of the areas that I see that appears to be very lacking is standardized genotype-phenotype interpretation. I believe that it would be valuable for the community to agree on the genotype-phenotype
00265
interpretation, and for all subsequent studies to use the same standard.
DR. GOODMAN: Genotype and phenotype differentiation all to use the same standard for the CYP2D6, correct? Thank you, Dr. Mansfield. Dr. Teutsch is next.
DR. TEUTSCH: Two suggestions. One is, it would really help to have quantification of the absolute benefits and harms for each of these things and the downstream consequences so we could see what the tradeoffs actually were, and particularly for relevant subgroups. The other is because we're interested in decision science here as well as evidence synthesis, it would be helpful to have some clear sense of what the decision models look like and what the consequences are likely to be, so we can appreciate the relevance to clinical practice and be clearer about the scenario in which they are appropriate.
DR. GOODMAN: Okay. Dr. Teutsch, I want to make sure we don't lose that. Can you just, I think you made four or five requests. Can you run them through them again?
DR. TEUTSCH: I can only count two.
00266
DR. GOODMAN: I thought I heard more. Give us what you can recall, and I just want to be sure we're clear about this.
DR. TEUTSCH: One was to have some outcomes tables or whatever you like, but a table that will illustrate what the absolute magnitudes of benefits and harm were for each of these tests. And they can be done in various ways, we get the sense of NMCs and that sort of thing.
The other was because we're in decision science in the clinical world, it would be helpful to have what the cascade of events were and how the decisions were going to influence that over time, so we could understand what the likelihood of errors or interpretation, other kinds of things were on the impact of the utilization in the real world practice.
DR. GOODMAN: Great, thank you. So it

would be outcomes tables with specific information about the outcomes, and what you called a cascade of events, which really can comprise a decision model.

DR. TEUTSCH: Yeah. I'm not saying

00267

that these don't have an economic aspect that's not within our standards on decision-making, but for the other aspects it would be helpful what the decision model looks like for the relevant tests.

DR. GOODMAN: Great, thank you very much. Dr. Eng is next.

DR. ENG: The areas that I would like to see, or I would consider gaps, and even the tests with the most robust and strong evidence are, in no particular order, I did not hear today nor do I see in the reading the impact of heterogeneity or ethnicity on response of any genotype toward a targeted drug. Most of the studies were in Caucasian populations, very few in Hispanic populations, and I think, you know, fewer in African-American, some Asian. And I think this becomes important when you consider the effect on the Medicare population, and we know that the Medicare population is increasingly more diverse. So that's one. The second gap I think, though it's not, it doesn't sway me from believing that these tests are applicable to the Medicare population, I do think that there should be

00268

more studies in those who are older, because there are so many morbidities that happen in the older population, not just cancer but other illnesses, so the conditions might intersect and the comorbidities could confound, you know, the response to the targeted treatments. And finally, I do believe that we need more studies on the drug-drug interactions, particularly if we are going to be looking at the Medicare population, older population. We don't know whether the multiple drugs that the elderly are taking for their other chronic illnesses will in any way enhance the benefits or enhance some of the harms of the targeted treatments.

DR. GOODMAN: Thank you, Dr. Eng. So then, it was capturing heterogeneity including ethnicity, ethnicity in an area where there's already little evidence for anyone, let alone the groups that you cited. The Medicare beneficiary population, including the aged. And of course, drug-to-drug interactions and

the issue of comorbidities. So nearly all these deal with the matter of heterogeneity and more evidence needed there. Thank you. Dr.

00269

Janjan is next.

DR. JANJAN: Thank you. My thought on this is that we need better data with regard to functional outcome. The whole point of doing this personalized medicine is to reduce toxicity, improve function, and overall outcomes. The cost of cancer care is greater in lost productivity than it is to deliver care, and I think if we're giving therapies that cause toxicity, especially if they have no effect, that's the worst of all situations. So the goal of all of these studies is to identify the patients who are going to respond and in who we can avoid toxicity. That should be the focus of every clinical trial, and how patients are responding to these drugs for their own personal view of their quality of life under these treatments. That should be included in all upcoming trials because, you know, futile care is probably the worst we could probably administer, so I would appreciate future trials to address that.

