Transcript of November 7, 2000 Meeting

Please Note: This transcript has not been edited and CMS makes no representation regarding its accuracy.

00001

HEALTH CARE FINANCING ADMINISTRATION
Medicare Coverage Advisory Committee
Executive Committee Meeting

November 7, 2000

Baltimore Convention Center
One West Pratt Street
Baltimore, Maryland

.00002

Panelists

Chairperson
Harold C. Sox, M.D.

Vice-Chairperson
Robert Brook, M.D.

Voting Members

Leslie P. Francis, J.D., Ph.D.
John H. Ferguson, M.D.
Robert L. Murray, Ph.D.
Alan M. Garber, M.D., Ph.D.
Michael D. Maves, M.D., M.B.A.
Frank J. Papatheofanis, M.D., Ph.D.
Ronald M. Davis, M.D.
Joe W. Johnson, D.C.

HCFA Liaison
Sean R. Tunis, M.D., M.Sc.

Consumer Representative
Linda A. Bergthold, Ph.D.

.00003

Panelists (Continued)

Industry Representative
Randel E. Richner, M.P.H.

Executive Secretary
Constance Conrad, R.N.

Expert Consultants
Kathy Helzlsouer, M.D., M.H.S.
Ellen G. Feigal, M.D.
Manuel Cerqueria, M.D.

.00004

TABLE OF CONTENTS

.00005

TABLE OF CONTINUED

.00006

PANEL PROCEEDINGS

(The meeting was called to order at 8:25 a.m., Tuesday, November 7, 2000.

DR. SOX: I would like to welcome everyone to this meeting of the Medicare Coverage Advisory Committee Executive Committee. My name is Harold Sox and I am chair of the committee. To my right is Dr. Robert Brook, who is the vice chair of the committee. We are going to start off with introductions, and when you introduce yourself, I would like you also to state if you have any conflicts of interest so that we will all know about them. And what you need to comment on, I guess this is at your table, is whether you have any direct industry financial investments, whether you have any consulting fee arrangements with any FDG PET related supplier or corporation, and whether your institution has any significant support from a source of FDG PET. So with that as a request, Alan, would you please start?

DR. GARBER: I am Alan Garber of the Department of Veterans Affairs and Stanford University. I have no conflicts of interest.

DR. FERGUSON: John Ferguson, now a private consultant, former director of the NIH

.00007

consensus program. I am a neurologist and have no conflict of interest.

DR. FEIGAL: I am Ellen Feigal. I am a medical oncologist and deputy director of the Division of Cancer Treatment and Diagnosis at the National Cancer Institute. I have no conflicts.

DR. SOX: Before you go on, can everybody here okay, or is it just me that's having trouble

hearing? Whoever's in charge of AV, could you crank it up a bit please? Go ahead, Linda.
DR. BERGTHOLD: I am Linda Bergthold. I am with the Center for Health Policy at Stanford University and I have no conflicts, and I'm the consumer representative.
DR. HELZLSOUER: I'm Kathy Helzlsouer an epidemiologist and medical oncologist from the department of epidemiology at the Johns Hopkins School of Public Health. No conflicts of interest.
DR. FRANCIS: I'm Leslie Francis. I'm professor of law and professor of philosophy at the University of Utah, and I have no conflicts.
DR. CERQUERIA: Manuel Cerqueria. I am a cardiologist and nuclear medicine physician at Georgetown Hospital. I am a member of the diagnostic imaging panel. I have no conflicts.

.00008

DR. DAVIS: I'm Ron Davis, a preventive medicine physician at the Henry Ford Health System in Detroit. I have no conflicts. I am not aware of any relationship that my institution, the Henry Ford Health System, might have with the FDG PET industry, so if it does have any such relationships, I am not aware of them.
DR. PAPATHEOFANIS: I am Frank Papatheofanis, I am at the University of California at San Diego. I'm also a nuclear medicine physician, and I chair the diagnostic imaging panel. I wish our institution had relationships with the FDG PET industry.
MS. RICHNER: I am Randel Richner, from Boston Scientific, and as far as I know, we don't make PET or anything associated with that.
DR. BROOK: Robert Brook, from UCLA and from Rand. The only conflict that I have is that two of the speakers are also from UCLA.
DR. SOX: My name is Harold Sox. I'm a general internist and chair of the Department of Medicine at Dartmouth. I don't have any conflicts and PET has not made its way into rural America yet.
DR. TUNIS: I'm Sean Tunis, I am the director of the Coverage and Analysis Group at HCFA.

.00009
MS. CONRAD: Good morning. I'm Constance Conrad, I am the executive secondary of this committee.
DR. SOX: Okay. I have a few brief opening remarks to try to set the stage for today. Today, the Executive Committee convenes to evaluate the evidence about several applications of a diagnostic test, PET. This is not the usual function of the executive committee. That function is ordinarily reserved for the panels, and specifically the imaging panel.
In preparation for this meeting, we have, we, and I say principally Alan Garber and myself, have developed guidelines for evaluating evidence about diagnostic tests, and in the fullness of time, we will add these guidelines to the interim guidelines that we approved earlier this year. We will use these guidelines today to evaluate two and possibly three applications of PET scanning, colorectal cancer management, the differential diagnosis of dementia, and lung cancer diagnosis and staging.
Our purpose today is threefold. First, it's to advise HCFA on the quality of the evidence and the magnitude of the effect size for these
.00010
applications of PET scanning. Secondly, it's to give our new guidelines for evaluating diagnostic tests a workout, with the expectation that during the course of the day and afterwards we will refine them and they will then be available for use by the diagnostic test panel, the imaging panel, and particularly by HCFA staff as it considers other application of PET scanning. The third purpose, if we can, is to render an opinion about whether conclusions about PET are readily generalizable to other cancers and to other uses of tests besides the ones that we will consider today.
So, are there any questions from the panel or comments before we get started? Alan.
DR. GARBER: Yes, Hal, thank you. I just wanted to mention to the Executive Committee members

that these guidelines for evaluating diagnostic tests have not undergone review by either the Executive Committee or the subcommittee of the Executive Committee that is charged with making revisions to the existing interim guidelines and as such I think that Hal and I intended what we've written to really be a starting point for our discussions today. No one should have the impression that we believe these are in final form in any sense, and I think it would

.00011

be appropriate for the panelists to express any disagreements they might have or any changes that they think might be appropriate in the document that has been distributed to you.

DR. SOX: I would just like to underscore that. From the long range point of view, the most important purpose of this meeting is to give these guidelines a workout, to refine them, even though we have an urgent short-term goal o accomplish as well. Bob?

DR. BROOK: When are we going to do that? Do we have two minutes to just talk about a couple major issues with this paper?

DR. SOX: Well, the plan, if you can wait a bit, the plan is to hear from Dr. Phelps about PET scanning. Then I'm going to go over our framework for evaluating diagnostic tests, and then we'll have a full hour to discuss that.

DR. BROOK: I just wondered since what I -- are the panelists, are the people aware of what this document is, have they seen it? Since we're evaluating what they're doing, have they seen the document that they are going to be evaluated on?

DR. TUNIS: The document was just posted about a week ago on the web. Dr. Phelps, I believe

.00012

got a copy of it a week or less ago. It has only been drafted in the last 10 to 14 days.

DR. BROOK: I mean, I think there is a philosophical statement in this document that's going to be difficult to deal with, and I don't know when you want to get into that.

DR. SOX: Let's get into it after I get a

chance to lay out the framework, so everybody will be on the same page, if that's okay. Any other comments from panelists?
In that case, the next item on the agenda is to hear from Michael Phelps, who will discuss the science and biology of PET scans.
DR. BERGTHOLD: Hal, could you introduce the two people who just came, and have them do their conflicts before we start?
DR. SOX: Oh, I'm sorry.
DR. MAVES: I apologize. I'm Dr. Michael Maves, I am the president of the Consumer Healthcare Products Association, and I have no conflicts with regard to PET.
DR. SOX: Bob?
DR. MURRAY: Robert Murray, technical director of clinical laboratories at Advocate Healthcare. I have no conflict of interest on the
.00013
items that are noted on the conflict of interest statement.
DR. JOHNSON: Joe Johnson, chiropractic practice, no conflict of interest.
DR. SOX: Thank you. Anything else before we hear from Dr. Phelps? Thank you.
DR. PHELPS: Thank you very much.
DR. SOX: Sir, would you introduce yourself and give your affiliation please.
DR. PHELPS: I am Mike Phelps, I am from UCLA. I'm the chairman of molecular medical pharmacology and the director of the molecular gene institute and also the laboratory for structural biology and molecular medicine.
So before I begin, I would like to tell a quick story before we have to get very serious about all the things you have to do today. So the story is about three people who were riding in a car. In the back is a cardiologist and a microbiologist; in the front is a chemist. They stop at a light and a guy jumps in the back seat and puts a gun to the head of the cardiologist, and he says tell me what you do and why it's so important that I shouldn't shoot you. The guy says well, I'm a cardiologist and I save the

lives of people who have heart attacks, and bam, the
.00014
guy shoots him. So he puts the gun to the head of the molecular biologist and he says tell me what you do and why it's so important I shouldn't shoot you, at which point the chemist in the front seat says, for God's sake, shoot me first. And the guy says, why the hell should I do that? He says man, I cannot stand to hear another story about how great molecular biology is.
(Laughter.)
So molecular biology is great, it is changing the world we live in. There are 20 genomes that are being sequenced out and it is coming forward with medicine to form the new molecular medicine. And part of what I will show you today is in fact, molecular imaging technology is a part of that new movement. So if I could have the slides and the lights off?
Unfortunately I'm going to have to make the lights a little bit dark so you can see this, so let's begin with just the principles of PET. It is a molecular imaging technique, so we take molecules, in fact we can't form an image without molecules, we'll label that with a positron emitter of oxygen 15, nitrogen 13, carbolatimer fluorine 18. These isotopes will emit a positron that will move a short
.00015
distance, annihilate the two photons that are emitted back to back, and we use that unique property for the detection of opposing detectors that will register about 20 to 40 million of those electronic combinations simultaneously, and then we reconstruct the image. In these molecules they are injected intravenously, diffused throughout the entire body, and then participate in the process that they mimic. This particular molecule is the one that we are going to talk about today, deoxyglucose. It was originally developed actually at Washington University by the Coreys, a husband and wife team, they both won the Nobel prize, and they had developed deoxyglucose as a biochemical assay for glucose metabolism. And in fact, there has been about 25

years of work on this particular molecule, deoxyglucose. And not only was it used for a biochemical assay, but Lou Sakaloff at NIH developed it with carbon 14 as an autoradiographic technique which became the standard throughout the world for imaging glucose metabolism in animals with autoradiography.
You can see an example of the image. This is a tomographic image, which is typical of the studies that are performed with PET. They can either

.00016

be to an organ or to the entire body, and this is a longitudinal tomographic section, it's about five millimeters thick, it's a woman that had a previous resection ovarian cancer, and you can see the glucose metabolism in the brain, the arms, and also the heart, and then recurrence of her disease bilaterally in the lungs. So this is the general type of assays that we use for various types of compounds that we use, but in every case there's going to be a molecule that will originate from biochemistry, biology of the pharmaceutical industry.
Here you see two examples, and this is where the general concept is used.
Fluorodeoxyglucose competes with glucose for the transport sites within the tissue, and then hexacarnase needs to be phosphorylated to the 6-phosphate form, and that is not a substrate for further reaction, so it's retained in the cells so that a map now is provided of glucose metabolism throughout the body.
Here you see a patient with non-small cell lung carcinoma. The coronal and sagittal longitudinal sections, you see the tumor here as high glycolysis. So the trapped gulcose-6 phosphate now represents the glycolysis throughout the body. The

.00017

compound below is another analog of, it's a thymidine, it's a fluorodeoxythymidine so it's another deoxy analog, and in fact came from a group of compounds that were developed either to assay DNA replication of cell proliferation, or to therapeutically treat it. The most popular version

of that is AZT.
In this case, the fluorodeoxythymidine, though, is used as a biological assay, DNA replication, and here you see the full body distribution, the replication throughout the body in that same patient, so you see the tumor has high replication and high metabolism. And these are the general types of assays that are developed.
Now, when we do the studies, these are tracer studies, so the amount that is injected ends up producing a mass in the tissue, it's in picomoles or nanomoles or phenomoles that you program, so they're tracer levels without disturbance of the biological processes.
Now, just for a minute looking at glucose metabolism with deoxyglucose, the entomology of cancer cells has been known for about 50 years now.
As neoplastic degeneration occurs, glycolysis is amplified about 19 to 30 fold because the Creb cycle

.00018

is lost in the progressive degeneration of neoplasms, and in addition, glucose is actually used as a carbon skeleton for the DNA and RNA synthesis. So, there's a very high amplification of glycolysis that allows us to identify the tumors away from other tissue, and to see small lesions.
But just looking at how general this principle is, cancer biologists have established this as a fundamental issue in neoplastic degeneration.
But just looking at some examples of different primary metastatic disease, here you see an ovarian carcinoma as you saw before with metastases of the lymphatic system in the lower left quadrant, prostate cancer metastasis in the lymph nodes and also the lung, Hodgkin's lymphoma with lesions throughout the body in the skeletal and soft tissue, breast cancer with an 11-millimeter lesion and behind that, a 7-millimeter lesion in the primary breast, axillary lesions, lung cancer, primary metastasis in the lymph nodes, melanoma lesions throughout the soft tissue, indicating that in fact also in patients, we confirm what we see in cancer biology, that this is a general generic process for neoplastic lesions.

But you're not going to be able to see the lesions here, but -- actually, can we turn down the
.00019
lights a little bit more, is that possible? So one of the important aspects in cancer with PET is the fact that we can look at the entire body in one single procedure, so we can go in and inspect every organ system and examine for the primary disease and also metastasis throughout the body.
Here you see an example of a woman with breast cancer where the primary lesion is seen here, one of the primary lesions in this breast. Although you can't see it, some very small lesions in the axillary lymph nodes, and also in the internal mamillary, and another primary lesion in the opposite breast. But also in the liver, the lymph nodes and it lung an throughout the bone. So in a single procedure, we can quickly sort through all the organ systems, and identification of asymptomatic disease is a routine issue in examinations with PET.
Now with that issue in mind, I want to raise just a question about early disease and show you some examples that disease can be identified from a biological perspective many years before even symptoms occur. So, there were studies performed on symptomatic Alzheimer's patients in which cases the CTs and MRs were normal, and it was well established that PET could accurately identify the metabolic
.00020
abnormalities of Alzheimer's and in fact, the other organic dementias. And here you see a classic example of hypometabolism in the temporal cortex, the normal MR, compared to the normal HMX control, and the metabolic deficits extend from this level at the temporal cortex up into the parietal cortex.
Now, we wanted to show in fact that we could identify disease long before the symptoms occur. Stages were compensatory responses and reserves were being used to compensate for an error of disease. So we went to a genetic disorder, the classic one of a hereditary dominant disorder is Huntington's disease, so we studied patients for 15 years. It's a study by John Mazziota, and it was

published in the New England Journal of Medicine. Some of these patients had a normal study of metabolism in the caudic putamen that was the site of the expression of the hereditary disease, and we had known that from studies in patients that were symptomatic. But in these asymptomatic patients, some of them also had metabolic deficits in the caudic putamen and in fact the distribution of them was clearly and accurately correlated to mendelian predictions of who carried the bad gene. Now these patients had every psychological and neurological

.00021

exam that you can imagine to show that they were asymptomatic, but over the course of 15 years, every patient that went on to express symptoms, it was preceded by a time where there was some metabolic deficit in the brain, and that was the case for every patient. And in fact, the longitudinal nature of the study showed that we could identify the metabolic deficits about seven years before symptoms occurred. In a similar way in familial Alzheimer's, by Gary Small in papers that were published in JAMA, New England Journal of Medicine and Proceedings of the National Academy of Sciences, also showed in Alzheimer's that the metabolic deficit as shown in this patient in the parietal cortex, is shown here, and these early abnormalities actually tend to occur unilaterally and then with time spread to a bilateral distribution. And in fact in this study, which was correlated to the occurrence of APOE, it was shown that the metabolic deficits were occurring or could be detected about five years before symptoms occurred.

Now in a similar way, moving to the heart, glucose had been known to be a protective substrate in ischemic tissue, and that was used to identify patients who would benefit from revascularization

.00022

from those who had not. An example of that is shown here in a patient who has a left anterior descending coronary artery occlusion, in a superior and a midlevel cross-section to the heart. This is myocardial blood flow that can be either imaged with

ammonia or ribilliun, both of which are FDA approved. And you see the blood flow deficit here in the anterior wall, but looking over to the glucose metabolism, you see that that area is in fact, that there is an acceleration of glucose metabolism in that area that's ATP efficient, or is sufficient at producing ATP in oxygen limited states. So this is what was called a mismatch.

And it would predict that a patient would benefit from revascularization, as opposed to patients that had a match where there was a flow and metabolism were both reduced, and the patient would not benefit from revascularization. You can see, this patient has a very low ejection fraction, about half the normal value, and there's akinesis in the anterior wall. So this patient was taken to angioplasty. You see three days later, the flow has returned, glucose metabolism is fairly normal throughout the left ventricle, but the akinesis in the ejection fraction are still low.

.00023

Now from basic biochemical studies we knew that during this time was the time where restoration of cellular functions and membrane potentials were taking place before the heart could return to work. At seven days, you see that blood flow and glucose metabolism now are normal and the injection fraction and the wall motion are also normal. So this became actually a gold standard for predicting which patients would benefit from revascularization. This will be discussed by Jamshid Maddahi later.

In this article, the first article of this was also published in the New England Journal of Medicine by Jan Tillisch and Hank Schulberg and their colleagues.

Now in the last segment, I just want to look at the question that Dr. Sox brought up in the beginning, comparing different classes of tests, and we will break them into biological and anatomical, and we will look at some of the fundamental issues between these two tests. And of course, we should always keep in mind that disease is a biological process.

Now, the principles of anatomical versus biological imaging, anatomical imaging, x-ray films, CT, MR and so forth, have empirical relationships to

.00024

the detection of disease. Now that's not bad, that's just the way it is, and that's fine. There is no fundamental relationship between electron density with x-rays or CT, or hydrogen density with MR and disease.

In a biological test with PET, there has been a fundamental basis from over 80 years of biochemistry and biology that normal organ function and their failure in disease, so that's well established in the basic sciences. This was also a part of the basis for FDA's broad approval of FDG PET along with a literature based evaluation of the clinical research, and Dr. Love will go through that today.

PET molecular energy probes come from biochemistry and biology and the pharmaceutical sciences. This also provides a natural link to the biology of disease, as well as between molecular diagnostics and molecular therapeutics; that is, we don't actually develop the molecular energy probes. That is done and their proven principle occurs in basic biochemistry, biology and the pharmaceutical sciences.