DR. GOODMAN: Thank you. So when we do talk about health outcomes, be very specific with regard to functional outcomes, avoidance

00270

of toxicity, matters of quality of life and, as you said, keep that goal in mind.

DR. JANJAN: Right.

DR. GOODMAN: Thank you, Dr. Janjan. Dr. Cox is next, followed by Dr. Fischer.

DR. COX: One of the things that intrigued me today about the presentation from Tufts was the idea of repurposing prospective randomized trials. I made a comment to some of my colleagues at lunch that six, seven, just a few years ago, many of the trials that we participated in in community practice, we were not collecting tissue in the past. Now all of the trials we participate in, whether they're phased early trials or randomized trials, we're collecting tissue that can be tagged or tapped to do just what Dr. Janjan discussed. I just really see this as a well, especially when you're looking at hard questions, that this idea of looking in our genomes, or looking for polymorphisms in genomes when you need a large database, that you know the clinical correlates.

So I guess my question, where I see

the gap is maybe for folks like me. I need to
00271
understand the methodology a little better about what you mean by repurposing these trials, and think that could be a real help.
DR. GOODMAN: Thank you. So you're asking a methodological question raised in part by the discussion of the repurposing of RCTs. Dr. Trikalinos, I saw you stand up with an apparent attempt to answer Dr. Cox's point. Did you have something to say?
DR. TRIKALINOS: I was quick to sit down.
DR. GOODMAN: You were quick to sit down, and for you that's okay, since you've done a lot of standing up today. So you didn't want to comment on that necessarily?
DR. TRIKALINOS: I just wanted to clarify that when I was speaking about repurposed RCTs, and this was in the slides, I was speaking about their ability to inform us on pharmacogenetic associations, are they present or absent, as I clarified in the slide but I didn't dwell on it. They do not inform on outcomes and they do not inform on the impact of treatment decisions, so essentially the kind of data that you are alluding to,
00272
repurposing cannot give them to you.
DR. GOODMAN: Thank you, Dr. Trikalinos. Dr. Fischer was next and then back to Dr. Teutsch.
DR. FISCHER: As we talk about functional outcomes and as we talked about older patients with their comorbidities, it is remarkable to me that one of the -- and the Karnofsky scale in outcomes from the chemotherapeutic agent -- it is remarkable to me that one very very critical aspect of the patient is almost never mentioned, and that's nutritional status, especially in this particular disease.
Now you know, we used to, in the surgical literature one always relates it to nutritional status, to serum albumin or serum transferrin. Europeans have a very different view of serum albumin and of CRP, that the impact of inflammation, quote-unquote, whatever they call it, and in some of the European literature they write that cholesterol gives rise to inflammation, but there's this nebulous concept. I think the nutritional status or the inflammatory status or whatever you want to --
00273

one of the people who was in my lab for three years, now years later is the principal exponent of albumin as a surrogate measure for inflammation, or low albumin. So I think he's a traitor, but that's okay.

I think that's totally neglected, and as long as you're looking at functional outcome, I think that that particular aspect of the status at the beginning of treatment and what happens at the end of the treatment, if you want a reasonable functional outcome, that has to be included, I believe. And you have the patient who crawls in there at 120 pounds, you're not going to do much for him.

DR. GOODMAN: So nutritional status as a cofactor in looking at the baseline.

DR. FISCHER: Followed, at the baseline and in follow-up, with some very simple biochemical tests.

DR. GOODMAN: Thank you, Dr. Fischer. Dr. Teutsch is next, and then Dr. Juhn.

DR. TEUTSCH: We talked about the actual preferences and shared decision-making up front in the use of these tests. So we talk here about the importance of these tests in

00274

clinician decision-making, but that really needs to be done in the context of shared decision-making with patients. And particularly as you deal with the elderly, we talked about comorbidities, we talked about people with perhaps limited life spans, but clearly all of us have preferences and values, and it would be very informative to know to what extent the information about these tests actually use an informed patient's decisions that deal with the use of tests and as we would say with probably everything that is done with cancer therapy. So it's just an important and oftentimes we think that there's patient-related outcomes, but patient preference in terms of the decision-making process.

DR. GOODMAN: So we need enough evidence to inform those, to allow for those patient preferences to occur or be stressed.

DR. TEUTSCH: And we need to know how patients feel about it and how you can better inform them, and do these tests play an important role or secondary role, that sort of thing.

DR. GOODMAN: The impact of the test

00275

on patient behavior. Dr. Teutsch, before we leave you on this point, though, I want to

recall to you your concern about the lack of comparisons from one test to another. I wonder if you could just recapture that in a nugget for us before we move on.
DR. TEUTSCH: Sure. I mean, we were looking at these tests pretty much in isolation and whether they inform our understanding of disease. But most of these are happening in a clinical context where there are alternatives, and what we really care about is the incremental benefit or harm compared to the alternative process that would take place, you know, sort of what would be the next best alternative, so we can understand what the incremental benefit is, and as I said, that would be the absolute benefits and harms compared to alternative ways to manage the patient.
DR. GOODMAN: And you were or were not satisfied with the amount of evidence with regard to those concerns?
DR. TEUTSCH: I was not.
DR. GOODMAN: You were not, okay.

00276

DR. TEUTSCH: And again, I'm not an oncologist, so there may be, the oncologists that we heard from and others can inform these kinds of things, they may know. But at least for me as a person just looking at the evidence that was put before us, it wasn't as if you didn't use KRAS and you decided to, there was some other really great chemotherapy out there, you would ask why are you testing at all.
DR. GOODMAN: Thank you for that, Dr. Teutsch. Dr. Juhn.
DR. JUHN: So, I will revisit the item that I brought up this morning, which is related to some of the methodological issues such as heterogeneity, such as kind of the data mining concerns and multiple comparison concerns. This suggestion is not so much for other investigators doing studies for these particular diagnostic tests, it's really for our colleagues at Tufts and their colleagues in the evidence-based practice centers and perhaps this is something that AHRQ might want to look into, which is, are there different methodologic considerations for doing technology assessments for these types of

00277

gene-based diagnostic tests and future gene-based therapies, just given all of the unique statistical and analytical concerns.
The concern that I raised this morning

I'll just repeat, which is, are we trying to use a methodologic framework for doing technology assessments for the diagnostic testing that may work in a setting like A1c testing for diabetes, are we trying to use that same framework for a set of diagnostic tests that may have many complexities far beyond A1c?

DR. GOODMAN: Dr. Juhn, you still would agree, I think based on your earlier comments, the evidence framework notwithstanding, somehow it's got to show how to get from a test to a healthier patient, or better patient outcome.

DR. JUHN: Absolutely. It's not so much the causal change that I'm focusing on. What I'm commenting on is really how do we assign different criteria for the importance of the methodologic considerations. Because the way that, and I'm hoping that the technology assessments will be used by the investigative community, is that they will see what plays

00278

well or what scores well in the technology assessment and what doesn't, and by that type of understanding that they then will make, give more consideration to some of these factors in the absence of having that kind of framework.

DR. GOODMAN: Thank you for that. Next is going to be Dr. Mansfield, Dr. Pao and Dr. Hayes.

DR. MANSFIELD: So, this is a technical issue and I'm making this plea sort of as an FDA employee, but I think in these clinical studies, bio-specimens, appropriate ones, whether they be blood or tumor or whatever, to the greatest degree possible should be retained, well annotated with clinical outcome, polypharmacy information and so on, so that we can actually do more retrospective looks at these types of disease, or drug test associations.

I know the FDA is on the verge of requesting this for registration trials. I know that Carol Thompson at NCI is doing a lot of work on bio-specimen collection and handling, annotation and so on, and I think that will make some of these evidence questions

00279

a lot easier to answer in the future.