Some facts about glucose metabolism in FDG. Glucose metabolism is critical to proper cell

.00025

function. 95 percent of ATP for cerebral function comes from glucose metabolism, so it provides an excellent way to assess the functional or metabolic status of the brain. Glucose metabolism is protective in ischemic tissue. This is well established in biochemistry. I showed you an example in the heart, but other tissues have also been shown. Glucose metabolism increased 19 to 25 fold in cancer; that's what we talked about. FDG measures glucose metabolism, well established in biochemistry and also the PET literature. You can differentiate malignant from benign tissue. It's a fairly straightforward evaluation with PET, where it's not

with anatomical approaches. About 20 to 40 of biopsies in the lung, and 68 percent in the breast are benign. While there are some indications empirically in differentiation between malignant and benign, it is a difficult process, with anatomical techniques.
You can differentiate malignant tissue from adenous, necrotic and scar tissue. This is an issue in primary disease, metastasis, but also in recurrence and therapeutic evaluation. Differentiate reversible from irreversible tissue, as we talked about. Detect early disease, even asymptomatic

.00026

disease, without detectable anatomical changes. We know that most diseases go on for many years before they actually become symptomatic, so the biological nature of disease exists for years.
Now, the last slide simply shows you a way to look at a broader context of PET. We have developed not only in the clinical systems but also little systems that sit on bench tops, and we use them to study mice as a part of the genetic revolution to engineer disease in, mammalian disease into mice, to study it in terms of its biological nature and also therapeutics, and I don't have time to go through this, but this is an example of an approach with PET to measure gene expression, the imaged gene expression quantitatively in the living mouse, so to bring the genome to life.
Here you see a study in which we have transferred a gene into the liver of a mouse with an adeno virus, and then we use a technique called the PET recorder gene, PET recorder probe, to actually image gene expression in a living mouse. So here the genes have been transferred into the liver for the adeno virus, the PET recorder genes and therapeutic genes, and then any time we want we just inject a PET recorder probe to image the gene expression. In he

.00027

control study there is no reporter gene so there is no gene expression to image. Two days after we gave he virus, you see gene expression throughout the liver; four days it's decline and by two weeks it has

disappeared because the virus has terminated the transfer. But just to illustrate that there are many different probes that are being developed for cell communication, synthetic processes, metabolism, and all the way down to the level of gene expression. Thank you very much.

DR. SOX: Thank you very much, Dr. Phelps. Does anybody on the panel wish to address any questions to Dr. Phelps before we proceed?

DR. TUNIS: Just one question, Mike. Are there -- you mentioned in one of the slides towards the end that the PET imaging is good at distinguishing malignant tissue from adenous tissue and other differentiations. Are there any sorts of tissue normal or pathological for which PET has greater difficulty in terms of differentiating between malignant tissue and nonmalignant tissue?

DR. PHELPS: Yes, there are some nonmalignant inflammatory processes that in some instances do have a high glucose metabolism, so there is a false positive rate, it's fairly small, but it's

.00028

real.

DR. SOX: Bob?

DR. BROOK: Mike, can you give us some basic facts that are not in our material about how many PET scanners there are now in the United States, about how many total patients have undergone them, have there been any studies on the reliability of multiple readers in different centers reading these images at all in terms of what's going on, and are most of these people now on some protocol or research, or is a lot of it being done routinely? Just put us into some context in the year 2000 of what is going on at the moment.

DR. PHELPS: Okay. There are about 800 PET scanners in the world and they are about 50 percent in America. There are over two million studies that have been performed. The shift over the last five years has gone from research to clinical service, and has spread throughout hospitals and clinics to more routine base. There are educational programs in most of the major universities to educate

the general practitioner. And if you look at some of the clinical trials, for example in some of the publications, as Gary Small will mention, we actually do the evaluations with well trained physicians and
.00029
then we take a very short time, train naive physicians, and have them also read the studies, and the concordance is about 90 percent. So, the studies are actually quite easy to read, because the contrast is so high in the lesions. What was the other?
DR. BROOK: One last question. The average exam takes about how long to do?
DR. PHELPS: It varies. From the brain, about ten minutes; to the whole body, depends on now people will either do an entire body or will go down to below the pelvis, so those studies take 30 minutes to 45 minutes, with some of the systems an hour for the whole body.
DR. SOX: Any other questions? Well, in that case, the next item on the agenda is the discussion of the evaluation of the framework for evaluating diagnostic tests. And what I will do is to summarize what is in the material that you should have received on Friday prior to your review of the data on PET scanning. I thought what I would do actually is to go through, kind of stop after each transparency and have a chance to discuss it, so that perhaps we can sort of conflate the presentation and discussion together, and then of course there will be more time at the end.
.00030
So Connie, or somebody, can I ask somebody to show these transparencies? Does anybody have a laser pointer that I can use?
DR. PHELPS: Yeah, here.
DR. SOX: Thank you.
Well, we want these guidelines for evaluating diagnostic tests to fit into the framework that we developed for evaluating other technologies and therefore, our basic question is, is the evidence adequate to conclude that the use of the test will lead to a clinically significant improvement in health outcomes as compared with the use of either

established tests or nothing.
Now, ideally, we would, the form of evidence that we would have would be a randomized trial in which patients are assigned either to get the test under consideration or the established tests. And then these patients would be followed through for a period of time to allow outcome events to accumulate, and then you would compare the frequency of outcome events in the two groups. There are relatively few studies of this type. The best example certainly are the eight or nine randomized trials of screening mammography which have been done over the past 40 years involving probably 40 or
.00031
50,000 women overall.
And -- but we don't have very many examples of that and so what we do know about diagnostic tests is mostly their test performance, how accurately they detect patients with disease and how frequently they have false positive results indicating disease in people who don't really have it. So that's the information we have about tests, and the challenge for MCAC panels is to see if we can infer effects on health outcomes from what we know about test performance, so it's a much less straightforward problem. Any questions about this one before we go on? Ellen?
DR. FEIGAL: Yes. I have a question about what you mean by health outcomes. What I would like to be clear is, does the panel think there is intrinsic value in having an accurate diagnosis regardless of what the treatment options are? I mean, we don't have to discuss that now, but I think that's an issue to raise.
DR. SOX: Yeah. Many people would classify that as an intermediate outcome that may or may not be linked to an outcome that really makes a difference in terms of the patient's sense of well being or their emotional well being. Anybody else
.00032
like to comment? Alan, you helped me on this, so I want you to be -- I don't want to be -- I'm supposed to be the chair, not an advocate here of this.

DR. GARBER: Ellen's question is a very important one and I think that from my point of view anyway and I am only speaking for myself, that health outcomes may not be limited to something like effects on mortality or even measured morbidity, it's a broader sense of well being. So in my own opinion, we should have an expansive view of what constitutes a health outcome, but once we have that view, the test should be demonstrated to improve that set of health outcomes.

DR. SOX: So for example, a sense of emotional well being after having an accurate diagnosis could be a health outcome, if you could measure it. Bob?

DR. BROOK: I'm just wondering how we got into this box of the wording of that first item. I think the first question that we need to answer with a diagnostic test, is there evidence adequate to conclude that the use of the diagnostic test leads to the same accuracy as the previous materials that already are here? In other words, we have been excluded from covering costs or any of these kinds of

.00033

questions under the stuff that we have been dealing with. I'm not sure that, the initial question ought to be a very simple straightforward one, is there enough evidence that this is a reasonable alternative to what exists now in the diagnostic processing of diagnostic testing? So that, I mean, we would like to know the answer, or I would like to know the answer to the question you raised, but for the purpose of this panel, we're missing the first priority, which is, is there evidence here that you know, this is at least as good as what you got and, in terms of what's going on, in terms of accuracy. Then I would like to make it formally known that in terms of health outcomes, I think things such as the reliability of the -- the ability to transport the test into the community versus in the laboratory is extraordinarily important in that kind of question. And also, the convenience and the ease that the patient -- is this a test that is more comfortable to the patient? So if we have two tests

that were basically equivalent in terms of diagnostic accuracy, not even talking about outcome, and one, the patient just had to appear and somebody used that laser pointer and got the answer, which we are going to get to sooner or later, and another that you had

.00034

to open them up, and even if they produced the same long-term outcome, I would view that as a significant breakthrough.

So, I'm not sure that the phrasing of this question is the way that we ought to have it.

DR. SOX: Well, just one comment on the point about evaluation of effect size, our interim guidelines first ask, is the evidence adequate to conclude anything about the effect size, and then we have a hierarchy of effect sizes that go all the way from breakthrough down to causes damage, and it would seem to me that hierarchy would embrace something that doesn't really change health outcomes.

DR. BROOK: I'm not disagreeing with that, but I would love the first question to be asked in an unbiased way.

DR. FEIGAL: I would second that.

DR. BROOK: I think HCFA needs to know what we feel about the evidence that exists there from the -- first and foremost, you got something out there, it may cost a trillion dollars. We've been told not to consider money, but as far as we can tell, if it's safe, it's effective as anything out there, and we ought to say that loud and clear as the first comment. We may then say look, there's no

.00035

evidence to say this is better than, or used in combination it's better than, or any of these kind of things, but there needs to be an a priori statement made here about something that relates to the first priority that hey, you know, it's a reasonable alternative.

DR. SOX: Well, maybe one way to frame it is, is the evidence adequate to reliably measure effect size, and then the second step in the process is, what's the size of the effect. Alan?

DR. GARBER: Well, I'm not certain that I

understand Bob's question, but I believe that if you go on further in the document, that is the question that's asked. And maybe it would be helpful if you went through the entire document and we have discussion at that point, just to insure that we don't quibble over points that might come later in the document.

DR. BROOK: I'm sorry, Alan, I don't believe this is a quibble. I will shut up, but if you look at the phrasing of question one, I've read this document carefully. In question one, question two, we are talking about evaluating diagnostic -- the things that are bold, I always look at bold things first, and the bold things are all reflecting

.00036

something better than, significant improvements. And I agree by the way, I mean I would agree that those are the right questions to ask. But that's not the mission we were charged with when somebody stepped forward to us and said you know, we can't consider costs, you can't consider these kinds of issues. In that case, we ought to become true to the mission, and the mission is really, the first issue is, is there enough evidence out there that this is reasonable for people to use, it's safe, it's effective, it looks like it's as good as anything else. Is the evidence for this about the same as it was for other tests like CAT scans and MRIs, and where are we in that continuum? We need to answer that question before we can then take -- I think we ought to answer the question, one of the two, and I love the document, but I think we have to answer that a priori question.

DR. GARBER: Well, Bob, let me ask you a question. The first subquestion under the bold face, and this is in italics, not bold faces, is the evidence adequate to determine that the use of the test provides more accurate diagnostic information? Is your point that it should instead ask, provides at least as good as?

.00037

DR. BROOK: Yes.

DR. FEIGAL: Exactly.

DR. BROOK: That's it. I mean, the tenor of the document is technically superb.
DR. GARBER: Okay. Now I'd like to suggest that we put that there. This question will be dealt with, however, also in the classification of effect size, as Hal was alluding to. It is intended to be part of the main interim guidelines document which has the seven categories of effect size, so you can assign it to either -- if the evidence is adequate, then you can assign it to a category that says equal effectiveness, or greater than or less than with some other benefits, and so on, as you were alluding to before.
So that would change question one, that the test provides diagnostic information that is at least as accurate as standard alternatives, or words to that effect.
DR. FEIGAL: Or offers some other advantage.
DR. BROOK: The reason I'm saying this is that when I read the TEC assessments report, they compared it to a gold standard as opposed to the use of other technology, and I was a little bit -- and
.00038
there was no statement that I could see in how the sensitivity and specificity of this was similar to or better than existing modalities, and it was all phrased in better than and in terms of gold standards, at least as I read through these things. And I just wondered if we should at least point out that we want an answer to the first question first. It's not sufficient, but I'd like to see us make a statement regarding the answer to that first question.
DR. GARBER: Well, Bob, let me just point out that the rewording, which in principle I appreciate, does have practical implications. And one issue is, to prove at least as good as, that means it is sufficient to prove that it is no worse, at least in my estimation, and if you have a series of small studies that are inadequately powered and from them you cannot conclude that it's any worse, does that constitute adequate evidence or is that not

adequate evidence? Obviously, underpowered is a value judgment.

DR. BROOK: I believe that we ought to answer the first question first and it may be as good as, better, because there really is evidence, it may be good as, because we can say that the studies are

.00039

bad in both cases, but I think we ought to answer the first question first, because that's to me a very important question. Now, is that the right policy, I'm not going to get into a policy debate because that's not our consideration. But what this tells me is the only thing we are going to address is whether, the major emphasis is on whether we get a better outcome. The outcome may be no good because there's no therapy, the outcome may be no good because once you do it, the radiologist has to turn it over to me the internist, and I screw up with the results. There is a lot of reasons why the outcomes are allowed to be lousy.

I really want to know the answer to the first question, given the evidence of why we have CAT scans and x-rays and MRIs and all those other things that we do, spiral CTs and everything else, the question is, is this as good as an alternative, is it reasonable for a reasonable man to conclude that this is as good as anything, as good as the other alternatives that currently exist? That's the first question. We then can ask whether it improves significant outcomes, we can ask about the evidence, we can ask about the compound, you know, prior posterior probabilities, we can go into all of that,

.00040

but we ought to come clean on the first question.

DR. GARBER: I would like to make a suggestion, since we're leaping ahead a bit here in the document, but one point to make is that if we change the wording this way, I think there should be a clear understanding that adequate evidence means based on sufficiently large and well designed studies to conclude whether or not two tests are at least equal ones, or whether the new test is better. So there is a clear idea that there should be an

adequate evidence based and technically speaking, the adequacy of the evidence base is much more difficult of an issue when you're trying to assess equivalence than when you're trying to assess superiority. But it should be incumbent upon the panels to make the judgment that they are well designed studies of adequate size to be able to draw conclusions about whether the two technologies are at least equal in accuracy.

DR. SOX: Okay. Somebody -- I got a fair number of nods when somebody suggested that I just crank through the talk, and so why don't I go ahead and do that, and then we can come back and kind of go through it piece by piece. That way everybody, particularly everybody in the audience who hasn't

.00041

seen this can see where we're coming from. Ellen?

DR. HELZSOUER: Yeah, just one question of process. This is the first time I think it has been brought up that these have been publicly aired and this is the first time that this framework has been publicly aired. It's a little bit of an unusual circumstance in that we're setting the framework on the same day that we're evaluating an application. It's unusual. But regardless, that's how it's being set up. So my next issue after raising that problematic issue is of process. If HCFA or if this panel decides that this framework is worthwhile to use, does it just get adopted or does it go out for public comment to the technology developers, or to patients who might be the subjects of this diagnostic test? Maybe Sean or somebody from HCFA could just answer that question.

DR. TUNIS: Yes. This framework is a piece of the interim guidelines for evaluating effectiveness that are being developed, have been under development for the use of a coverage advisory committee. And as you know, Ellen, the coverage advisory committee is advisory to HCFA on coverage decisions, so the entire process today in terms of what the panel does or doesn't do in regards to the

.00042

framework, or in fact even applying the framework to

the couple of case studies that we may be able to get to this afternoon, that whole thing is sort of advisory to HCFA in terms of wrestling with the coverage decision around PET.

What you raise in terms of whether or not this would be subject to public comment, et cetera, there is a process separate from this which you also know about where we're developing the process of developing or predeveloping the coverage criteria for Medicare coverage which will be done through a regulatory process and with a proposed regulation, et cetera. The information we get here from the MCAC obviously will be closely tied in terms of us saying, you know, they will be covering the same sort of territory, but the terms of Medicare's criteria for making coverage decisions, that's a separate process that will go through a regulatory process, and there will be opportunity for public comment, et cetera. So the framework we're talking about here is for the purposes of the process of the coverage advisory committee only. Does that answer your question.

DR. HELZSOUER: Sort of. And I will try to talk into the microphone here. I guess what I was getting at is FDA has guidance documents, so that the

.00043

people who are submitting applications know in advance what the rules of engagement are, and that's sort of the process I'm bringing up, are these guidances, are these guidelines going to be something that is broadcast?

DR. SOX: Well, you know, we're doing this because there is a lot of intense effort, interest in this process, and we're doing our best with a situation that is not the usual process for this organization. The interim guidelines we have already developed have been out on the web, we have got public comment. These have been on the web for a little while and we will revise them and put them out again for public comment, so there will be a lot of opportunity for people to give input. Does that deal with your questions?

DR. HELZSOUER: Yes.

DR. SOX: Thank you. Okay, so I'm going

to go ahead now and go to the next transparency please.
So, the first question is, and I will try to edit here to reflect the earlier discussion. Is the evidence adequate to measure accurately the use of the test on health outcomes? I think that's really what we're talking about. So the first step
.00044
in that process is to evaluate the quality of the studies and test performance to find out whether the measurements of sensitivity and specificity are valid or whether they are biased, and if they're biased, to try to decide in what direction they are biased.
There is a -- many individuals have developed guidelines for evaluating the quality of studies f diagnostic test performance and for the purposes of this document, I summarized five of them by noting first the characteristics of the ideal study, and then the characteristics of the study that we all too often find in the literature, and then to show the direction of the effect of the studies that we actually get on what the ideal study would show.
So the next please.
So first of all, study subjects should be consecutive patients seen in a typical clinical setting with a chief complaint or with a well defined clinical problem. Very often the study subjects instead of being consecutive patients are patients who were selected because they had the reference test, and by choosing only people that have the reference test and ignoring people who are, who have a negative result on the index test which is the test under study, you can overestimate sensitivity and
.00045
underestimate specificity.
A second criterion is that everybody, which is related to the first, everybody who -- ideally, everybody who gets the index test should also get the reference test, but what all too often happens is that patients with negative results on the index test don't get the reference test. In the ideal study, the person who interprets the index test is blinded to all other clinical information so that

he or she doesn't, in the situation where it's a close call, doesn't tend to make the call in the direction suggested by the other clinical data. What often happens is that the person who interprets the index test knows the clinical history and often the results of the reference test, and that tends to overestimate the correlation between the reference test and the index test, and overestimate sensitivity and specificity. Next please.
And then the converse of that is the person who interprets the reference test should not be aware of all other information and the reality is that frequently they are not, and that has the same effect. Then on to the next transparency please.
Finally, the reference test should be a valid measure of the disease state but in reality,

.00046

the reference test measures the disease state itself instead of being, reflecting the deeper truth of the situation. So, for example a coronary arteriography is the gold standard for studies of exercise testing. It really doesn't measure coronary ischemia, which is the critical disease state that you're trying to defect. So, let's go on.
So, the step two in the process of evaluating the ability of the test to detect disease and discriminate between patients with and without disease is to evaluate the extent to which the test under consideration, the index test, correctly identifies patients that the comparison test fails to identify as diseased, and that's clearly pertinent in PET because as Dr. Phelps pointed out, its basis is biological rather than anatomic, basis for detection.
So, one point would be if the sensitivity of the index test is substantially greater than the comparison test, it clearly identifies patients that the comparative test fails to identify as diseased.
However, sometimes the sensitivity of the index test can be similar to that of the comparison test, or in principle, even lower that the comparison test, but it can still identify patients that the comparison test fails to identify as diseased. And so if two

.00047

tests have similar test performance, then you have to look carefully to see if the two tests complement each other.
The best way to demonstrate the complementary function of two tests is to do both tests and then the reference standard, and then to display the results of the test under consideration in patients with a positive result on the one hand and a negative result on the other, on the comparison test. So that you can actually look at the ability of the index test to pick up people that are negative on the comparison test, and that's shown on the next slide.
So what we would like to have is a table like this that shows test one results positive, test one results negative, as the two major columns, and then within that the results of the reference test, and then test two results are the rows. Now if test two picks up patients that test one fails to pick up, then A-prime will be greater than zero and therefore, the sensitivity of test two in patients who are test one negative will be greater than zero, A over A-prime. In that case we can conclude that test two is complemtary to the comparison test, it picks up patients that the comparison test does not pick up.