DR. GOODMAN: Great, good point. And it has methodological relevance because if you don't want to have to do prospective trials or RCTs in many instances, then you've got to have some rigorous data from retrospective studies,

and in the case of laboratory testing and pharmacogenetic testing in particular, bio-specimen archiving can come in real handy. This is a good use for those kinds of retrospective applications. Thank you, Dr. Mansfield. Dr. Pao.

DR. PAO: I just wanted to make a point that the data has been very eye opening and informative, but it's been under, I guess in this room, it would be great if all this information could be sent to the people who are actually running the trials so they run them in the proper manner. In conjunction with that, there's about 861 drugs in development right now for cancer and so there's many many diagnostic tests and platforms coming out, and to have each one of these assessed in this manner is going to take a lifetime. So if there was some kind of dissemination or

00280

agreement upon how trials could be run and how these specific genetic tests should be used in these trials, it would be great, so we wouldn't have to reinvent the wheel every time.

DR. GOODMAN: Thank you, Dr. Pao. Well, if nothing else, but I'm sure there will be more, the comment you just made has been captured by our court reporter which is going to show up in a transcript somewhere, which means that it's going to be citable, but we hope that you will do much more of that. Dr. Hayes, and then I believe Dr. Jeter.

DR. HAYES: I have three comments/suggestions. The first, I think we should all take Dr. Voigt's comments this morning to heart, which is that sometimes I'm concerned that a committee like this will truly dampen enthusiasm for innovation by being too regulatory. The flip side of that is that these tests are becoming increasingly more important in taking care of patients. It's not like a hemoglobin where you can repeat it a few times, but in fact we're talking about either withholding or treating patients with very

00281

expensive and toxic, but quite effective drugs, and so I have no problem actually raising the bar to a level that almost equals that of a new drug.

Having said that, I think what Dr. Teutsch has talked to us and taught us about over the last two or three years regarding the three cornerstones of

diagnostics, analytic validity, clinical validity and clinical utility, are terms that need to be really ingrained in everyone's brain who is doing this research and taking care of patients. Over and over again I think we see lousy analytical validity so we can't figure out whether the assay is any good or not, we heard that today. We see confusion between clinical validity, gosh, I see a P value of .04 separating two curves, and clinical utility which is, gee, I actually designed a study to tell me whether or not this test helps me take better care of the patients.

And it is the latter that is much more important but it's the former that gets promoted, and I fear those two get confused often. This committee I think could go a long

00282

way in making it clear that we're looking for clinical utility and not clinical validity. That can be prospective, and as we heard it can also be used in archived samples. And Dr. Mansfield's comments were terrific. Rich Stein, Cindy Bacon and I just published a paper about it's okay to use archive samples, but there's a hierarchy there as well, and some of those studies are lousy and some of the studies are very good, and you have to be aware of those.

My second comment is how these results are published, there's an incredible publication bias. And it's interesting that we discipline ourselves here based on using only peer reviewed published data, which I'm all for, but the data I think are poorly edited. And in addition, the editors frequently will not take negative studies, they only want positive studies, especially the high impact journals. Which means that the lesser, the negative studies which are very important, are usually published in a lower impact journal where you tend not to see them. And I think this committee has gone a long way in sending a

00283

message to the editors of the high impact journals that a negative marker study is every bit as important as a positive study, to encourage investigators to focus on their negative data.

And to use the so-called REMARK criteria that Lisa McShane and her colleagues worked very hard on developing. Several journals have said that they will use these REMARK criteria for publication, and yet the

editors ignore them completely.
DR. GOODMAN: Dr. Hayes, let me just stop you. The REMARK criteria, could you explain the acronym, please?
DR. HAYES: It's R-E-M-A-R-K, and it has to do with reporting tumor marker data in a way that you actually tell where those patients came from, where the samples came from, how they were stored, how the assay was done, a whole set of things, and Lisa McShane, et al., there was a committee that put these together. They've been published in five or six journals now that have adopted these, but they actually have ignored them once they've adopted them, if you want to know the truth. And over and over

00284

again I see journals that say we've adopted these, and then a paper is published that has completely ignored the REMARK criteria.
DR. GOODMAN: Okay, the REMARK criteria. You were about to close on a final point?
DR. HAYES: Yes. Finally, of all the things we've seen, obviously we'd all like to see the CYP2D6 data come. I can tell you with a conflict of interest, those data are coming down the pike and we're going to have a lot more data regarding the choice of tamoxifen versus other agents using archived samples from randomized trials in the next year or less, so we should have data for that.
The second real gap, I think, is how to use intermediate or, if you will, the clinical scores for HER2. We've assumed that these drugs don't work in patients who are one plus or two plus positive and FISH negative, but we actually don't know that, and there are a couple of signals from the randomized trials that patients who got into those trials because they were called positive somewhere else, when tested centrally, were negative with no

00285

benefit. This is a huge area, because it could be really expanding the education for these very effective drugs, so that's an area that I would like to see proper analysis.
DR. GOODMAN: Great, thank you very much for those three points, Dr. Hayes.
Dr. Jeter.
DR. JETER: Thank you. I have been quiet to this point because I felt I was a guest and this was a first-time experience for me, and I thank Louis for the opportunity. One of the comments that I want to make is that we

received a huge volume of material to go through to make assessments today for five assays. As a contract medical director I'm here to tell you that in the pipeline right now, there are easily 1,500 molecular assays that everybody is dying to get coverage for. And I know this is not a coverage meeting, but I think that the extent to which this committee has gone to, the MedCAC, to pull data together, to have these TEC assessments, gives you an idea of what really needs to be available to the contract medical directors, not necessarily to this extent. And what we're

00286

seeing at our end is, you know, one or two marginal to poor articles that are published in the literature with, you know, minimal sample size, that kind of stuff, with everybody clamoring.

And I understand the whole concept that there isn't a pile of money to run all these trials and everything, but this gives you an expectation of what CMS and the whole medical community wants and needs, and that is evidence-based decision-making. With that, I'll end my comment.

DR. GOODMAN: Dr. Jeter, we're not going to let you off the hook there. So there are many tests in the pipeline that are fast on the way to your desk, and you expressed in a very nicely summarized way how limited the evidence is typically that you see. Does the absence of evidence or those limitations in evidence, are you ready to say no in a lot of those cases?

DR. JETER: I have. We have.

DR. GOODMAN: We have. Thank you for that.

DR. JETER: And we're under the gun

00287

for that. I mean, we've got non-coverage policies out there that are in draft right now, and any number of organizations are clamoring, you know, that for whatever reason, that there's sufficient data, and there isn't sufficient data for many of these. And almost all of the molecular assays, none of them have, or I should say all but a couple of them have any clinical utility. They have the analytical and the clinical validity, and some of that isn't even published, you have to beg, borrow and everything to get that out of the companies, because they're claiming that it is all proprietary, and we understand the whole

concept of proprietary. But without that, we can't make an assessment or any kind of determination.
DR. GOODMAN: Okay, thank you, Dr. Jeter. For those in the marketplace, innovators, manufacturers, patient groups, physicians, providers, that sounds like a pretty clear signal to me from a well-recognized payer. Thank you for that. Dr. Mansfield has done a little side research on REMARK, and we're doing this in

00288

particular for our court reporter. Would you tell us what REMARK stands for?
DR. MANSFIELD: Yes. REMARK stands for REporting recommendations for tumor MARKer prognostic studies, and it's a paper by Lisa M. McShane, M-C-S-H-A-N-E, et al.
DR. GOODMAN: Thank you for that, Dr. Mansfield.
With that, as we want to make sure, the MedCAC wants to make sure that the very fine people that have spent their full day with us here in the CMS auditorium have been heard. Have we as a panel missed or not heard any very important points that are directly addressing our evidence question today? Did we miss anything big or important? We hope you won't walk out of the room saying well, they forgot to talk about X, and that should have been right up their alley. What important things might we have missed today with regard to answering these questions about the evidence with regard to those five tests or their generalizability, anything important here? Yes, sir, and please do come to the mic with a brief comment about that. There's