.00048

Next please.
Now we move on to the next part of the evaluation. Our second major question is, if the test in fact has improved accuracy, is the evidence adequate to conclude that the improved accuracy will lead to better health outcomes, or as we would reframe it after our early discussion, is the evidence adequate to make conclusions about the effect of the improved accuracy on health outcomes, and then we would characterize the magnitude of that effect.
MS. RICHNER: That's fine.
DR. SOX: Better?
MS. RICHNER: Yes.
DR. SOX: Okay, good. So, to determine whether a difference in test accuracy would lead to differences in health outcomes, the panels may find

the following steps useful. First, to calculate the post-test probability of disease, that is, the probability of disease after the test is done, and then secondly, to evaluate the potential impact of differences in post-test probability on the management of the patient.
An example of that is shown in the next transparency, which is just a little bit too big, but
.00049
what we have on the horizontal axis is the post-test probability of disease and on the -- correction -- on the horizontal axis is the prior probability of disease, and on the vertical axis is the post-test probability of disease. And the data used to calculate each point on these curves is base theorem, which requires pretest probability and sensitivity of specificity of the test.
So for example, here we have a situation where we have two tests, CT scan represented by the smooth lines, and PET scan related by the lines that connect the dots. So for example, let's imagine that the pretest probability of lymph node metastasis was 70 percent; after a negative result on CAT scan, CT scan, the probability of having positive lymph nodes would be over 50 percent, whereas the probability of having positive lymph nodes after a negative PET scan would be about 30 percent, so that's quite a large difference in probability of disease. The question is, is that difference in probability of disease likely to alter management strategies in a way that would actually improve health outcomes.
Specifically in this circumstance where you're trying to decide whether or not to do to a thoracotomy for lung cancer and you're using the PET
.00050
scan, a negative PET scan to tell you that there aren't lymph nodes there that have malignancy in them, and therefore it's reasonable to go ahead and do a thoracotomy, one might reasonably ask well, if the pretest probability of lymph nodes was 70 percent, would you do a thoracotomy on a patient who had a 30 percent chance of having malignancy in the lymph nodes. Or alternatively, would you do another

test like media stenoscopy before going for thoracotomy. So that's how the post-test probability of disease can be related to management decisions that themselves can affect health outcomes. So the two questions we could ask are, first, does the test under consideration raise or lower the probability of a disease to an extent that is useful in decision making? And it could be that one test would be a lot better than another but still, the post-test probability would not be low enough to alter management strategies. And then secondly, does the post-test probability of the two tests differ to a clinically important degree? So, let's go on.

Step two, in trying to estimate, trying to infer the effect of differences in test performance on health outcomes is to evaluate the potential

.00051

impact of the difference in post-test probability on management and health outcomes. So a test result is likely to improve health outcomes under these circumstances, when it distinguishes, when the test distinguishes very well between patients with disease and those who do not have disease and also, when the test is effective in patients with the disease, or the treatment does not benefit patients who do not have the disease. Under those circumstances, it could be very useful to distinguish clearly between patients who don't have disease and those who do. It could result in improved health outcomes if the treatment did not benefit patients without the disease, and it would not, it could also be useful if the treatment posed significant risk to the patient so that it's very important to avoid unnecessary treatment and therefore, to clearly distinguish between patients with and without disease. Anything else?

So, just to summarize where we have come from, first in evaluating, trying to evaluate the effect of diagnostic tests on health outcomes, you should start by seeking high quality studies that provide direct evidence that test results improve, or that test results affect health outcomes, and then

.00052
measure the effect of that size as compared with the established tests, to characterize the degree to which the test under consideration really adds to what we have.
If there is no high quality direct evidence, as there will not be for most diagnostic tests, then you have to evaluate the indirect evidence, first deciding whether studies of test accuracy are sufficiently free of bias to measure test performance accurately, and to be able to compare it with the established test. And then second, to evaluate the potential impact of differences in accuracy on health outcomes, first by evaluating the effect of effect of test accuracy on post-test probability and second, deciding whether changes in patient management that could lead to improved health outcomes are likely to occur as a result of the test results.
So, that's the framework that we have developed and that we will be in the process of trying out today and trying to improve it as we have already tried to do. So, why don't we just start by, in terms of trying to frame the discussion, why don't we go back to the second transparency, and we will put that up and discuss that.
.00053
First, any overarching comments before we kind of go through it piece by piece?
DR. PHELPS: Harold, could I ask a question?
DR. SOX: Please.
DR. PHELPS: I would like the committee also in their deliberations to look at something that we struggled with with HCFA and that is, when we look at the broader indications and uses, we had tried to figure out where do you draw the line, and we also looked just fundamentally, if you start with a clean piece of paper, if you had empirical tests, you know, the bias would be you should do them indication by indication because it is an empirical issue. As opposed to, if you had a broad biological or fundamental basis, then the question is how many

indications would you have to look at to try and realize three or four broader indications. So, I'd like to ask also that the Committee consider that.

DR. SOX: Thank you, Dr. Phelps. We understand that part of our process, or part of our charge today is to try to advise HCFA whether it is reasonable to generalize from a few applications of PET scanning to all applications of PET scanning and to all cancers, so that is part of our task.

.00054

So, first step, evaluate the quality of the studies of diagnostic tests, and I think it's implicit that any evidence report should address these major characteristics of a high quality study of diagnostic test. Any comments on this one?

DR. FRANCIS: Where do you put in questions about discomfort, what it's like to have the test performed, all those sorts of issues? Where do you put in things like risks?

DR. SOX: This would be a logical place to do that, to look to see if the studies comment on that issue. Alan?

DR. GARBER: Well, maybe I can reframe that question a little bit and ask a question of the Executive Committee. The amended version that we have of this question is basically, is it at least as accurate as some alternative, and if the test is not at least as accurate, the Executive Committee has to ask the question, would you still want someone to go through this process if it were clearly not as accurate, yet it provided some other benefits that could be quantified, which is what Kathy was referring to, that is, more comfortable, in some other way more advantageous when compared to the standard tests.

.00055

DR. SOX: Well, our hierarchy of effect size includes some things that are, as I recall, are perhaps a little bit less effective than the established technology but have some other advantage that might make them preferable for some patients.

DR. GARBER: This question -- no, no, that's right, but you may not get to question two if

your answer to question one is negative, so -- and it's not you may not, you will not get to question two if the answer to question one is negative, so if you cannot determine whether it's at least as accurate, is it the sense of the Executive Committee that nevertheless, it should proceed to question two and be classified? So, question one could be rephrased, is the evidence adequate to conclude anything about accuracy, basically, and clearly that's the single most important feature of the test. So that's how it would be rephrased, if it was the sense of the Executive Committee that tests should pass a barrier of having adequate evidence to say something about accuracy, positive, negative or indifferent.

DR. SOX: So, any comments on that? It seems like a reasonable rule of, operating rule, that if you can't conclude anything about the accuracy of

.00056

the test, whether or not it happens to be more comfortable or more convenient for the patient isn't germane, is I think is what Alan is saying.

DR. GARBER: Let me try it with --

DR. SOX: Alan, try to frame your question in a way that people can object to, and if they don't object to it, we can assume that we agree.

DR. GARBER: Well, I'm actually going to try to reframe it in a way that nobody can object to.

(Laughter.)

Is the evidence adequate to determine how the accuracy of the test compares to alternative diagnostic strategies, and that includes other tests and things based on clinical characteristics and so on.

DR. FEIGAL: And it's not implying it has to be better or worse, it's just can you evaluate the accuracy of this test?

DR. GARBER: Right.

DR. SOX: Bob?

DR. BROOK: Why don't we just use the first question of what is known about the accuracy and reliability of the test? Why don't we answer that for HCFA? Part of the subquestion becomes is it

25 better than, compared to what, compared to an
.00057
1 alternative, but the first question is, what is known
2 about it? I mean for instance, the first speaker's
3 comment that we have had two million PET scans done
4 in the world and that we're going to be examining
5 technology assessment reports that deal with hundreds
6 of patients and less, I think at least we ought to
7 make a comment that there's a huge missed opportunity
8 in this field for producing the kind of data that you
9 have talked about. And that is the -- I mean, the
10 discrepancy between the stuff we're looking at and
11 what's happening is, you know, the difference between
12 a pilot and a microbe or something like that, and I
13 am really concerned about that.
14 So, I would like us to have some of that
15 in the evidence based report here about what we're
16 doing, and I'd like to make that first question
17 pretty neutral, what is known about this, so we ought
18 to start out with something about this is the current
19 state in our report that's not in any of these
20 technology reports, these are the numbers, these are
21 the machines, these are the millions of patients that
22 are getting this. Our evidence is based on hundreds
23 of patient, and why is it only based on a hundred
24 patients and then what's the evidence, what do we
25 know about this reliability?
.00058
1 And so one of the things they can't tell
2 us is that we don't know more about reliability and
3 accuracy because it hasn't been done. I mean, that
4 becomes an obvious conclusion all of a sudden that
5 hey, it's been done two million times, and look at
6 the miserable small amount of information that we
7 know about reliability, or here, maybe we know a lot.
8 But the bottom line is that if we set it in this
9 context, it would be a much less biased presentation
10 than either comparing against alternative or against
11 improved outcomes. That's all. The technology here,
12 the academic technology here is great.
13 DR. SOX: Bob Murray?
14 DR. MURRAY: All of these questions are
15 phrased to elicit a yes/no answer, but the reality is

it is rarely black and white, all of the studies have some value, all of them have some weaknesses. Alan asked a number of questions basically which boil down to, do we go forward, and I think unless a study or a question, unless a study of the evidence is utterly devoid of quality, yes, we do go forward.
DR. SOX: But if we don't know enough to assess accuracy because of poor quality studies then we don't go forward, right?
DR. MURRAY: My point is that the
.00059
question, do we know enough, is a difficult question. Do we know enough, yes, no, well, we know something and unless we know virtually nothing -- in other words, I suggest that we set the bar fairly low so that we don't exclude or we don't prevent ourselves from looking at all of the evidence.
DR. SOX: Well, anything more on this one before we move on? John?
DR. FERGUSON: I guess this is an old saw, but before I had suggested that rather than, is the evidence adequate, what is the evidence, and what is the evidence to determine that this test is comparable, less or more accurate, I think is a better way to discuss the evidence than is the evidence adequate, but I said that before.
DR. SOX: Alan?
DR. GARBER: Just one point. We had a discussion on basically this same issue when the Executive Committee unanimously approved the interim guidelines of the Executive Committee, and this document and the questions were drafted to adhere very closely to the format and the wording of those questions. Now, we could always revisit that in a more general way, but I would like to suggest that whatever we decide to do, we try to be pretty
.00060
consistent between diagnostic tests and all the other kinds of health interventions that the panels will be evaluating.
And I actually do think breaking things up into questions this way, the first one being about the adequacy of evidence, has been very useful and it

does not imply that evidence should be overlooked. There should be a complete cataloging and evaluation of evidence in the course of responding to question one. All of the issues that Bob Brook raised are relevant, important and should be included in the process of answering question one. I don't think that implies a rewording of question one. That's how it has been interpreted in the technologies that we have studied on the medical surgical panel.

MS. RICHNER: I disagree to a certain extent. Based on what Dr. Murray has just said, if you answer a yes or a no, you stop, and that's the problem. So I think that we have to get to an equivalency point here with the questions, so I disagree, I think it needs to be reworded.

DR. GARBER: Well, Randel, I should amend that a little bit. I just realized that you but not most of the members of the Executive Committee have seen the early drafts of the revised guidelines which

.00061

unfortunately I guess we won't have time to discuss today, but there were some other approaches to dealing with this issue that have been suggested, and maybe Hal, if there's an opportune time later today, we could discuss those approaches, but they are designed to deal with the issues you raised.

DR. TUNIS: Can I also just, you know, kind of impress on the Executive Committee a little bit just from the perspective of what I think would be helpful for HCFA, and I do think that Bob Murray's comments about answering yes/no to the question of, you know, is the evidence adequate, you know, buried beneath that is probably an even more important question, and I don't know exactly how to phrase it, but it's something about like qualitatively, how good is the evidence? And it's going to range from either, you know, nothing, to you know, every study is ideal.

And it seems that the aggregate of evidence sometimes is going to be suggestive and sometimes it's going to be very suggestive and sometimes it's going to be almost definitive, but you know, when you look at the body of it, there is going

to be a spectrum of the overall evidence and for this committee, I think to think about how to characterize
.00062
that as an end point or at least as part of question one would be helpful.
And then so the second question really is the committee's view on whatever it is, is that adequate or is that good enough, or some judgment about where this threshold, should some magical threshold in there should be. I think we need -- you know, there is a lot of information in this sort of gradients of quality of the overall evidence that might be helpful, and at least I would throw that out.
DR. FRANCIS: You bring it up nicely in the comments but maybe it's worth underscoring also here that it's probably relevant whether the evidence goes to the likelihood of false positives or the likelihood of false negatives, because that might be awfully relevant depending on what management is there.
DR. SOX: Why don't we go on to the next transparency and see if there is any discussion about that. This transparency and the next basically are sort of a very concise version of tables that are seen in many studies of evaluating different studies of diagnostic tests. And clearly as part of our homework for preparing these guidelines, we have to
.00063
get a table that is more complete than this and more precisely phrased, but I wonder, is anybody concerned about the concept of using established measures of the quality of studies of diagnostic tests as a way to answer Sean's question, which is how good is the evidence? Anybody got any trouble with that? Bob?
DR. BROOK: I don't understand what the purpose of this is, Hal. Is this to tell -- if we go through our usual process, we are actually going to commission these technology assessments. Is this to tell the person who does it that we want the evidence presented to the panel in this way? I mean, is this a statement of just here are some issues? For instance, I agree with -- I think this is beautiful,

but let's -- you say the study subjects are consecutive patients seen in a typical clinical setting for the chief complaint. I might make this that they ought to be, that the test results ought to be interpreted by a typically trained person in the profession that is probably going to do that, so -- but that's nitpicking.

What I'm asking is, is the purpose of this to say unlike what we got today, which are, we commission these technologies, so part of this document is written to the preparers of the evidence

.00064

report. Can I suggest that be separated out into a document that says we ought to produce guidance to what the -- I mean, are we going to produce a document that is a guidance document to the preparers of the technology assessment, or is this sort of a, you know, cheat sheet to have the Executive Committee know that when people talk to them, they ought to look at least at some of these issues. What is this? That's my concern with it. I have no problem with it, it's wonderful science, I just don't know what it is.

DR. SOX: Well, I think the main purpose of it is to instruct the people who present the -- create the evidence report, to give us information that will allow us to decide whether the sensitivity and specificity are valid measures.

DR. BROOK: Well then, you see, I come back, because under step one you say the panel should first address the quality of the studies. Now I don't think the panel can do that. I think that's why we have a technology assessment report. That's why I'm nitpicking about this thing, is this really -- I mean, I agree that we ought to agree on a standard format so it should make it much easier for the panel and the presenters, and we ought to ask the

.00065

presenters also to adhere to this when they talk to us or we ought to say basically, you're out of order. I mean, if you can't do it this way, we're not going to listen, because it makes no sense to us and we can't hear it.

So the question here is, I would suggest that we do prepare a document at some point to help deal with the technology assessment report in a way that would make it more useful to us, and I think it's especially important when we do a diagnostic test.

DR. SOX: Okay. Anybody -- Bob has made a specific suggestion. Anybody have any objections to our doing that? Alan, objection?

DR. GARBER: This isn't exactly an objection but a friendly suggestion for a change in that. I think Bob, the intent of this is not just to guide people who write evidence reports. It's to guide everybody involved in the process and that includes the panel members, it includes public presenters --

MR. BROOK: Then it's amended. We ought to write something that does that.

DR. GARBER: Right. And your point is, it should be much more complete, absolutely, and it

.00066

should function as a stand-alone document. And of course there is a great deal of material out there in the literature, some of it's been distributed to the Executive Committee, that we can draw upon. But this was a shorthand way of trying to accomplish that, and I agree with your suggestion.

DR. BROOK: It would be helpful to me because I can't, when people present the data from the floor, it would be helpful is somebody said okay, this is in line with what we have proposed or not, so that we at least know and that people who are going to present know, and the people that write the technology, that we expect information to come to us in a format that we can understand.

DR. SOX: Ellen?

DR. FEIGAL: Yeah. This gets to the issue I was bringing up in the beginning, is the guidance to the people who are trying to develop the evidence, do they know what the rules of engagement are? The second issue I wanted to bring up was in terms of the issue of bias, and that some of this implies as you go through the table, and so what I

wanted to know is in part of this -- it's easy to find bias. Is there some way to make some estimate of the magnitude of the bias and whether the bias

.00067

will qualitatively change the results, or simply quantitatively change the result. And I don't know how to get that in, but it's easy -- it's not easy but it's often the case that you can find problems with the methods in which studies were conducted and so you can say there was a bias, but is there some way to quantitate the magnitude of the bias and whether or not it's qualitatively, not quantitatively going to affect the results.

DR. SOX: Yeah, there are some techniques that have been worked out that apply under some relatively limited assumptions, so I guess the limited answer to your question is yes, there are, and we certainly would want to ask the folks who prepare the evidence report to do whatever they can to characterize the effect of the bias on measures of test performance.

DR. HELZLSOUER: Qualitatively.

DR. SOX: Qualitatively or quantitatively, if they can. Alan?

DR. GARBER: Just for the benefit of Ellen and other people who have not participated in prior proceedings of this panel, that is an issue that is ubiquitous, certainly not limited to diagnostic test literature, and the Executive Committee decided to

.00068

leave it up to the panels to decide whether the results could be explained by bias. And as Hal says, there are quantitative methods for attempting to do that but if you dig beneath the surface, they are all Bazian methods, which essentially means that you have to guess at some point what the magnitude of the bias is. And so it remains, even with these quantitative methods, a heavily subjective process. So, I think the Executive Committee decided it made most sense to allow the panels to just draw their conclusions after looking at all the evidence, without necessarily using a quantitative technique for doing that.

DR. SOX: The best approach clearly is to

design a study that minimizes bias, rather than trying to measure it. Kathy?
DR. HELZSOUER: Yes. Along those lines, I guess if we're going to have something in here as a guideline, I agree it has to be more detailed. The ideal study to me is never consecutive patients. You have to have a (inaudible) that as you say, minimizes selection bias, and I think that's what you want to say. For example, particularly in a cancer setting, you may have everybody with advanced stage disease and it tells you nothing about early stage, and that's what you might need to know about an
.00069
evaluating diagnostic test. So I think it should say minimize selection bias and cover a wide range of presentations, as opposed to how it's written now.
DR. SOX: Yeah, I kind of think of this as a cartoon that's meant to get over a point, and suggestions like that are very helpful and --
DR. GARBER: Maybe we should call it better study versus usual, or rather than ideal, because Kathy is quite right, that is not an ideal study design.
MS. RICHNER: I want to also drive home the point about the process issue that Bob brought up, because I think that's a very important point, and we discussed that in the subcommittee guidelines and once again, was what are the instructions that we are going to give to the body that develops the technology assessment. And from my perspective, this kind of information should be what we would give any guidance for the technology assessment report, but not for essentially what our panel needs to do and that needs to be separated, and I want to make that point very strongly.
DR. SOX: I thought I heard you say this, we should tell the folks who prepare the evidence report to pay attention to these issues.
.00070
MS. RICHNER: Exactly. There should be a separate --
DR. SOX: But when we get around to evaluating the evidence report, I thought I heard you

say we should ignore it.
MS. RICHNER: No, not ignore them. Certainly we're going to be drawing up whatever the guidelines are for a robust technology assessment so we'll have a part of that, but this is supposed to be a recipe for how we evaluate the information that comes to us in a succinct manner, and this is too detailed essentially for what we need to do as a panel.
DR. SOX: Well, perhaps, but as you can see, it may make it more complicated trying to make an inference about effect on health outcome.
MS. RICHNER: What we're giving in evaluating the technology assessment is essentially given to an outside body to conduct, so those are almost separate guidelines than this.
DR. SOX: Well, hopefully there is concordance between what we ask the folks who make the evidence report to do, and the standards that we are going to use in trying to decide whether the evidence is adequate to measure test performance
.00071
accurately. Well, should we go on? Anything more on this one?
DR. BROOK: I'm going to raise the -- Hal, I need to raise the other side of this in a diagnostic test because now I'm really confused. When I looked at all these evidence reports, we are now beginning to break, we're moving towards the objectives, the way that we moved to the random appropriateness method 20 years ago, just to sort of basically start to break people into homogeneous groups of indications. And I don't know how far we are going to go down that with the evidence. We've moved very far down trying to evaluate the evidence, but for whom becomes the question. A 90 year old with a history of ovarian caner 20 years ago with what looks like a scar on a CT in the chest may be a very different person to do a -- and requires a different set of evidence, as you said, to look at a PET scan, versus somebody that has a much higher likelihood of having a pretest probability of having something there that's important.

And all of these, what I'm really asking now is what are we, when we are doing a diagnostic test, evaluating the evidence for? How fine groups of patients and indications are we going to break
.00072
this into. And when you look at the literature and these small studies, they're all broken into very small groups, and HCFA has to make a major decision of what to do. So if we say that anybody can, that the evidence is that -- are we going to say that the evidence is that anybody that has anything on CAT scan or MRI of the chest is fair game for a PET scan, or are we going to do this in more homogeneous indications, and I don't know the answer to that question, but I'm confused now, with a diagnostic test.
As I understand it, right now, if you do endoscopy for instance, a simple standardized test, you can do it on anybody and get paid, you can do it and you'll get paid at this moment. It doesn't matter. It's a standard procedure, everyone can get it, even if the person is asymptomatic they will get paid in the standard fee for service Medicare environment. So what I'm asking is what are we doing, and how does this evidence cut across the clinical homogeneous nature of patients?
DR. SOX: Well, I think we want to ask HCFA what information will be useful to them, which will probably vary from application to application. I'm a little mindful of the time. While
.00073
this is a good discussion -- pardon?
DR. CERQUERIA: I would just like to make one comment as sort of the clinician in the group. And the point that was just made, that a lot of, if you took these criteria and tried to apply them to things that we are currently doing out there and getting reimbursed for without question, I think we would have problems getting those things clearly approved. And so here we are, we're trying to come up with a prospective system that doesn't really factor in the whole issue of clinical judgment, clinical assessment of a particular patient with a

particular setting to make a decision for which test to use. And I am obviously not a health policy expert, which is what we've been talking about here, but just on a clinical basis, you know, it has been said that there have been two million PET studies done out there, and either we believe that those were all done fraudulently without any clear indications or we have to trust the fact there was some clinical judgment that went into making those decisions. You know, if we took this and went retrospectively back to what we are currently reimbursing for, would the things that we're reimbursing meet those standards? Because I think we

.00074

have to look at that if we are going to set up a prospective system.

DR. SOX: Thank you. Ellen, I think you were next.

DR. CERQUERIA: Well, aren't you going to follow up?

DR. FEIGAL: Yeah, I just want to give a concrete example. A pathologist reads a microscope slide, gives you a diagnosis. That in itself has value. It tells you a diagnosis. It may not impact how you, you know, the patient may have early stage or late stage cancer, or they may not. They may have adenocarcinoma of the lung or they may have squama cell carcinoma of the lung. It may not impact on how you treat that patient. You may still treat them with the same type of chemotherapy. But what I'm saying is, I think we have all generally accepted that what that pathologist is doing is of intrinsic value; it's helpful in terms of the patient, informing them what they have, and it's helpful in terms of the doctor, informing them of what the potential options might be. But I'm just saying that some of the things that we're doing today are setting a very high bar, and maybe ideal, but I don't know if it's where you want to go based on some of the

.00075

technology and useful items, useful tools we currently have. As you said, would some of the things that we currently use and that we currently

find useful meet your new bar?
DR. SOX: Well, I think it's implicit in this document that we're trying to set a higher bar than simply making a diagnosis, trying to ask whether that diagnosis is likely to lead to important health outcomes, and to try to make some inferences about the accuracy of that diagnosis. Alan?
DR. CERQUERIA: Well, in an abstract way I agree with that, but if you look at the practical applications of it, and you know, the fact again that you have all of these things that you're reimbursing for which we find medically important but wouldn't meet the standard that you're prospectively establishing.
DR. SOX: Well, maybe I should ask Sean or Dr. Kang to comment. We've been asked to give HCFA advice and we are trying to do it in a way that makes thoughtful use of the evidence that's out there, and in order to do that, we're following in the footsteps of other organizations that have tried to create a systematic approach to looking at evidence, and not simply do it in an ad hoc fashion.

.00076

I think Alan Garber was next, and then Ron.
DR. GARBER: Well, I would like to briefly answer that question and get back to Bob's comments, if I might.
We have had extensive discussions about this very issue in past meetings, and maybe the simplest way to state it is that this committee as Hal says, is advisory to HCFA. We do not make the coverage determinations. If something has gone through this process, that is the MCAC process, it is deemed to have met certain criteria, and those criteria are what we are trying to -- I should say, it has met certain standards of evidence and so on, which is what we're trying to hammer out here.
A negative determination I presume, by the MCAC process, does not automatically mean something is not covered. There are all kinds of other information that we presume HCFA will take into account, and although the coverage determination

process is still as I understand it, undergoing revision by HCFA, it is very likely that what MCAC says will not ordinarily be the final word, many other kinds of information will be taken into account.

.00077

Could I briefly address Bob's original question, or do you want to continue on with this?

DR. SOX: Yes please, briefly.

DR. GARBER: I think Bob was making two points. One is really about generalizability, that is, do these results apply to the Medicare population? I think, my understanding anyway is that these are supposed to be inserted into the interim guidelines which say that the panels do need to draw conclusions about whether the results apply in the Medicare population.

The second question was about how finely you divide the questions, and in the end that's not really the panel's charge, that's HCFA's charge in posing questions to the panels. And we hope that HCFA would make reasonable decisions about how to ask the questions, and they may solicit input from some of the panelists, but that's not really something that our guidelines should necessarily go into. We presume that HCFA figures out what question is relevant for their purposes and they pose that to the panel.

DR. BROOK: Could I just --

DR. SOX: Please respond.

DR. BROOK: Alan, I just want to -- the

.00078

panel, we made a decision not to look at all cancers today because we thought that was too big, but let's take a look at lung cancer. Would it have been better to have ten questions for ten of these different subgroups of patients with lung cancer and ask the question, is the evidence there to say that this test does something reliably and accurately, whether the tumor on MRI is this size, that size, or it's this way, peripheral, centrally? I don't know what are the critical questions, but I'm sure we could find those out quickly.

The bottom line that I'm asking is, you're right, and all I'm suggesting is that the evidence that we have spent a lot of time looking at one side of this, we haven't spent a lot of time looking at the framing of the population to which this is generalizable to. We have talked about over 65 and those things, but we've not talked about the clinical characteristics actually of the patients that actually come. That's the first thing.

I, by the way, want to support Hal, and I think our role here is to raise the bar of what we know so that we can practice better medicine in the future from what we've practiced in the past, when we've made a lot of mistakes because the evidence is

.00079

inadequate to make good clinical decisions. So I'm not afraid of raising the bar, I just want to make sure as we do this, we've got it right in how we raise that bar.

DR. TUNIS: Hal, may I just make one comment, sort of responsive to Dr. Cerqueria and Dr. Feigal. It's on the issue of, I think Ellen, you were sort of framing the issue of the pathologist looking at a slide and obviously, you know, buried in there, is there is some knowledge about the likelihood that a particular reading of a slide is in fact accurate versus not accurate, and then how that does or doesn't factor into the treatment, the diagnosis and then the treatment decisions. So the same issues really apply there. And I think what we're trying to get at here, and the panel is clearly wrestling with this in a helpful way, I think, is what is the minimum or the optimal amount of information that you need to have about the accuracy, you know, of the objective information that tells you about the performance of whether it's reading a slide or reading a PET scan, that allows one to make some kind of sensible decision about, you know, should this be broadly -- you know, is it ready essentially to be broadly available across the country, you know,

.00080

from a payment perspective.

So that's obviously a complicated both

policy and methodologic problem, which is what makes this difficult.

DR. FEIGAL: Not to belabor --

DR. SOX: Ellen, I'm going to cut off discussion of this issue now. I think we really do need advice on how to do this well, but the previous meetings of this group we have made a pretty firm resolve to try, as Bob said, raise the bar, let people know what the standards are that are going to lead to a smooth and easy assessment on our part and a positive recommendation about the quality of the evidence. So that's -- we have been through this in several previous meetings.

What we need to do now I think is to focus on this set of guidelines, because that's what we're going to try to use this afternoon to make some sense of this PET scan business. Manuel?

DR. CERQUERIA: But if we're creating a set of guidelines that aren't applicable to what's being done, and I agree that all of us have to have standards of what we do and we have to make certain that things are being done accurately, and you've set a high bar, but you're giving yourself an out saying

.00081

that you advise and HCFA makes decisions. And obviously I haven't been part of the Executive Committee discussions that have gone into this, but somehow you're creating a very abstract concept that doesn't really get at what is being done.

DR. SOX: This is work in progress and we're going to try it out today, and I suggest that you play along. Ron?

DR. DAVIS: Just to, I think maybe recapitulate where the committee is coming from, and I don't mean to extend this beyond where you want to go, Hal. But I am a physician and I'm all in favor of providing substantial deference to physicians' clinical judgment, but I think when we have a new technology, especially an expensive one, we have to set the bar somewhere. And if we simply had HCFA and Medicare cover everything that physicians believe is appropriate according to their clinical judgment, Medicare would probably be insolvent tomorrow. I

think we heard that two million PET studies have been done worldwide, half or so in the United States, and if each one costs about a thousand dollars per study, then that's a billion dollars right there, so I think we have an obligation to set the bar somewhere, and what we're struggling with is where to set it.

.00082

DR. FEIGAL: I do just want to make one comment, Hal, and that was, the issue of example pathology was not to say set the bar low, the issue was just to say there's intrinsic value in getting an accurate diagnosis, regardless if there's a treatment option or other treatment that you can give that patient. That was my only purpose in giving that as an example.

DR. SOX: Thank you. Well, I think we need to continue to discuss this framework, because we're going to use it this afternoon, and if we get it wrong we're going to potentially make wrong decisions about the technology this afternoon. So let's go on to -- go ahead, skip on to the next one. So, this next part of the process deals with trying to evaluate the possibility that two tests complement each other, and the starting point is that there's a big difference in the ability of let's say PET to pick up disease, as compared to CT scan, that's evidence that the two tests complement each other, PET is able to pick up more patients than CT. But if the two tests have fairly similar sensitivity, then the issue is whether PET might be picking up patients that CT is missing, and that's the reason for placing some emphasis on the issue of

.00083

trying to see if the two tests are complementary. And there's, in the Blue Cross/Blue Shield assessment, I think we all saw some evidence of trying to show that two tests were complementary, for example, CT and PET were discordant in a number of cases; most of the time according to what the evidence report stated, CT actually was correct and -- or PET was correct and CT was wrong, so that would be a clear evidence of two tests complementing each other. Any questions about this aspect of it? Bob?

DR. BROOK: I'm confused what you mean by complementary. I would have asked a further question, evaluate the possibility that the new test will replace the reference test. Now, if you mean by complementary, that that's what it is, but I mean, we used to test urine by testing urine for diabetes, as you know, by testing it as opposed to testing it, so -- by tasting it, I suppose is what I wanted to say. So now, so the question is here, should we -- what I'm worried about -- like I say, the science is fine, but should the first question be to evaluate the possibility, should we give HCFA an answer to he question, do we think with this group of patients that this new test will replace the current existing reference test?

.00084

DR. SOX: Well, that's on the top of our hierarchy of effect sizes, it's a breakthrough technology.

DR. BROOK: Okay, Hal, as long as we can get there, as long as this thing all becomes internally consistent.

DR. SOX: Okay. Frank?

DR. PAPATHEOFANIS: I have a concern about the use of sensitivity in this setting. To me, that harkens to sort of a screening approach to testing, and you've discussed notions of disease prevalence from the document. Why can't we use predictive value instead of sensitivity in these studies?

SPEAKER: Because not all these tests are going to be screening tests.

DR. SOX: Well, predictive value is a function both of the performance of the test and the population prevalence, whereas sensitivity and specificity are supposedly independent of the population prevalence. And as later on we get to looking at differences in post-test probability, but it's pretty well accepted, and I'm sure you know that you characterize a diagnostic test first by its sensitivity and specificity, and then calculate post-test probability, which is the same as

.00085

predictive value.

DR. PAPATHEOFANIS: Right. It just struck me that it seemed more of a screening sort of a framework.
DR. SOX: I don't think so. Bob?
DR. MURRAY: The comment on the slide, the second to the last paragraph that suggests that complementarity be identified by doing the two tests, the reference test and the test under consideration, as well as a diagnostic reference standard seems to me impractical or at least not often done. Usually what we see being done is the reference standard is the comparative test. So do I understand correctly that you're suggesting, or perhaps Alan can comment on this, that the complementary issue would require doing three tests, is that the suggestion?
DR. SOX: Alan.
DR. GARBER: Yeah, this only refers to the study setting and unfortunately as you're all aware, there are a few versions of this document floating around, and I think the one distributed today does not correspond to my final version, and let me read to you the change in the last version for the reference test.
It says, the reference test is a test
.00086
that's considered gold standard. Tests commonly used as reference tests are coronary angiography, et cetera. Then the last sentence says, reference tests can be interpreted more broadly to mean any method that is considered the definite basis for determining whether a disease or risk factor is truly present.
So in other words, it's -- yes, I know this was not distributed to you, and I'm sorry about that, but -- and these were mainly minor changes, but this is one that might help move the discussion along a little bit. The point is there has to be some method for ascertaining whether the disease or indication is present, and it's a very reasonable standard to actually have some form of confirming a certain result after testing for is present or not, and if you want to find whether a new test under consideration is better than an old test that is not

the reference standard, yes, you would use another method of ascertaining presence of disease in addition to the two tests.

DR. SOX: And in fact, many of the studies of PET scanning include CT, doing PET scanning, CT scanning, and then sampling, you know, biopsy or something of this sort, which is the gold standard or

.00087

reference standard. So it's not at all as infrequent as all that. The problem is that too often, the reports don't display the results in a way that allows you to see where there is a complementary character.

Well, it's 10:20 and we are going to take a ten, not a 15-minute break. We'll resume at 10:30.

(Break taken at 10:20 a.m.)

DR. SOX: We are going to continue to plow through the framework that we are going to use. Before we resume that discussion, however, Sean is going to make a few remarks.

DR. TUNIS: Yeah. An issue I just wanted to lay out clearly, and maybe differentiating some of the tasks that we are continuing to look forward to the EC's help with, and some tasks that we know are internal HCFA tasks and in the context of this framework and applying this framework to PET, or particularly in the context of this framework we're discussing now, we understand that what we're getting from the EC and are asking the EC for is sort of an optimal approach to looking at scientific evidence about diagnostic tests, hopefully to be adopted and applied this way, applied in a prospective fashion. As Ellen Feigal and others have pointed

.00088

out, given that this framework at least in the context of for HCFA is now in the process of development, the issue of how this should be applied retrospectively to technologies such as PET or other diagnostic technologies is a policy decision that we understand is on the shoulders of HCFA, and I just wanted to be explicit about that, that in the form that we're discussing it, this framework is intended for prospective application, and to what extent

elements of this framework are also determined to be useful and helpful in terms of making judgments about technologies now on the table, that's something that we are not asking the EC to help us for, we will be doing that in the context of policy development at HCFA, so I just wanted to be clear about that.

DR. SOX: Well, could you skip to about two transparencies ahead please? Now, we're running a little bit behind now. It's going to be important that we discuss this framework, so I urge you to think of good questions and say them succinctly. Okay.

So, the first part of the discussion was about trying to evaluate the evidence about the accuracy of a diagnostic test. The second part of the assessment is trying to make an inference about

.00089

health outcomes from knowing only the performance of the diagnostic test. And the first step of that is to calculate the post-test probability of the target disease for the test and the second is to try to make inferences about the potential effect of the probability of disease on management strategies and on health outcomes. So, next please.

Now, aside from making people maybe feel better about themselves, the main purpose of a diagnostic test is to move probabilities of disease around, to go from uncertainty to certainty about a diagnosis, that's what tests do. So we felt in sort of teeing up this strawman for the committee to digest and to modify that we would start by looking at the effect of the diagnostic test on the probability of disease by calculating the probability of disease by calculating the probability of disease after a positive test and a negative test for all possible values of the pretest probability. That would be the first step toward trying to decide whether the probabilities of disease after the test is close to some threshold for making a decision that might affect health outcomes. And so, we have proposed that at least for some instances, in fact providing a plot of pre versus post-test probability

.00090

can be helpful. So, any questions about this or comments about this as sort of a heuristic to help us think about the effect of the test on diagnostic certainty and sort of be the jumping off place for trying to decide whether the tests might affect management strategies that might affect health outcomes? Alan?