00289

another woman here who will go second, and if you could keep it to a sentence, we would be very appreciative, or two short sentences, and give us your name first.
DR. AVERBUCH: I'm Steve Averbuch, I'm a vice president of oncology clinical research and head of pharmacodiagnostics at Bristol-Myers Squibb. Just that I wanted to make a couple comments about specific cells in the BCR-ABL discussion.
DR. GOODMAN: First give us the main point.
DR. AVERBUCH: The main point is that in terms of the TEC assessment and the point mutation discussion, we talked about the T315I but we didn't discuss the other point mutations

that lead to therapeutic decision-making. So for example, Dasatinib was actually, the scientific rationale for its development was based on mutations that conferred resistance to imatinib, so I just want to make that point. I think it also skews the technology assessment in terms of the assessment in terms of those point mutations.

DR. GOODMAN: Thank you.

00290

DR. AVERBUCH: And then with respect to the molecular monitoring assay for BCR-ABL, I wouldn't dispute anything that people said here, just remind the panel that the gold standard is still cytogenetic outcomes, and that's the gold standard for regulatory review and approval as well as other hard outcomes such as survival. So whereas there may be clinical application of molecular monitoring, the health outcome data in terms of hard endpoints of survival and total outcomes is not there for the molecular monitoring.

DR. GOODMAN: The evidence is not there for molecular monitoring, and you made your point about the point mutations. That was acknowledged by our panel, we did say we were going to focus on that one. Thank you for that. Yes, your name?

MS. COLLINS: My name is Sarah Collins, I'm president of PharManage. I think this has been addressed but I want to be a little more prespoken, or blunt. I think this is all, this issue that there is an original risk to the originators, and I think this is spoken to in the importance of FDA approval,

00291

since there are imitators, frequently called home brew, and so whether a Quintiles or LabCorp do these and the accuracy is increased, the importance of that is increased by managed care or other contracts with the large labs. So I just wanted to make that point as much for the people in the back, as well as for this audience.

DR. GOODMAN: Great, thanks. There's no front and back, we're all in the same room. Thank you very much. The last point of business for the panel, and I will give Dr. Pao a warning here, everybody on the panel gets the last word. And so what we want to ask at this point is, in one sentence or bullet point, is this. What's, even if you said it before, what's the single most important point you want to make about

evidence for these tests to either CMS or those in the market who make or use these tests? So it's about evidence, it's about these tests; what's your last word, that single most important point you want to make to our host here at CMS or to those assembled here? And we'll start with Dr. Pao and come back this way.

DR. PAO: Well, it's become evident from today that if you're going to have a test with clinical utility, it better affect a patient's health outcome.

DR. GOODMAN: Thank you, Dr. Pao. Dr. Mansfield.

DR. MANSFIELD: I think it would be immensely valuable, to the degree it's possible, for CMS and FDA to work together so that registration trials for drugs and diagnostics will also yield information that will fulfill our evidence requirements.

DR. GOODMAN: Thank you, Dr. Mansfield. Dr. Jeter.

DR. JETER: Just that we have to have clinical utility for coverage.

DR. GOODMAN: Thank you, Dr. Jeter. Dr. Juhn.

DR. JUHN: Tip of the iceberg, this is just the beginning.

DR. GOODMAN: Thank you very much, Dr. Juhn. Dr. Teutsch.

DR. TEUTSCH: I think we still need more work on the evidence standards, particularly how they need to be adapted to different clinical situations.

DR. GOODMAN: Thank you, Dr. Teutsch. Dr. Scheuner.

DR. SCHEUNER: I think my comment is similar to Dr. Teutsch's. I think we need to have studies that examine the alternatives so that we can really have a bigger picture and put this in context when we think of clinical utility.