DR. GARBER: Just a point of clarification. I believe, Hal, that the lower solid line in that figure should be dashed, corresponding to the negative tests. I'm sorry. From here you can't tell that it's dashed. Is that dashed?

DR. SOX: Yeah, it's my laser printer. But basically the lines that look smooth are CT, and the lines that have got the little dots, that's PET scan in this example. And the results that indicate a negative test are the ones that are concave upwards, and the ones that indicate a positive test are concave downward.

Any comments from anybody else about this? I don't want to limit the discussion entirely to the panel. Let's move on then to the next transparency. This is the point which for me at least, it's very difficult to be very specific about how to proceed, and perhaps the most important thing to say

.00091

is the direction that you ought to be aiming, because I think the specifics will vary so much from clinical application to clinical application that any sort of general recipe is not going to work. But the basic idea is to try to evaluate potential impact of a post-test probability on a choice of management strategy, and then to infer whether that management strategy would in fact alter health outcome. Any comments about this? Confusion, disagreement? Well, in that case, let's go on.

And so that's basically a summary of the process. Now, we'll get a chance to go through this process this afternoon. I've tried to frame the evidence at least dealing principally with the Blue Cross/Blue Shield evidence in the context of this series of steps, and when we get to our first application of colorectal cancer, we will have a

chance to walk through this process with some transparencies that show the evidence as suggested by the Blue Cross/Blue Shield folks, so if it still looks a little vague now, I think it will be more specific when we actually apply it to a specific instance and so forth. So if there are no more comments then -- Sean?

DR. TUNIS: I'm wondering if this would be

.00092

a good time, if this is what we're going to use this afternoon, I know there's been some suggested modifications, you know, in the conversation so far, and I wonder if this would be a good time to try to summarize what those are to see if we want to actually change this before we try to use it, since this is what we started with before the discussion. But if you just want to use this the way it is this afternoon, that's fine, but if we want to change it, maybe we could actually between now and this afternoon make a different slide that you will use to do your evaluation. I know Bob and others had some suggestions about how to modify this, so maybe this would be a good time to make sure we have got those.

DR. FERGUSON: The panel does not have this summary, Hal, I guess you know that, I mean I don't have it, this series of slides that you just showed us.

DR. SOX: That's correct. I made it up yesterday morning.

DR. FERGUSON: Would that be useful for us to have if we're going to try to follow it?

DR. SOX: I think that would be a good idea, we could maybe have copies made during the lunch hour, a good suggestion, although the

.00093

transparencies I've prepared to guide us through some of the specific examples kind of repeat these points, so I think that will be helpful too. Ron?

DR. DAVIS: I just wanted to get another question out on to the table. If we modify question one like we were talking about earlier, so that we're looking for whether a test is just as accurate or more accurate than standard alternatives, then I

wonder if that would push us toward considering a modification for question two as well, where we would talk about health outcomes that would be as good as or better than health outcomes associated with other tests. And if we did that, then we get to that hierarchy that we've approved before, where we could have a test that would be as accurate, leading to a health outcome that is as good as the health outcome from another test, but all of that might be more comfortable or less risky to the patient than an alternative.

DR. SOX: Alan?

DR. GARBER: I like Ron's suggestion; I'm going to suggest an amendment though, which is instead of alternative test, alternative diagnostic strategy. And the reason for saying that is sometimes this will be additive to another series of

.00094

tests, sometimes it will be instead of another test, and we can encompass all of those things under the term diagnostic strategy.

DR. SOX: Bob?

DR. BROOK: The summary is I think much less a problem than the document. I think the summary as stands is perfect, except I would probably add a prior question or another question. I think these ought to be the things that they do. I think that there should be something like, that we also ought to describe the state of the evidence in relationship to the state of the practice, something like this that is more descriptive. This is all evaluative, and I think there probably needs to be a descriptive step that the panel ought to describe the state of the evidence relative to the state of the practice for those patients on whom they think these tests ought to be done, so something like that. But these -- I mean, there is nothing here. It says seek, it doesn't say is there evidence adequate to move (inaudible) clinically significant improvement. I like the wording in the summary. The summary, I think that's a great summary.

DR. SOX: Well, perhaps on that note we ought to stop the discussion and move on.

.00095
(Laughter.)
In response to Sean's question, my read is the main difference that emerged out of this morning's discussion was a change in the frame of reference, instead of improvement, we are talking about as good as or better, and I think that doesn't materially change the way we would use these guidelines this afternoon. Alan?
DR. GARBER: Well, there is one issue that will come up this afternoon. In some situations we will consider the diagnostic test instead of another diagnostic test, in which case that at least as good as applies. But does it not seem appropriate to ask that it improve, if it's to be used in addition to something, as compared to not doing anything at all? In other words, if the PET or another test is being considered instead of directly moving to some management strategy without any further diagnostic testing, in that case is it sufficient to say that it's at least as good as doing nothing?
DR. SOX: Why not, why wouldn't it be?
DR. GARBER: Well, if the diagnostic test adds no value compared to not doing any further testing and just moving on to treatment, are we prepared to say that that's sufficient to go ahead
.00096
and go through this whole apparatus if all we can say is it's no worse than not testing?
DR. SOX: I'm not sure. I'm not following you.
DR. GARBER: In some situations you will perform a diagnostic test after you have already performed a series of -- you're considering performing the PET or any other diagnostic test after you have already performed a series of tests, so at this point your clinical decision is do I get yet another test or do I not.
The alternative is not testing, it's not another test, it is not testing. In that case, is it sufficient to say that performing this diagnostic test is at least as good as doing nothing, that it's going directly to treatment without further

diagnostic testing of any kind? Maybe I'm not being clear. This is compared to a strategy where you don't do another test.
DR. SOX: So what we're trying to evaluate is whether it is, whether the test adds values compared with nothing?
DR. GARBER: With no further diagnostic testing. It is not another test that is the alternative under consideration. So is it sufficient
.00097
to say that it's at least as good as not doing a test, or does it have to be better than not testing?
DR. TUNIS: My sense of that is it would be, you know, useful, if you feel that you can come to the conclusion that in fact the test, you know, doesn't add information and therefore is no better than no additional test. That's a useful conclusion and I guess the decision about whether or not that test should be covered, you could leave to HCFA. I mean, it's enough for this committee to come to the conclusion, but I'm not sure if what you're asking is -- I mean, as long as you're clear about that conclusion, I think that's useful. Whether or not that means you decide that it should or shouldn't be available might not be where you want to go with the committee.
DR. SOX: I would like to move on now. We have an opportunity for public comment on discussion that you have heard today, and I ask that anybody who wishes to make a comment, please step to the microphone, identify yourself and whom you represent, and try to if you would, make your questions or comments concise and to the point so we can get as many people as possible up to the microphone.
MS. CONRAD: Let me call Peter Valk first
.00098
please, Dr. Peter Valk.
DR. VALK: I wanted to -- is this working now? Thank you. I wanted to say a couple of words about an aspect of technology evaluation that has only been touched on so far, but I'm sure will come up again before the day is out, and that is the use of randomized control trials in technology

evaluation. Evaluation of a new imaging technology by direct comparison with a standard technology in a single group of patients has been criticized because it isn't based on the randomized control trial. I think such criticism sometimes results from a failure to appreciate some of the differences between therapeutic and diagnostic procedures and as such, is not appropriate.

The randomized control trial or RCT is well established as the most valid means of comparing two therapeutic modalities. You cannot treat a single patient by two different methods at the same time, which means that to compare therapeutic modalities, you have to go to two different patient populations, and this immediately raises issues of random variations between the populations and the possibility of bias in allocating patients to the two study groups.

.00099

As you know, appropriate large patient numbers are used to try to reduce the effect of random differences and randomization is used to try to reduce bias. All of this when you put it together gives you a test that requires great resources in terms of money and manpower, and time.

When it comes to comparing two diagnostic technology, this does not have the same problems. You can in fact do two tests in one patient essentially at the same time. And all of the problems associated with studying two different populations completely go away. The number of patients that's needed is markedly reduced and so is the cost of the entire procedure.

Now, for evaluating diagnostic accuracy, this direct comparison method is in fact more accurate and less expensive than the RCT. But of course, it doesn't work if you want to go to direct evaluation of the effect of the imaging technology on patient outcome because now you will have to evaluate the outcome for both technologies separately, and you go back to the two patient population model, if indeed you think that direct evaluation of outcome by trial is even appropriate in this context of

diagnostic tests.

.00100

In practice, an RCT for evaluation of the effect of a diagnostic modality is in fact hard to do even if you consider it desirable. For example, in cancer management, it's rarely possible to initiate an RCT where the only difference between the two arms is a single diagnostic test. Even if you manage to initiate such a study, other problems follow. For example, the effect of a therapeutic modality in a blinded trial is independent of the physician, whereas the effect of a diagnostic modality is dependent on the physician's thinking, diagnostic thinking. It's also -- it's hard to blind a physician to the modality that's actually being used because as part of patient management it's frequently necessary to look at the images.

You really then can't expect that the physician will use data from a new and unfamiliar modality in exactly the same way as data from an established and familiar modality and in fact there is a large possibility there for physician bias, and there is really no effective way of taking care of this.

There are more problems still with the RCT in the diagnostic framework, but fortunately we don't often have to tackle the RCT or its problems because

.00101

of basic differences in the diagnostic and therapeutic fields. A therapeutic modality is intended to change patient outcome and this change must be evaluated by clinical trial, there is absolutely no other way to do it. A diagnostic modality has no direct effect on outcome whatsoever. Rather, it gives more precise, more accurate information on the presence and extent of disease which may then lead to change in therapeutic modality and eventually to change in patient outcome, but the actual change in outcome is not a result of the test itself.

In fact, you can look at the evaluation of a diagnostic modality in a given clinical situation for a particular indication as two questions. The

first is, how accurate is the modality for making the diagnosis. The second is, how important is the diagnosis for patient management and outcome. The first must be answered by trial, and it reflects a relationship between the technology and the disease. The second which reflects the disease and the therapeutic approaches that are available can be evaluated by decision analysis modeling, because these data must already be there in the published literature, having been gathered at the time the
.00102
approach was developed.
So I think in general, you can say that a randomized control trial of a diagnostic imaging technology only occasionally makes any sense at all and in fact, the rest of the time it should be avoided as much as possible because of its great cost, complications and sources of potential sources of bias, which generally are not recognized in such discussions. In fact, it's rather unfortunate that the general enthusiasm for the RCT, which has come from its success in the therapeutics sphere, has spread to all spheres, sometimes inappropriately, and I think that includes the diagnostic imaging sphere we're talking about.
DR. SOX: Thank you very much, Dr. Valk. Would anyone like to address questions to Dr. Valk or comment? Thank you very much.
MS. CONRAD: Jeff Kang, please.
DR. KANG: Mr. Chair, first of all I would just like to say -- my name is Jeff Kang and I am director of the Office of Clinical Standards and Quality at HCFA, and coverage is one of my four or five responsibilities, and I just wanted to say that I appreciate the Executive Committee today working so hard, and this is very important for us obviously, in
.00103
the future of coverage.
I just had actually one question for clarification on your interim guidelines, and if I could just have the first overhead, how I just wanted to make sure here on the last three bullets, is it your view here, or the Executive Committee's view

that these are all ands, so that the likelihood of an improved health outcome associated with increased diagnostic accuracy is when the treatment is effective and it doesn't benefit those people without disease, and imposes significant risk. And I wanted to be clear, because I think it's an and, but it suggests, your written material suggests an or, and so I just wanted to get a clarification on that.

DR. SOX: Alan, do you want to comment?

DR. GARBER: Jeff, thanks for the question. They are ands, at least I think that was our intent, except the third one is perhaps redundant with the second, so -- because it's implied that it may not benefit either because it doesn't work or because it imposes significant risk. But it's an and for the first two bullets.

DR. KANG: See, it's interesting, I was actually thinking the first and second bullet are redundant, and the first and third are the ands.

.00104

DR. GARBER: No, the first and second are two different populations, those with disease and those without disease.

DR. KANG: Okay. Let me deal with the first and third then, if I could have the next overhead, because I do think the first and third are ands also. If you look at treatments, treatments can be divided into effective treatments and risky treatments, and they could be both effective and risky, effective but not risky, not effective but risky, and then neither. And when you think about this, the issue of the likelihood that a test with improved accuracy or complementary information with improved accuracy, incremental, I'm talking about incremental accuracy, will change management or improve outcomes, that's certainly true in the first where you're really concerned about minimizing your false positives and false negatives, both for effective and risky treatments. But if you have a treatment now which is effective but not risky, there the clinician is faced with an issue of boy, I don't want any, I really want to minimize all of my false negatives. But if the test is only incrementally

changing your false negatives a little bit, they are going to ignore that second test and still treat.

.00105

Likewise, in the other scenario where it's not as effective and it's risky, so I just wanted to, I think those really are ands, and it's very important. But this is for your consideration and I just wanted to make sure of that clarification.

DR. SOX: Thank you very much. Any comment? We'll work on that to try to make it more explicit and more logically consistent. Is there anybody else from the audience who would like to comment before we move on to the next stage.

MS. CONRAD: You can use either of the aisle mikes, or the podium or the table.

DR. SOX: And please identify yourself.

MS. CONRAD: And you each have five minutes.

DR. SOX: Maximum. I'd prefer it to be less, because we probably should move on in about ten minutes to the next scheduled presentation.

MS. TESSER: My name is Ruth Tesser. I am an employee of CTI and I am a PET imaging center director and also past president of the Institute for Clinical PET. Due to the time frame between the announcement and this meeting, some of the surgeons and oncologists that would like to come were unable

.00106

to attend, so I've got four letters that I will try to read quickly for you, in support, and just discussing their feelings about, or their thoughts about broad coverage an PET.

The first is, I actually had a mix between academic centers and community based centers. The first is from a community based center that actually holds, they have at-risk contracts for patients, so they actually had to make decisions about whether they were going to be using PET or not. This is from Dr. Cargiano (phonetic). He is the director of Sutter Cancer Center in northern California. As the medical director of a large not for profit cancer center and medical oncologist with

eight years experience with PET scanning usage in oncology, my colleagues and I have seen thousands of patients with a wide range of cancers, and have found the PET scan to be invaluable and essential for correct treatment decisions. I and my colleagues in medical, surgical and radiation oncology strongly support broader coverage for PET scans similar to the process of coverage for CT and MRI scans. We have found PET scan to be useful in correctly staging a wide variety of cancers, including but not limited to breast cancer, pancreatic cancer, brain tumors,

.00107

hepatoma, and head and neck cancers. I am familiar with the PET scan literature, especially from the Northern California PET Imaging Center. These data are quite compelling and in my experience support broader coverage for PET usage. On a practical note, PET scan use may actually reduce costs associated with complex cancers, especially important in a capitated health care environment.

The second letter is from Dr. Thomas D'Amico, assistant professor of surgery, medical director of the clinical oncology services, and co-director of the thoracic oncology research lab at Duke University.

This letter is in support of broad coverage for positron emission tomography scanning in patients with known or suspected malignancies. I'm a practicing thoracic surgical oncologist in a large academic medical center as well as the medical director for oncology services within the Duke Comprehensive Cancer Center. As the literature demonstrates and our experience supports, PET scanning has made a tremendous impact on the practices of medical and surgical oncology. In addition to the accepted and supported indications for PET scanning, this technology is in fact useful

.00108

for virtually all patients in oncology and has been shown to improve the staging of cancer and decrease the overall costs of patient management in patients with suspected malignancies, and to decrease the overall cost of patient management. In patients with

suspected malignancies, a negative PET scan may curtail unnecessary follow-up and unneeded further scans to exclude malignancies. PET scans have the ability to address the primary tumor, to assess possible lymphatic involvement, to evaluate the entire body for potential occult metastasis, and to detect recurrence after treatment. For patients with occult metastatic involvement, a positive PET scan may prevent unnecessary exploratory surgery in a patient with unresectable disease. While all diagnostic procedures have their strengths and weaknesses, I strongly believe that while positron emission tomography is an invaluable study of patients with oncologic disorders, owning to its sensitivity, specificity, and the ability to evaluate the entire body. In our institution it has replaced galleon scanning for lymphoma, bone scans for metastatic lung, esophageal and breast cancer, and adjusts our treatment plan in a significant number of patients with all types of malignancies. Broad

.00109

support of PET scanning for patients with known versus suspected malignancies would improve the quality of care and by reducing the number of multiple other organ specific staging studies, have the ability to reduce overall costs. If you have any questions, please contact me.

You've got copies of each of these letters. The next letter is from Dr. Hilliard Sigler, chief of surgical oncology, professor of surgery, professor of immunology at Duke University. I'd like to take this opportunity to express some views concerning the clinical utilization of PET scans. My position at Duke University Medical Center is chief of surgical oncology and my university titles are professor of surgery and professor of immunology. Over the past several years, clinicians involved with neoplastic disorders have come to depend heavily upon MRI and CT scans. More recently we have evaluated the utilization of PET scans. My own experience with PET scans now numbers more than 300 clinical patients with patients being diagnosed with malignant

disorders facing potential major abdominal or thoracic operative procedures. We have determined that PET scans are 90 percent accurate in terms of

.00110

sensitivity for occult phacitis of neoplastic deposits. Often times we will alter our clinical management based on the findings of the PET scans. We have deferred radical neck dissections, pulmonary resections, hepatic resections, adrenalectomies, and partial bowel resections to remove occult neoplastic disease when PET scans define distant sites which render the patients not operative candidates but candidates for systemic chemotherapy and/or immunotherapy. If we can save patients major operative procedures, not only are we reducing health care costs, we are more accurately defining those patients who will benefit from surgical procedures and those who should be subjected to, should not be subjected to an unnecessary surgery because of distant disease not defined by CT and MRI scans.

DR. SOX: Excuse me, your time is up, so I wonder if you could just kind of hit the absolute minimum high points that you think you want to get across.

MS. TESSER: We have -- in your file also that you have, you have another letter from Dr. James Fleischman from University of Washington, who is professor of surgery, who goes through his points in a two-page document.

.00111

DR. SOX: Before you sit down, I wonder, does anybody on the panel want to comment about these letters and what you think of them, any take on this? Anybody want to say anything? Bob?

DR. MURRAY: Are any of the authors of those letters, have any of them published any of the studies that we have in our materials? That's a rather vague question, but they speak very strongly, but if they haven't published, that certainly has to be taken into consideration.