DR. GOODMAN: Excellent. Dr. Matuszewski.

DR. MATUSZEWSKI: My comment is for CMS and FDA to partner with the Medcos and Caremarks of the world. CVS Caremark went out and bought their own company's generation house because they think that's how important it is. And again, it's amazing that sort of the private side of healthcare payers is way, way ahead of CMS in this case. Usually it's sort

of everybody waiting to see what CMS is going to do, but not this time.
DR. GOODMAN: Thank you, Dr. Matuszewski. Dr. Kaul.
DR. KAUL: I think we heard some
00294
really eloquently done TEC assessments comparing disparate testing data to try to make sense out of this. The pathologist in me would remind people to take a hard look at how samples were collected, how they were sub-dissected and what the technology for the analysis is, because sometimes just the analytical factors are going to affect the results.
DR. GOODMAN: Thank you, Dr. Kaul. Dr. Janjan.
DR. JANJAN: Clinical utility is where it is and if we can't make good clinical decisions based on this data, it's useless.
DR. GOODMAN: Thank you, Dr. Janjan. Dr. Hayes.
DR. HAYES: I made my comments earlier.
DR. GOODMAN: Thank you, sir. Dr. Fischer.
DR. FISCHER: Thank you. As somebody who sits on multiple editorial boards and from what I saw today, there does need to be some uniform way of reporting for patients exactly what the outcome of these 1,500 tests that
00295
everybody is rushing to put together, and it either comes to us or somebody else to approve or disapprove, but you can't if the data is not there. So clinical outcomes, survival, recurrence, we all mention all these things, and I wonder whether CMS or this group can say okay, if this is what you want, this is what you're going to have to tell us, and that would help everybody.
DR. GOODMAN: Thank you, Dr. Fischer. Dr. Eng.
DR. ENG: When I started reviewing the literature for this MedCAC I began to worry that the field of personalized medicine might become elite for those with money. And then I thought about, well, in order to make this equitable and be available to everyone, we really do need evidence, because good evidence will convince providers, and here I will say physicians, to be able to say, to be able to provide that kind of care to all their patients, not just the ones who come knocking

on their door saying I read this somewhere, give me this test. And the physician may be able to say well, I know you read this

00296

somewhere but here's the evidence, it won't work for you. So I think that we're very far from that time, but to actually have reviewed the tests that we did review today, because there is some good evidence, but I think going from here to the point where we make this available to all our patients, not just Medicare patients, is a long way to go.

DR. GOODMAN: Thank you, Dr. Eng. Dr. Cox.

DR. COX: It's hard to add anything to what's been said. I appreciate CMS using this as a focal place to cry for evidence. I also appreciate the Agency and FDA continuing to work together. One of the confusing things for docs in practice is to be challenged by a laboratory who presents into my office hawking a test and claiming that it's FDA-approved, when it's very difficult then to look beyond the FDA approval to its real clinical utility.

DR. GOODMAN: Thank you, Dr. Cox. Dr. Atkinson.

MS. ATKINSON: I just want to add on to what Dr. Eng had said earlier about making sure that research gets done in this Medicare

00297

population. But in addition to that is making sure that when it's done in the Medicare population, that we're not just looking at the healthy robust older adults but the frailer older adults as well, and really looking at barriers to practice. Why, if the research is out there and the evidence is strong, why are we not doing this, so what are the barriers to practice and then the barriers to acceptance by the population that we're serving.

DR. GOODMAN: Good, thank you, Dr. Atkinson. Dr. Satya-Murti.

DR. SATYA-MURTI: I have the last word, all right. In addition to everything that has been said, I think in clinical oncology the outcomes are not equally weighted. I think it's a special field where deaths and progression-free survival, and patient-reported quality, I think we need to have some kind of weighting on these in some measured nuanced way of what's important, and just survival may not be as applicable in certain circumstances. So certain weighting of the outcomes might be good.