MS. TESSER: I can speak for the one physician in northern California, and he has not been an author, has he? The ones at Duke have been

authors, and I don't know if Dr. Fleischman has been an author.
DR. MURRAY: I'm sorry, the names of the ones who have authored?
MS. TESSER: Dr. Fleischman from Washington University, Dr. Sigler from Duke, and Dr. D'Amico from Duke have all been authors.
DR. SOX: Frank, did you have a comment?
DR. PAPATHEOFANIS: Just very quickly then, I think that it's important to include this sort of information obviously, because the folks that have written these letters have taken the time to do
.00112
so, should be acknowledged. Obviously they are thought leaders in their institutions and I think that what I was impressed with was just how specific they were in their comments and I appreciated that, rather than very broad general statements about PET being great or something. I think the letters offer very specific examples of where the technology is being used.
MS. TESSER: Well, this also speaks to the question that Dr. Brook was mentioning, that this has been a long history, that each of these physicians have had a long history in dealing with PET, it's not just over the past year, so I think that's important for the panel to know.
DR. BROOK: I would like to make one comment. The letters could have been far more useful if the authors had actually gone through and been more specific. It would have been interesting how many times they used them, did it replace any other test. It could have been much more quantitative, and for what kinds of patients, and did they really think that -- so they would stand by and actually state that if PET was approved they would have -- had they already moved to the point -- it would have been very interesting to know if they had moved to the point of
.00113
giving up some other tests.
MS. TESSER: Yeah. We had a --
DR. BROOK: I know the letters had to be drafted hurriedly because -- but if one's going to

evaluate one's clinical experience, I think there would be some nice guidelines to actually show that because actually when you go back quantitatively and look at some of this sometimes, it's based on a N of two or three or four, and it would be very interesting to know that in a little bit more specific detail.

MS. TESSER: Absolutely. All in time.

DR. SOX: A brief comment from Dr. Valk.

DR. VALK: (Inaudible comments; speaker did not go to a microphone.)

DR. TUNIS: I guess for, you know, for the Executive Committee to help us with later obviously is that, you know, as you're going through the framework that you're developing, to the extent that the conclusions you come to from applying the framework to the empirical evidence that we've got, to the extent to which that's consistent or inconsistent with the strong feelings, consistent feelings we hear expressed in this sort of letter, I think it would be very useful for us to hear you

.00114

discuss how those things should be reconciled with our own deliberations about this, because I have a sense that maybe the direction might be from reading some of the material that we will be hearing presented later that there maybe is more questions about the solidity of the empirical evidence and yet there is a fairly strong statement from the clinicians, these clinicians about the clear value of the technology.

DR. SOX: One thing I would like to suggest is that next time we announce we are going to evaluate something, we could also state on our web site description of the announcement, people are encouraged to comment, but please, and then give some suggestions about how to make those comments as focused and useful to the panel as possible.

DR. BROOK: Hal, can I just emphasize that again. I think we have a total disconnect between the hundreds of patients in the assessments and the millions of patients that have gotten this. And I don't for one have any sense of where the standard of

practice is right now, especially in organizations like Kaiser which would be at risk for actually doing these additional, have -- are there whole groups in the country that have replaced doing something with
.00115
PET, where are we. There is no summary of that kind of evidence and you get these letters, and it would be great if somehow the clinical evidence and the clinical standards could be put in a little different way.
DR. SOX: Thank you. Second commenter please. Actually, I'm sorry, Linda.
DR. BERGTHOLD: I don't know if this will be relevant to the second commentator, but you know, usually the letters that we get, particularly from those who would benefit from having HCFA covering this, is to sort of tell us all the good things, marvelous wonderful things that this treatment or tool will do. And I think in the future, I just want to echo what Bob said, sort of what I was going to say, is it would be extremely helpful for us to get from practitioners some sense of the relative merits of various diagnostic tools, and what are some of the weaknesses. Because in fact, we will find out what the weaknesses are probably somewhere else, and it would be very helpful if the practitioners could say, you know, we don't use this for everything, or it's not helpful in every case, these are the cases where it's most helpful.
DR. SOX: Thank you.
.00116
DR. WALL: I am Richard Wall. I'm chief of nuclear medicine and director of PET at Johns Hopkins, and also vice chairman of radiology. I have a conflict of interest; my son goes to Dartmouth and I'm collaborating on a PET project with one of your faculty because you don't have PET at Dartmouth.
(Laughter.)
I also gave a lecture there a couple weeks ago and received an honorarium, so I just wanted that background out. All right. So we have some respect for the institution.
In any case, as far as background, I am

also previous president of the Institute for Clinical PET and prior chair and member of the American Board of Nuclear Medicine, and certainly on our board for nuclear medicine exam we test on PET, and we think it's an important field. I personally have a 15-year experience with PET and was involved in some of the early studies, preclinical studies, showing the potential of some of the PET agents for tumor imaging, and also some of the earlier clinical studies showing the feasibility in humans of doing those studies.
And I wanted to say just one thing. I have been involved in therapeutic and diagnostic
.00117
studies because part of my interests lie in therapeutic radiopharmaceuticals, but the clear thing is, small studies are needed if the effects and powers of the test are large, and large studies are needed if you're trying to see small effects. So in most therapeutic studies, particularly in the cardiovascular area, to see an effect of a couple percent is very important, but you need thousands of patients to do it. To see a difference in sensitivity or accuracy between 50 percent and 80 percent, 50 percent and 90 percent, you need studies of 20 to 30 percent.
When we published in Radiology in 1994 that PET was more accurate than CT for staging lung cancer, that was a 23 patient study and the power of it was like .01 because the difference in performance was substantial. There have been about nine other studies, including the one recently published in the New England Journal of Medicine, showing the same sort of thing.
But I think you do have to keep in mind that if something works really well, you don't need the large numbers you typically need in therapeutic studies, and I know that this audience is aware of that, but it seems to come out repeatedly in PET. In
.00118
the FDA, this has been taken into consideration for instance in approval of some biologic drugs, wherein 60 to 100 patients have been sufficient for approval

of a drug, and you know, in well designed controlled studies.
I just wanted to say that in the studies we did showing PET to be more accurate than CT for instance in lung cancer, they were carefully designed so that we were blinded as to the results pathologically. We also blinded our referring physicians to the PET scan results because we didn't want to introduce bias. But in a study like that, you really can't look at management effects, because you are blinding the referring physician to the results of the new test. As Dr. Valk pointed out, the may not be confident in the new test and in the RIRB's approved view, it would have been inappropriate to use the results of a new and unproven test to change management.
So I think if you ask for accuracy of blinded tests that changing management in the same test is not possible. The problem we faced is once you prove the test is significantly more accurate in a prospective blinded study, it's hard to convince your referring physician to use the test that's less
.00119
accurate in a comparative study to show change in management. So you know, if it's really good, it's hard to go back to use something that's really bad. I mean, once you have driven a Lexus, you don't want to drive a Hyundai, and I mean, not to impugn certain manufacturers.
(Laughter.)
But as far as the view outside of this literature, which is admittedly not as big as we might like, major societies such as the Radiologic Society of North America last year chose PET as the topic for their plenary new horizons lecture. Similarly, the ASCO had a major focus on this, the American Society of Therapeutic Radiation Oncology had this as a major focus, and the Society of Nuclear Medicine. This is a major part of medical meetings and there's a huge growth; over half the abstracts in the Society of Nuclear Medicine are on PET.
At Michigan, where I was until I recently joined the faculty at Hopkins, in 1990 none of our

studies were clinical PET. Now 80 percent of our PET studies are clinical, of which 95 percent are oncologic. So there's been a huge growth, even in places that traditionally do research on PET in the brain, on the use of PET in oncology. At Hopkins, a

.00120

similar growth is occurring. Since I have been there, there has been about a doubling of clinical PET volume, and extensive use in a variety of diseases. I think on a national basis, 30 to 40 percent growth is being seen.

Now, just to look at major cancer centers, if you look at the top funded cancer centers, Memorial Sloan Kettering has gone from one PET scanner, now they're moving to four; Johns Hopkins has moved from one to two and now we're looking at three. M.D. Anderson has gone from one up to apparently three; Dana Farber has installed these. Major cancer centers are installing PET. They're probably not doing it because they just want to spend money, they're doing it because they want to use the technology for both the surgical and clinical conditions.

So, I think that outside of the literature, there's a lot of evidence to suggest there's a lot of growth in the use of PET in places that try to make rational medical decisions.

The other concern I have is, if you have a rare cancer, you are really in trouble by these criteria. We did work prospectively on testicular cancer, showing that PET worked very well. It took

.00121

seven years to acquire the data, and I think we had 23 patients. It just takes a long time. What if you have adrenal cancer? PET seems to work very well. 900 cases a year. They will never prove it to these standards, and I don't think this committee is, or at least I should say, the committee needs to be cognizant of the issue of low frequency tumors in questioning or determining coverage guidelines, because you just cannot get enough cases.

Particularly annoying to me, I got a call a couple nights ago on a patient with an adrenal

tumor and unfortunately doesn't qualify in general for coverage under Medicare guidelines. So it does come up on a daily basis. And some of the conditions like esophageal cancer, head and neck cancer, I think the evidence is rather strong that PET is superior and I would just simply say that I think if these were being done with a data management safety board that the DMSB would probably have stopped the studies because PET is superior to conventional methods.

DR. SOX: Thank you for your comments. Anybody want to respond or ask questions? Bob?

DR. BROOK: I'm really a little bit upset about your testimony. The technology assessments show that virtually all of the technologies are

.00122

single site studies, which I would interpret as zero cooperation across these esteemed institutions that we all fund, and I'm just really wondering why given the number of PET scanners everywhere and why the sample sizes of testicular cancer could not have been accumulated with cooperation among different investigators across site. And I really am a little concerned about, that small sample sizes for lung cancer works fine if you consider lung cancer a homogeneous entity. But as you know with subgroups, you need larger sample sizes to look at different subgroups of patients. And could you just fill me in on why the field operates in this single site single investigator manner in terms of producing evidence, so that we're in this quandary, and HCFA is in this quandary of what to do here?

DR. WALL: Well, I don't know if I can speak for the entire field but I can speak a little bit as to our own experience. I mean, we mounted and have recently just completed accrual of a prospective multicenter study for PET in staging breast cancer, and that's just under analysis. So with the NCI's support, a collaborative study across --

DR. BROOK: I understand that, but you started out by saying you've been in this field for

.00123

15 to 20 years.

DR. WALL: Well, I've been doing clinical

PET for 11 years.
DR. BROOK: Everyone has come in front and testified in front of us that this is not a new technology, and you have all these machines and all this stuff, and you see at least in the technology assessments so little, now you're beginning to see it, and one of our responses could be well, you guys sort of screwed up, so why don't we just wait another three years until all this multisite stuff gets done. And I'm really wondering why this stuff has not -- what is the impediment here, is it industry, is it the NIH, is it HCFA? What's this impediment that you couldn't mount a better scientific story here earlier and quicker? What happened here?
DR. WALL: Maybe Dr. Phelps would like to try to address that and I'm sure he will in a moment, but in lung cancer, solitary pulmonary nodule, the Institute for Clinical PET did mount a multicenter study and I think that was reported in ASCO in the last two years. But I think clearly, individual site studies were performed first to show proof of concept. The first proof of concept papers, for instance in breast were '91, melanoma '93, so it does
.00124
take a while after you have individual center studies showing efficacy, it does take a while to move those forward into clinical studies. From our own experience in breast, we had to move forward to do the multicenter studies. Dr. Phelps?
DR. SOX: Very briefly, Dr. Phelps, because we do need to move on.
DR. PHELPS: Very briefly. First of all, when CT and MR were introduced, CT very quickly went into clinical utilization, MR paused for a little bit and then went into clinical utilization; no clinical trials of any substance were even done. In PET we did not begin to look at the issues of clinical medicine; we began to do the basic science to develop biological assays and biochemical studies. We were interested in the basis of disease, not in clinical use. In fact, not until the late 1980s and the early '90s did clinical trials begin.
And I would also wait on your small sample

questions to give Sam and Ed the possibility. I would say that if you had ten places that did 100 studies a piece, that's a lot better than one place that does a thousand, in terms of randomizing out biases and variables. If you look at the literature, they come from institutions all over the world. They

.00125

are published not only in imaging, but primarily in nonimaging journals. So let's just be patient a bit and go out to the rest of the day and look at some of the evidence more carefully.

DR. SOX: Thank you, Dr. Phelps. We will now move on to hear from Patricia Love, who is going to describe the FDA approval of FDG PET.

MS. HALLIDAY: I had signed up in the beginning to do a public presentation. Do you want me to wait until the end?

MS. CONRAD: What's your name?

DR. SOX: Well, we are going to have another opportunity for comment later on. We want to hear from everybody; at the same time, we've got to have deliberation time at the end. That is the problem we've got.

MS. HALLIDAY: Inaudible.

DR. SOX: Why don't you -- we'll make sure that you get a chance at the second public comment period to be the first person. What's your name please?

MS. HALLIDAY: Sue Halliday.

DR. SOX: Sue Halliday, thank you.

Dr. Love?

DR. LOVE: Thank you very much. While

.00126

she's putting on the projector, my name is Patricia Love. I am director of the division of medical imaging and radiopharmaceutical drug products at the Center for Drug Evaluation and Research at the Food and Drug Administration. I do not have any financial relationship to any PET center.

As you know, the FDA as well as HCFA, has been considering PET products for a number of years in trying to determine exactly what we were going to do, and we grappled with some of the types of issues

that you've been discussing this morning. My comments today will be as brief as possible. Just quickly from a historic perspective, as was mentioned earlier, the FDA also considered PET as primarily a research tool earlier and it has moved into clinical practice over the last several years. In 1993, we did recognize the need to regulate PET, and there have been a number of approaches that were published in 1995, but also as well recognized, those approaches were not well received. And in 1997, with the Food and Drug Modernization Act, Section 121, the FDA was directed to withdraw specific prior documents and to develop approval procedures for the approval of PET drugs and in so doing, we considered several approaches that

.00127

already existed. One was using something called the 505.B.2, which is a literature approach to approving a product, and a J, which is a generic approach to a proven drug product. Also, the Agency was required to develop current good manufacturing requirements for the use of PET products, and we would allow the USP approach in the interim while we were doing these developments. The Agency was to consider relevant commercial and nonprofit differences, identify any and consider those that might be relevant. Certainly we were involving stakeholders. The Agency had two years to establish these developments at least as a preference, and there was another two years for implementation. Some things have been developed within that time, some things have not. Where we are right now is with the stakeholders as listed here, the Agency has been developing an approach to these different drugs, and I know today we are talking specifically about FDG, and the types of discussions that I will be summarizing today are available on the FDA web site. Under FDAMA there is a specific PET page, which includes various reviews, literature, various guidances and regulations, in relationship to the FDA

.00128

approach.

In discussing the issues with the PET community, our decision was to initially focus on various commonly used PET drugs and to develop other approaches later. In so doing, as you, we looked at what was available in the public literature. We also considered what the FDA already knew based on FDG approval for epilepsy at one clinical site and an old approval for sodium fluoride in the 1970s, using F-18 for bone imaging.

But now just then specifically, this is just a list of what was initially looked at as far as PET overall was concerned. The first set looked at FDG, ammonia, water and sodium fluoride, and we're currently looking at you F-Dopa. We sought guidance from (inaudible) biologics that was initially published in 1998 and a revision was published in June of this year. From the guidance for establishing clinical effectiveness, that guidance obviously describes what might be done prospectively from a standpoint of clinical trials that are under development, but it does contain a section for how one might approach a literature review and use literature to establish evidence of safety and effectiveness when other data and detailed trial data

.00129

are not available for us.

And some of the key points are just highlighted here. One is that we looked very specifically to insure that there are multiple studies. Sometimes, as was mentioned earlier, we do not have trials that are multicenter studies so in that situation we look to make sure there are a wide variety of studies with different authors, different investigators representing a prospective across the board.

We also look in the methods section for each clinical study to be sure that there is a prospective design that is detailed in the method section, that there is a full accounting of all patients that were involved, so we can look at both an intent to treat type of analysis as well as an all evaluable type of analysis and be able to make decisions about bias.

We looked to see whether or not the information in that clinical trial might be useful in consideration of the indication that we might be considering. We certainly recognize that clinical trials that are done and available in the literature are not necessarily done for the purpose of supporting an approval, so we have to look very

.00130

carefully at whether or not the information will support a labeled indication of proposed use and in so doing, we consider the clinical trial setting that was studied, are these patients as was mentioned, that are just being enrolled in a sequential manner, is there a particular question that is being asked, is this a screening study, is this a study that is going to be used just before one makes a major decisions to go forward with a biopsy or an invasive procedure, or a diagnostic or therapeutic study. And we look to see whether the end points that are identified in that clinical trial will be relevant for the proposed indication.

The medical imaging guidance discusses how that might be used in a diagnostic indications study. We look at what you have termed a reference standard, this is our standard of truth or gold standard that's used to establish the diagnosis. We certainly would like to have other controls also in the article but that's not always present, but we certainly at least require the presence of a truth standard.

The analytical plan for handling the images must be clearly described and that would include the discussion of blinding, how are blinded images used. Clearly we require blinded images for

.00131

the basis of our primary decisions in evaluation of the primary end points. We also want to see how the, what the statistical analysis that is identified in the literature and a discussion of sample size, is it relevant and how is that determined.

And then the results for the primary identified prospectively stated end point would need to be robust and based primarily on a prospective analysis, not a retrospective ad hoc analysis of the

data.
So in looking at the literature again, these are the drugs that were identified for PET and from now on my comments will specifically focus on FDG looking at myocardial indications and oncology, and how that led us to the approval process that the Agency published earlier this year. For the literature search for FDG for the myocardial indication, 632 articles were identified and 10 met the set of criteria that I just mentioned, criteria for review to determine whether or not they would lead to the type of indication that was being considered. And I might add, specifically the indication that we were considering in this context was whether or not FDG would be beneficial for the evaluation of myocardial viability, in that context.

.00132

And then for oncology, 150 articles were identified and 16 were identified as meeting the criteria for review.
Focusing on oncology for the moment, of those 16 articles, they involved at least 50 patients, pathology was the standard of truth; here's a statement of the doses, the ranges across the 16 articles. These arms evaluated a variety of different cancers, non-small cell cancer, colorectal, pancreatic, and others that you see listed. And there were a number of different metastatic sites involved in the different articles.
The articles also specifically were used in a clinical setting where there was an abnormality either already identified by a prior test and the patients were being imaged to seek a diagnosis, or the patients had an existing diagnosis of cancer and were being imaged for further workup or monitoring.
None of the 16 that we were reviewing in this context looked at FDG as a screening test in healthy asymptomatic patients.
Again, of the 16, 2 were considered adequate and well controlled in the Agency's traditional test. These were articles by Vallo in the Journal of Clinical Oncology, and Dr. Carr in

.00133

Blood, both in 1998. The other articles were considered supportive for our purposes. This is just a very brief summary of the two key articles, again, all having greater than 50 patients, histopathology, other modalities as controlled or blinded read. There were lesion criteria specifically identified in one article. Prospective design, the dose was identified and the data allowed us to do additional analyses to determine the sensitivity and specificity, looked at positive and negative predictive values and the like. Here is just a summary of the sensitivity and specificity results by a visual analysis and by an SUV analysis. This slide is derived from the primary presenter's review of the data.