DR. GOODMAN: Thank you,
00298
Dr. Satya-Murti. I am going to give a few closing comments before I turn it back to Dr. Jacques. First of all, I thank the panel very much. I think it's been enlightening for all, and hope it's been enlightening for everybody in the room today.
And just a couple of closing remarks. Tip of the iceberg is right, and the tip of the iceberg that we've seen so far in many ways has been scientifically extraordinary. The sequencing of the human genome is now seven years past, I guess it is, and there can't be a person in this room who has not been thoroughly impressed by that, except that now that's not enough. And when you think about the kinds of signals that come out of a meeting like this, they're very clear with regard to science is great, but what we also need is clear evidence about analytical validity and clinical validity, and clinical utility, where clinical utility embraces direct evidence of or clear evidence of impact on a decision, an impact on an outcome, the way we heard outcomes described here today.
The signals that come from a meeting
00299
like this should help to shine a brighter light for innovation, redirect it in ways, not quash it, not put it away, but help shine a light towards science and new technologies that will improve patient outcomes with substantial rigorous evidence in support of that. Not just guesses based on sensitivity and specificity, but evidence that shows improved patient outcomes in the very ways that you heard today.
And the suggestions that were voiced today come not from government employees, they come from a diverse range of people in the healthcare community. So we're not the green eyeshade people, we're not the beloved bureaucrats, these are people in the community who work with these every day.
I would add just one more point that wasn't made explicit today, and I did hear the phrase pile of money, I did hear some discussion about how much clinical trials and other rigorous studies cost. I think I recall that the two largest companies in the laboratory sector have a combined, or had a combined revenue in 2008 of about $12 billion, and chances are it's a little bit higher this
00300

year. And so, $12 billion is not all profits, but as I read in the Wall Street Journal at one point, $12 billion would be the envy of a lot of global pharmaceutical companies for revenue, so there may be more funding available to do the kinds of rigorous studies that we need here. And with the signals that have been given today and the indication about the need for evidence, there can be an efficiency of those rigorous studies and the investment in those, and there will be a return on the investment that will be payable to Medicare beneficiaries and other patients.

So, thank you, panel, thank you, participants. I want to thank in particular our initial speakers, Dr. Freedman and Dr. Trikalinos and his team. I want to thank the nine very patient and to-the-point prearranged speakers, who were superb. And our two signed-up speakers from today, and our two last commenters from the audience, this has been a superb input, very helpful.

Dr. Jacques.

DR. JACQUES: Thanks. First I want to go ahead and thank Cliff for running a very

00301

good meeting and also echo his thanks to the panelists as well as the presenters and attendees.

I will reiterate, there are currently no open NCDs on these topics. Don't read too much into this. I'm not suggesting something's going to happen tomorrow, so go to sleep tonight, okay? It's okay. We are periodically asked, though, whether we'd consider doing NCDs in this particular space around genetic testing, and so far the only one we've done pharmacogenomic testing to determine was to predict warfarin responsiveness, and we chose that one intentionally rather than wading into cancer, arguably the warfarin issue is much simpler clinically.

At the same time, one of the reasons why we wanted to have this meeting intentionally without an open decision is to get a sense of what would be the challenges that we would be facing if we actually chose to actively engage in this particular space, and look at evidence that might inform coverage policy related to genomics and cancer for various indications.

00302

I think some things are quite clear. One, this space is so nuanced that even the

development of the question itself, as well as the breadth or the scope of the question in some cases will dictate that they'll get one reply or one answer rather than another. So this has been extraordinarily helpful in that case.

We intentionally chose topics where we had a sense that there would be some heterogeneity in the evidence, in fact to see how the panel and by extension the public and the stakeholder community would essentially respond when questions like that came up.

I'll close with this one comment. If one views the development of genetic testing as something of a train, the community can either pull that train or CMS can push that train. Now if you pull fast enough, we will never catch up to be able to push you, so I will leave that as my advice to you.

DR. GOODMAN: And with that, is the meeting adjourned? The meeting is adjourned.

(Whereupon, the meeting adjourned at 3:46 p.m.)