The safety for FDG, certainly we already had a product that was approved so we had a great deal of safety data already available and the doses that were being used were in the same range as those that were available for the previously approved epilepsy indication. The Agency was also required on the basis of a pediatric rule in December 1999 to determine whether or not any of the information was relevant to the pediatric population, and it was determined that on the basis of the original FDG

.00134

approval for epilepsy, we had a great deal of information; that approval was also including a pediatric approval. We had no information on glucose utilization in the pediatric population and no data on radiation dose symmetry in pediatrics.

So, for oncology indication, it was determined that PET was approvable for assessing glucose metabolism to assist in evaluating malignancy in patients with known or suspected abnormalities found by other testing modalities or in patients with an existing diagnosis of cancer.

Well, how did we particularly arrive at that indication labeling? There was a radiopharmaceutical rule also that was derived from FDAMA, and although I'm not going to go over all the issues on this slide, specifically that rule included a discussion of indications, how one would evaluate

effectiveness and safety, and these data were clarified in the guidances that I mentioned earlier. One specific thing in the guidance is how we look at different indications and as mentioned earlier, there is a structural or an anatomic type of delineation that's usually more of a nonspecific characterization of a mass delineation features and the like. There are functional physiologic or

.00135

biochemical aspects of imaging. There's disease and pathology detection, and diagnostic or therapeutic management.

The Agency has often been asked about how do we relate this to the management relevance or clinical outcome, or clinical utility or benefit in all of the indications. From our perspective, all of these have clinical utility and benefit, but it's in the context of the clinical study. As was mentioned earlier, there is very definite information that can be derived in the use and evaluation of the patients if you're simply looking at structure and delineating an outcome or an outline of a given mass. Functional information, our classic example is an ejection fraction or renal function. Again, this type of information has great benefit from a diagnostic utility without necessarily knowing the specific disease that may have caused an abnormality and an ejection fraction.

For disease or pathology detection, we're looking more at the traditional diagnostic, what is the cancer, what is the pathology. Disease would be a more specific type of an assessment to us and pathology a bit more general in the detection sense. And then for diagnostic or therapeutic

.00136

management, from an Agency perspective we're looking at the actual labeled indication that's printed in the package insert, so a diagnostic change might be one where given this test, the result, one can make a specific determination in a sequential diagnostic algorithm that one might be using, or if you're looking at whether or not a patient might have a different therapeutic intervention or perhaps may or

may not respond to a coronary artery bypass, this would the type of labeled therapeutic management indication. So it's a gradation or degree of how the different types of indications are used in a clinical benefit scenario and how the indication is actually construed on the package insert.
So, this indication for oncology then is somewhat of a composite. It has glucose metabolism so this is a functional utility; as well as use in a particular setting for evaluation of malignancy, this is a pathology detection type in a setting of patients who have suspected abnormalities by other modalities or existing diagnosis of cancer. The clinical trials labeling section of the package insert does describe what we generally know about the sensitivity and specificity based on these trial analyses, and it gives some of the caveats on false

.00137

positives and false negatives that were mentioned earlier, and specifically addresses the inflammatory processes, fungal infections and others that might need to be considered in the overall assessment and use of this particular imaging modality.
Just shifting briefly to myocardial viability, in an analogous fashion we reviewed 10 particular articles that met the criteria. These articles were actually a bit smaller but they were all very consistent. They all looked at hibernating myocardium or viability assessments in comparison to a functional outcome of left ventricular function by another measurement. The function was evaluated before and after coronary artery bypass. Some of the other articles also among these 10 looked at other types of clinical utility established by perfusion, other approved perfusion agents.
This is just a summary of the 10 articles, the sample sizes across the board for each individual patient as well as a segment analysis of the heart.
DR. SOX: Excuse me. I wonder, could you cut to the chase for the myocardial, since we're not going to be really discussing that today?
DR. LOVE: Okay, fine. The indication, coronary artery disease and left ventricular

.00138
dysfunction, again looking at glucose metabolism, and used with other myocardial perfusion agents to identify myocardium with reversible loss of systolic function. And again, it had false positives and negatives that are also available in the clinical trial section.
For our approval process then, developers of PET agents and PET centers are encouraged to submit a 505.B.2 or a 505.J application. This would be based on the chemistry. One of our concerns was that there are various FDG products available across the many centers, and we needed a way to insure that all centers are producing the same drug product, so that would be based upon the chemistry for each particular site. If the chemistry is identical to the one approved NDA that's already available, then someone would submit a J application; if there were slight differences, then one might submit a 505.B.2 application to document any chemistry issues. New clinical studies would not be needed. The FDA published a Federal Register (reporter changed paper while tap was changed) gave sample information for the CNC and specific formats for the labeling that included all details that would be necessary.
And I will stop there for the relevance of
.00139
this particular discussion.
DR. SOX: Thank you very much. Does anybody have any questions they would like to address to Dr. Love? Sean?
DR. TUNIS: Correct me if this understanding is wrong, but in terms of the -- so the FDA's determination of the effectiveness of FDG PET for broadly in oncology was based on the two studies that you mentioned and the 14 supporting studies, and that is the basis for the determination of broad conclusion about effectiveness of FDG PET in oncology?
DR. LOVE: Yes, the two key articles and the other 14 studies which were very consistent in their results.
DR. SOX: And did I understand correctly

that there are really only two articles that met all of your criteria to really be considered first rate evidence?

DR. LOVE: The two articles that met the bulk of the information. There were other articles that may have had pluses and minuses but again, we looked at the overall weight of the evidence from the other 14 to make sure they were going in the same direction. The Agency's standard at the moment for

.00140

safety and effectiveness is generally two adequate and well controlled trials, and so we did have two from that perspective.

There was a discussion about randomization. Although we don't necessarily look at randomization for each individual patient, we look at randomization of the blinded imaging reading protocol to make sure that that's sufficient from a standpoint of eliminating bias.

DR. SOX: Okay. Any other questions from the panel? In that case, we are going to move on. The next topic is presentation of the coverage request, and we are going to start with Dr. R. Edward Coleman, who is going to present on lung cancer and colorectal cancer. And Dr. Coleman, since we're not going to be discussing lymphoma, I hope you will just not discuss lymphoma.

DR. COLEMAN: No problem. I am Ed Coleman. I am professor of radiology and director of nuclear medicine at Duke University Medical Center and am here representing the Institute for Clinical PET and the Society of Nuclear Medicine, American College of Nuclear Physicians. I will keep my comments short, I know we're running behind schedule. On this slide I just want to make two

.00141

points. One is, there are now several imaging instruments out there, Mike estimated 400 today. If there's 800 worldwide, the United States has slightly over 50 percent of those. PET has become a routine study in nuclear medicine. In institutions like Duke where we have been doing PET for a while, it's no different than ordering a bone scan than ordering a

PET scan.
Mike went through that we're imaging biology, we're measuring function, and we've had a high accuracy for many diseases. We're using fluorodeoxyglucose; it has a 110-minute half life, cleared from the blood like glucose, once phosphorylated, not further metabolized. FDG is readily available commercially now; we no longer have to have the cyclotron in our facilities, it can be purchased just like a bone scanning tracer.
PET in lung canter has been shown to provide information in the evaluation of focal pulmonary opacity, staging of lung cancer, and evaluation of the effect of therapy, and I'll go through those indications.
Start off with the case example. This is a 50ish year old lady, admitted to the hospital for a gynecologic problem, a benign disorder, had this

.00142

chest x-ray, has a pulmonary nodule in the right upper lobe. The next procedure that was done was a CT scan; it's an indeterminate pulmonary nodule and the CT scan could not determine if it were benign or malignant. Interestingly, the next procedure that was ordered in this patient was a radionuclide bone scan, thinking that it was a very high likelihood that this patient was going to have cancer, and we saw an abnormality on the bone scan in the left iliac crest, which would be worrisome for malignancy but certainly not diagnostic. Got a plain film, it did not show any lesion.
Then a PET scan was ordered and these are the images of the chest, showing on this posterior coronal cut, this pulmonary nodule which was cancer, multiple lymph nodes within the chest, which were not seen as abnormal on the CT but were involved with disease, and multiple vertebral body abnormalities which were not seen on the bone scan. And here shows a sagittal cut showing the multiple bony sternal vertebral body mediastinal disease that had not been previously suspected or detected by the conventional imaging modalities.
Michael Gould and colleagues from Stanford

and the VA at Palo Alto Health Systems, have recently
.00143
presented at a chest meeting a meta-analysis of PET for diagnosis of pulmonary nodules of mass lesions. Here they found that 34 studies met their preestablished criteria for inclusion, about 1400 nodules of mass lesions, the maximum joint sensitivity and specificity, which is the upper left-hand point on the ROC curve, which has a relationship with the area under the curve, was 91.2 percent, with a sensitivity of 97 percent, specificity of 80 percent in that population. PET has been shown to be very accurate in staging the mediastinum. Dr. Sox showed some data earlier today. Here's an example that by CT scanning the mediastinum was negative. CT looks at size of nodes, the PET looks at the metabolism within the nodes, looks at the biology of the disease, and here we can clearly see two abnormalities within the mediastinum on this coronal as well as on the sagittal images. The PET information is not only used to determine if there is disease in the mediastinum, it's used by the surgeons to direct their mediastinoscopy or lymph node sampling if they can't get to the lymph nodes by mediastinoscopy. In patients who cannot undergo mediastinoscopy for contraindications for mediastinoscopy, the PET then
.00144
will be used in the management of that patient. Dr. Wall and his colleagues had a meta-analysis on mediastinal staging published last year in Radiology. PET on 514 patients, CT scan 2000 patients, and you can see the 19 percent better sensitivity, about 91 versus 77 percent on the specificity.
There have been several studies looking at PET in staging the whole body. This is a study that we did from Duke that was published last year; there was a very similar study published in the New England Journal of Medicine earlier this year. There's about six or seven studies out there now on around 100 patients who have had PET scans, chest CT, bone scans, and what we found is very similar to the other

studies, that the PET is more accurate than conventional imaging, here 83 patients versus 65 patients. Nine patients had metastases detected only by the PET scan and furthermore, 10 patients who had suspected metastases by conventional imaging did not have metastases by PET or subsequently by biopsy or clinical follow-up. So it upstages some patients, downstages others, but it puts them into the right stage.

Looking at the effects of therapy and

.00145

looking at prognosis after therapy, 113 patients with non-small cell lung cancer, had PET after initial therapy, 100 patients had positive PET scans, median survival 12 months; 13 patients negative PET scans, 11 patients alive, median follow-up of 34 months. So it's able to stratify the patients after their therapy for their lung cancer.

Another study in the European Respiratory Journal, 126 patients, non-small cell lung cancer, Stage I to Stage III-B, studied before and after therapy. 58 with curative therapy, 68 percent with palliative therapy; follow-up period was 8 to 40 months. And in this series, PET was very accurate in determining who had residual disease and who did not, and PET correctly identifies responsive therapy in 121 out of the 126 patients.

So there are several studies now not only looking at the diagnosis and staging, but looking at the affects of therapy.

There is data on looking at colorectal cancer. A lot of the data has been in lung cancer, we're using that as a model. There's more data in lung cancer because it has been paid for a longer period of time than other indications. If we're not being able to have these studies paid for, it's very

.00146

difficult to gather the data in these patients.

Here's a patient with a known colorectal cancer, has a low attenuation lesion in the liver, a single lesion thought to be an operative candidate. A certain percentage of these patients can be cured by surgery. We did the PET scan and there were

multiple lesions. And there's several studies showing that the PET scan is more accurate than the CT scan in detecting metastases in the liver from colorectal cancer, and not only did this patient have multiple liver metastases, but had multiple para-aortic nodes outside the liver. Again, the multiple liver metastases and the disease outside the liver would make this patient not be an operative candidate.

In patients who have colorectal cancer, rising CEA with negative CT scan, we're able to identify metastases in a high percentage of those patients, and this is an example of such a patient. These are lymph nodes within the abdomen. Frequently if you go back on the CT scan, you may see the abnormality that was thought to be nonyl pacified bowel, but it could not be made on the CT scan and the diagnosis is made on the PET scan.

The group from UCLA has done a

.00147

meta-analysis of PET in recurrent colorectal cancer. 281 patients, had a 97 percent sensitivity, 76 percent specificity, and then in 7 studies had a change of management in 29 patients. There's now an article in the literature that has been accepted in the surgery literature showing the cost effectiveness of PET in evaluating patients with recurrent colorectal cancer.

The group from Washington University knew that I was going to be presenting here and sent this slide to me from the surgery group there that's in the Annals of Surgery, it's in press, and what they have found in a group of 43 patients who were being evaluated for metastatic disease and thought to have a single metastasis and to be a surgical candidate had a PET scan, and the PET scan demonstrated disease outside of the liver in 7 of those patients, and they did not undergo surgery. Overall, three-year survival using Kassen Meyer plots was 77 percent in those patients compared to 30 to 64 percent by conventional methods, and the three-year disease free survival is 40 percent, and again, 15 to 28 percent by the conventional method.

Well, we have relooked at the data in the literature, Sam Gambhir will talk more about the data
.00148
this afternoon, but now we have data on 24,000 patients in 643 studies in the literature. The sensitivity of PET overall has been 84 percent, specificity 88 percent, and change in management 32 percent.
I should say, and make the panel aware that the data that had been submitted in the original document to HCFA that came from UCLA and Duke was at the request of Dr. Kang in a meeting that Dr. Phelps and I had with him in March or April of last year, and asked us to summarize the literature and what was in the literature, and to give some intermediate outcome. That data was not submitted as a meta-analysis and was not meant to be used for that purpose. It was a survey of the literature. He asked us to include abstracts, which we did, and we clearly identified, as well as a survey of the literature. And this is an extension of that data set.
Then if you look at the patients that had both CT and PET scans, the overall sensitivity of PET was 85 percent, CT 66 percent, specificity 89 and 76. And again, these numbers are no different than the population who had reported a PET without having a CT scan, but you can see the data show that the PET scan
.00149
is more accurate than the CT scan. Furthermore, the change in management in patients who had both CT and PET was 31 percent, again, no difference in those that had the PET alone.
So, in summary, PET is a molecular imaging technique, it images biology. It's accurate detection of multiple disease and it impacts patient management, and I will stop there.
DR. SOX: Thank you very much, Dr. Coleman. Now there's a bit of time for questions and comments. Ron?
DR. DAVIS: Dr. Coleman, you mentioned at the beginning that PET is considered a routine study in nuclear medicine. Can you give me an idea of who

pays for it?
DR. COLEMAN: Well, Medicare pays for solitary pulmonary nodules, staging of -- initial staging of lung cancer, detection of recurrent malignant melanoma, detection of recurrent colorectal cancer with a rising CDA, and staging and restaging of lymphoma. Other third party payers have policies to pay for at least those indications and generally more than that, so third party payers and Medicare pay for most of the studies that we're now doing. A lot of the patients who could benefit from the PET
.00150
scan are not being studied because there is not policies for reimbursement.
DR. DAVIS: Thank you.
DR. COLEMAN: We do have several patients who do pay on their own, and I certainly should include that. That's, I don't know, that's probably about 15 to 20 percent of our patients will pay because they want to get the study done, their surgeon or physician thinks it's necessary.
DR. CERQUERIA: Dr. Coleman, I wonder if you could comment, we've heard different criteria used for selecting studies in the literature, and our reviewers must be doing a very bad job. Dr. Love told us that of 150 published studies, only two were appropriate. The New England Medical Center reviewed the data that you presented and really cut it down. What do you think is a reasonable criteria, and why is there such a big difference between what some find acceptable and others don't?
DR. COLEMAN: Well, I think it depends on what are your criteria that you're setting for the publications and you know, just the size of patients, how the patients get into the study, the way the studies are read. And if you take the extremely tight situation where you have to have hundreds of
.00151
patients and blinded readings, there just hasn't been a large number of studies performed like that at this point in time. There certainly are some, you know, there are several multicenter studies in the literature, not a huge number at this point in time.

A lot of that, there just hasn't been the money to get these studies organized and get them performed.
DR. CERQUERIA: So does that mean that all the stuff that's out there that doesn't meet the criteria is worthless, or is there any value in those studies, and what is the value?
DR. COLEMAN: Well certainly there is a lot of value in that, and I think Sam Gambhir this afternoon will be going through that and will be going through the document in more detail, and will address that probably better than I can.
DR. SOX: Frank?
DR. PAPATHEOFANIS: Dr. Coleman, several of the letters that were read by Ruth Tesser were from clinicians at Duke, and I just wanted to go back to the Duke experience, and if you could be a little more specific, I know Bob Brook stepped out at sort of a good time, but one of his criticisms were if those letters could have been more specific, they would have been more helpful. So one question is,

.00152

can you be a little more specific since these are your colleagues?
The other is, from your vantage point at the S&M, RSNA and so forth, what's going on out there? I mean clinically, are these protocols just being used in general, or are there any limitations?
DR. COLEMAN: Well, speaking for my colleagues, the one, Dr. Hilliard Sigler is a surgical oncologist who deals primarily with melanoma, he now has two publications in the literature and has shown that PET is more accurate than CT scanning, and it changed management from surgical to medical management which he mentioned in the letter, in about 15 to 20 percent of the patients. He has another abstract on 300 patients showing very similar data.
Dr. Tommy D'Amico, who is a surgical oncologist, uses the PET scan in evaluating the patients for metastatic disease to make sure they're surgical candidates. And as he said, he no longer gets bone scans or abdominal CTs in his patients; they are staged with a chest CT and a PET scan to

determine if they are operative candidates. So it has decreased the utilization of other procedures. I should also mention that Dr. Sigler uses

.00153

the PET scan as the surveillance procedure. He no longer follows these patients with CT scans. We're starting to see that with our lymphoma patients; the PET scan is being used as a surveillance procedure and not repeating the CT scan.
Was there another question?
DR. PAPATHEOFANIS: Just a broader comment on practice in general.
DR. COLEMAN: Well, I think that we're seeing the type of practice that we've been developing at Duke over the last four or five years extended throughout the United States. We are seeing that PET is being more widely utilized, more widely utilized in primarily the malignancies that we're getting reimbursed for, and we're seeing it being used I think very effectively in the management of patients and changing patient management.
DR. SOX: Are there any other comments before we move on. In that case, thank you very much, Dr. Coleman. We'll now move on to a discussion of Alzheimer's disease by Dr. Gary Small.
DR. SMALL: Thank you. Let me begin with my conflicts. First, I do not have any financial relationships with PET centers. I am a consultant to several companies that make drugs for Alzheimer's

.00154

disease, including Jansen, Pfizer, AZI and Navartis. I'm a geriatric psychiatrist at the UCLA School of Medicine where I am also a professor of psychiatry. I direct the Center on Aging there, and I'm a clinician and clinical researcher, and I'm going to talk about the use of PET for evaluation of dementia. If we could just move the slide over and maybe bring the lights down a bit.
To begin with, a couple of points that 8 percent of people 65 and over have dementia, 25 percent over the age of 75. Two-thirds of the cases are eventually diagnosed as Alzheimer's disease by autopsy. The annual estimated cost in the United

States, if you include both directs and indirects, is over $100 million. Most cases go unrecognized. And the accuracy of clinical diagnosis can be as low as 60 percent.

So, we know that Alzheimer's is prevalent, it's costly, but it can be treated, especially in the early stages. The current approach to dementia diagnosis involves multiple often costly assessments performed over the years, yet PET provides early positive differential diagnosis for Alzheimer's and other dementias. In fact, the classic Alzheimer's PET pattern will appear years before the disease can

.00155

be confirmed clinically. In fact, we have found that over 90 percent of the cases can be, are accurate with PET three years before the clinical diagnosis can be established, and I will show those data in just a moment.

So if we look at a differential diagnosis of dementia, we find that it often involves these multiple clinical exams over the years, CT and MRI are normal, or show nonspecific atrophy or focal lesions, but they fail to provide a positive diagnosis of Alzheimer's disease. Despite the fact that CT and MRI rarely help in the differential diagnosis, we know they're reimbursed. And in fact, CT and MRI can actually reduce diagnostic accuracy because of the high rate of Alzheimer's with incidental infarcts and the low rate of true vascular dementia.

Here we see some examples of what the PET scan can show in various dementias. First, in a normal person you see normal glucose uptake in their gray matter and the deeper structures. With Alzheimer's there's a typical pattern of parietal deficits. Vascular dementia, you see both cortical and subcortical deficits. With frontal temporal dementia it's a fixed disease, there's a frontal

.00156

dementia, and with Huntington's dementia, there is loss of caudate metabolism. In all of these cases, except for vascular dementia, CT and MRI are normal, or they show nonspecific findings.

Now this week, John Hoffman and colleagues published a paper in the Journal of Nuclear Medicine where they looked at 22 patients with dementia, and they found that PET provided greater sensitivity and specificity in predicting neuropathological diagnosis than conventional clinical examinations, but our group wanted to expand the sample size and also include several centers, so Dan Silverman got together a consortium of clinical facilities that contributed autopsy and FDG PET data.

The mean follow-up after PET scan was about three years. Of 284 scans, we had neuropathological data on 138 of them; the other 146, we had longitudinal clinical follow-up and we classified the scans according to whether there was a progressive nondegenerative dementia. And all these assessments or classifications were made with the PET reader blinded to the neuropathological diagnosis and also, the neuropathologists were blinded to the outcome or the results of the PET scans.

And here are the results. First, looking

.00157

at the accuracy of PET for assessing presence or absence of Alzheimer's disease, and we have the Alzheimer's disease, whether Alzheimer's disease was present on PET yes or no, and whether Alzheimer's diseases was found on autopsy, and here you have the result of the two-by-two table so that one can see that in this study there were 6 false negatives and 30 false positives, yielding a sensitivity of 94 percent, specificity of 73 percent, and overall accuracy of 88 percent.

If you look at the presence on PET of any neurodegenerative disease, not just Alzheimer's disease, and compare that with the autopsy results, then we find sensitivity of 94 percent, specificity of 78 percent, an overall accuracy of 92 percent.

Now looking at the longitudinal clinical data, if we have here the progression predicted by PET versus the clinical progression documented, clinical outcome, we have a sensitivity of 91 percent, specificity of 75 percent, and overall accuracy of 84 percent.

If we put all the data together, the overall accuracy, we see all over 200 some odd patients, we have sensitivity of 93 percent, specificity at 76 percent, and overall accuracy of 88 .00158 percent.

We know that as the disease progresses, we see an increase in the deficit in the parietal area, the temporal area, and frontal deficits with sparing of the sensory motor strip in the deeper structures as well as the occipital area, the visual cortex. Late stage Alzheimer PET scans look very much like what we see in children. We have, as Dr. Phelps mentioned earlier, we've looked at people who don't have dementia, with very mild memory complaints, and we combine the PET information with information on APO-E4 genetic risks, we can actually see these patterns in people many years and even a decade or more before they reach the age of onset of dementia. So there is tremendous added value in early diagnosis. We can identify candidates for treatment intervention before there is extensive neuronal loss, we can begin a therapy early on, and have an effect not just on cognitive function, but also overall activities of daily living. We can save costs by avoiding of multiple diagnostic evaluations that are noncontributory, and also people can plan for their futures while their mental faculties are intact.

We now have several medications that are .00159 available that are effective for Alzheimer's disease, the cholinesterase inhibitors have been shown to improve memory and other cognitive functions, they stabilize the disease, they delay functional decline, people maintain autonomy, they stay in the community longer, and they have a positive benefit on care giver burden.

There are many different studies I could show you, but this is an example of one of the double blind placebo control trials comparing an active drug versus placebo in people with Alzheimer's disease mild to moderate, and so as you get higher up on the

vertical axis that means better cognizant performance, and this is time, and here you have the active drug group doing better than the placebo group. And what is interesting about this study, and there are others from other medications available, that after six months, the investigators put all patients on active drugs, so you have an improvement in the previous placebo group, but the improvement never quite gets up to the level that patients might have gotten to had they started six months earlier, and you see that difference is sustained out to 12 months. This is just the projected placebo decline, if they did not start drug.

.00160

Now we can talk about methodological issues such as dropouts during this first six months of treatment and that this is a later stage disease, but one obvious explanation is if you wait to treat people, you're going to lose ground.

So in summary, use of FDG PET for dementia diagnosis will improve early diagnostic accuracy, it will increase recognition of Alzheimer's disease and other neurodegenerative dementias. It will remove multiple years of ambiguity for patients and physicians, and it will facilitate earlier treatment leading to disease stabilization and improved quality of life. Thank you very much for your attention.

DR. SOX: Thank you very much, Dr. Small.

Opportunity for comment or questions from the panel? Yes, Kathy?

DR. HELZSOUER: Just one comment.

Although not an expert, I have reviewed a little bit in this area, and your clinical diagnostic accuracy I believe is an underestimate; there's only a few well designed studies that have looked specifically at well applied criteria with pathologic examination, I believe show much higher than 60 percent diagnostic accuracy.

DR. SMALL: That's true. There is

.00161

variability in the clinical diagnostic accuracy and it can range actually from as low as 55 percent to as high as 90 percent. It depends on the setting and

the actual criteria.
DR. HELZSOUER: Right. And I think that some studies suggest that if you apply them correctly, it's higher. The other comment maybe you could make is regarding specificity, which is 73 percent, which makes a fair amount of false positive test results, and with a fairly devastating disease diagnosis of Alzheimer's that despite some evidence you showed through therapy, it's not that striking for many people, the benefit. So, you have a 25 to 28 false positive rate for PET scanning.
DR. SMALL: You have a similar problem with specificity with clinical examination. In fact, I didn't present it here, but we looked at, we had in our neuropathological sample, we had clinical diagnoses on about 60 percent of the cases, and that we found that in fact, not just was sensitivity better but also specificity as well, compared to the clinical examination. So specificity is an issue both with the clinical exam as well as with the PET diagnosis.
DR. SOX: Yes?

.00162

DR. CERQUERIA: You presented some data in terms of treatment and response, but is there anything in the literature that suggests that PET would be useful to identify those people who would respond to treatment or benefit from treatment?
DR. SMALL: Right now we haven't done the studies in that way, that is, to use PET as a predictor of response. I think at this stage we want to do that and we are beginning to look at that. At this stage we're looking at PET as an accurate early diagnostic indicator. One of the big issues is that so many cases go unrecognized and untreated. There are many people where there is a stigma about having Alzheimer's disease, they avoid treatment. There's lack of knowledge among physicians. And when we have this accurate early diagnosis, people get treatment earlier.
DR. SOX: John?
DR. FERGUSON: Yes. Dr. Small, I think most neurologists anyway will use the CT and MRI, or

MRI, generally to rule out things that might be treated otherwise, and what's your feeling? Do you feel that PET will replace or should replace MRI or CT in the workup of a dementia patient? That's the first question.

.00163

And the second is, would you on the basis of the PET showing the temporal or parietal reduction in glucose immunization, tell a patient and their family that that's the diagnosis, and plan accordingly?

DR. SMALL: You're asking me two questions and testing my short-term memory, I think. First, I would say yes, I think PET should replace MRI in the differential diagnosis. As I said in my presentation, generally MRI and CT do not contribute to the diagnosis. Very rarely you pick up a tumor or you pick up some other kind of disease, but routine use of a structural scan like CT or MRI is not helpful, and in fact there's controversy in the literature and among thought leaders as to whether one ought to do that. On the other hand, PET does provide the accurate early information.

And the second, about the diagnosis, I do use PET scan to help me in the early diagnosis and defining treatment, and I share the information with families, and I find it tremendously helpful early on. For example, we see a lot of people who early on have a combination of perhaps mood changes, depression, and memory changes, and we're not sure whether to spend several months giving them an

.00164

antidepressant or to start a cholinergic treatment. If I get a PET scan that gives me the answer, I can initiate the treatment and avoid that problem of delaying treatment months, where I may lose ground.

DR. FERGUSON: Could I make a follow-up? Do you think that with PET scans you could rule out subdurals and tumor along the hydrocephalus easily as well as with the other?

DR. SMALL: Well, MRI is going to be helpful for some situations. There are situations where you want to get a structural scan and it's

going to be more helpful than PET, certainly.
DR. SOX: I have a couple questions about treatment. The first is, what's the average number of months that treatment will delay the passing of a particular milestone? In your study it was about nine months; is that about average for the course? And the second question is, the frequency of the people who drop out of trials of therapy as a proxy for how well they tolerate the side effect and so forth.
DR. SMALL: You're really testing me on this short-term memory test because I have two questions again. Notice I'm jotting them down. Number of months of treatment, the initial trials
.00165
were five to six months; data I just showed were up to a year. We don't have placebo controlled data beyond that, but we do have data up to two years in open labeling, and you can see that treatment is effective over two years compared to a naturalistic decline that you would expect. As long as there is some kind of a cholinergic system left, theoretically treatment may be helpful, but then later in the disease, it's difficult to say when treatment ought to be ended.
Frequency of dropouts is relatively low with some of the cholinergic drugs; people tolerate them very well. Others are more difficult, and now we have two products that are generally used, a third will soon be available, and probably the ones with less frequent side effects will be used more often.
DR. SOX: Thank you. Sean?
DR. TUNIS: I was just going to ask also, because it looks like the way I read the slide on the therapeutic effect, it looked like it was fairly dramatic even within the first couple of weeks, if I read that slide correctly for the cholinesterase inhibitors. And I guess my question is, could you just comment from kind of a clinical management perspective on a therapeutic trial with a
.00166
cholinesterase inhibitor versus a definitive PET scan essentially to make the diagnosis?

DR. SMALL: Generally the way we use the cholinergic drugs, if we think somebody has Alzheimer's disease, we will start them on a drug. If they tolerate the drug or if they get better, we keep them on the drug, because a certain percentage will not show obvious improvement but it will slow down the decline. As far as, I guess the question might be extended to say, well, should we put everybody on cholinergic drugs? I would say no. That would be very costly and probably even though they are relatively safe, you're going to see side effects. So I think that the PET scan is definitely helpful in those early cases where we're not sure.

DR. FRANCIS: I just want to ask you about false positives again. Would you recommend for people who are not symptomatic but have a positive PET scan for Alzheimer's, that they be put on treatment to delay the onset of symptoms, even with a 25 percent false positive rate.

DR. SMALL: Of course it depends on how you define nonsymptomatic. Right now, actually, we're studying questions like that with NIH support, where we have people with mild memory complaints, and

.00167

we are randomizing them to a cholinergic drug or other innovative treatments versus placebo. So I wouldn't generally recommend everyone should take these drugs if they are asymptomatic. But I think you get, there's a gray zone, there's a border zone, and when do you define, when is the cut point where somebody has early Alzheimer's disease or mild cognitive impairment? There's a lot of controversy there, and I think this technology helps us help the patient get started on treatment earlier.

DR. SOX: Alan, last comment.

DR. GARBER: I have a closely related question where the issue is not are the patients symptomatic, but what kind of data is there about the efficacy of the cholinesterase inhibitors in mixed dementias and non-Alzheimer's dementias? Has that been well studied and what kind of results?

DR. SMALL: That has not been well studies but there are emerging data with some of the

products; I have seen data showing that patients with dementia with vascular risk factors have a beneficial effect; patients with lower body dementia have a beneficial effect on behavior. So we don't have as much systematic data but what's emerging is that even if you have perhaps a false positive for Alzheimer's,

.00168

which can actually be a frontal temporal dementia or lower body dementia, if you treat you're probably going to benefit the patient.

DR. GARBER: Thank you.

DR. BROOK: Gary, can I ask you one provocative question? I know we're not allowed to talk about money, but I will. If you could give the money to your Alzheimer's patients to basically get more care giver services and relief from care giving, versus the PET scan.

DR. SMALL: Well, that's easy. I would just say PET scan. And in fact with families agreeing, we've started a memory clinic at UCLA and we give them the options, as most of the carriers do not fund it. We just started the clinic recently, and we find that 90 percent of families who can afford it will opt to get the PET scan to get the early accurate diagnosis.

DR. SOX: Well, we are about to break for lunch, but before we do that, I want to make a brief announcement. We are going to reschedule the public commentary so as to try to get the discussion of colorectal cancer and Alzheimer's disease started as soon as possible. So the way we're going to handle it is that people who have come here to make comments

.00169

on cardiovascular applications of PET, we're going to put their presentations off until we have had a chance to discuss and vote on the applications that we were asked to consider. So any of the scheduled public commenters who planned to comment on oncologic applications, perhaps you could just come up here and identify yourselves so we get the appropriate people lined up to make presentations before the panel starts its discussion.

Lunch time, it's called a working lunch,

which from my point of view means getting your lunch and working as hard as possible to get back here quickly. Connie is going to tell you exactly where to go for lunch. My instructions are to get your lunch as quickly as possible, butting in line if necessary, and not get into a wrangle with a cashier about paying, and come right back here and eat it here. We will start the discussion with Dr. Flamm's presentation as soon as we have a quorum, in order to try to keep things moving along. So please, get back here as quickly as you possibly can, so that we can have as much time as possible to discuss our assignment.

(Luncheon recess at 12:28.)

DR. SOX: Our first presenter for the

.00170

afternoon session will be Dr. Carole Flamm, who is going to be presenting the Blue Cross/Blue Shield technology assessment.

DR. FLAMM: Okay? All right, here we go.

First I would like to thank HCFA for inviting me on behalf of the technology assessment center of the Blue Cross/Blue Shield Association to come and speak today on PET. It's certainly our honor to be able to share our assessments on several indications with you today. The three areas that I have been asked to focus on include lung cancer, colorectal cancer and dementia. That still is a lot to over, and I'm going to try work this into 20 minutes, so hang in there with me.

I would like to first focus a little bit on why patient indication is important, and I think some of the discussion today has brought out some of the issues, but it just does deserve a little bit of emphasis. First, the patient indication determines what diagnostic imaging information we're seeking in doing the PET study. It lays out the clinical context and the frame of residence to determine efficacy, and whether there is added value by performing a PET exam. And what I mean by this is, in some circumstances, PET is going to be used as an

.00171

adjunct to a conventional diagnostic strategy as

Dr. Garber referred to, but in other cases it may be proposed as a replacement for conventional testing. And when it's being used as an adjunct to conventional testing, where the next step that it's being compared to is biological or histological diagnosis, a biopsy, the relevant question may be, is this good enough to replace the biopsy, and in that circumstance, biopsy is the standard by which it needs to be compared, the truth standard is relevant. Looking within patient indication, it permits assessment of how PET will influence patient management, specifically will PET findings result in not performing an invasive treatment or a basic diagnostic procedure, and it permits the assessment of the effect of health outcomes, weighing the benefit of correctly avoiding the invasive procedure, weighed against the harm associated with false test results that we've alluded to earlier. First, let me give you a brief idea of which indications within the three settings we're talking about. Our lung cancer assessment was published in May of 1997, and includes three specific indications, differential diagnosis of the indeterminate solitary pulmonary nodule, preoperative

.00172

staging of mediastinal lymph nodules in non-small cell lung cancer, and monitoring after treatment for lung cancer. Second, we're going to cover colorectal cancer, which was more recently updated, published in April 2000. That assessment covers staging of hepatic and extrahepatic metastases in patients who appear to have clinical evidence of resectable disease, differential diagnosis between local tumor recurrence and scar tissue. And the third indication was included within our May 1997 assessment, the use of PET in differential diagnosis of the cause of dementia in patients who have an unresolved diagnosis after conventional examinations. Okay. We're off. Hang on. Lung cancer diagnosis, we're talking about a situation where without PET, we assume that these

patients would ordinarily be referred for biopsy diagnosis. They have this solitary pulmonary nodule, we don't know what it is, we need to know, the next step is biopsy. So the relative comparator is, how well does this compare to biopsy? Other tests have already been indeterminate; CT has already been done. It doesn't matter how it does compared to CT. We

.00173

already know that our real question is, how does it compare to biopsy? So the reference standard is biopsy diagnosis, and it's intended as an adjunctive test.

We reviewed the literature within a minimum set of selection criteria, further looked at study quality among the selected studies, and the overall body of evidence included 18 studies which included almost a thousand patients. 13 studies focused on indeterminate solitary pulmonary nodules, so that was the main focus of the existing literature for diagnosis. Five studies on solitary pulmonary nodule, including 251 patients, met our quality criteria, all of them prospective, that sort of thing. And we concluded that the data was sufficiently free of bias to look at diagnostic accuracy.

When you pool the studies in different ways you get slight variations in sensitivity and specificity estimates, but just looking at the five highest quality studies, we are looking at about 95 percent sensitivity and 89 percent specificity. Another statistic you can calculate are likelihood ratios, and the likelihood ratio positive is 8.6, likelihood ration negative is 0.56, for those who

.00174

like likelihood ratios.

When we vary the pretest probability of disease over a broad range, you get predicted value positive or predicted value negative, which in this circumstance are the same as post-test probability, and it was decided in looking at this and weighing the benefits and the harm, that particularly in this low range of pretest probability, in the young patient who is a nonsmoker, a predictive value

negative of 99 to 100 percent was probably good enough to avoid doing the biopsy, and it was really that value determination that permitted the conclusion that health outcomes are improved through use of PET.

So looking at the way it changes management, a positive result on PET suggesting a malignant lesion, the patient would still proceed to biopsy and you wouldn't experience a change in management, and the only harm experienced is really just that associated with having done the PET test. The real change in management is in the patient who the PET suggested benign lesion, patient avoids the biopsy. You have to weigh the harm of delayed diagnosis since the false negative rate is high, and so there is the -- focusing on the low pretest

.00175

possibility group, and that's indication specific, that you can define that group.

In conclusion, PET for evaluating indeterminate solitary pulmonary nodule does appear clinically effective, and I'm going to kind of skip through the details, and this is already an indication that HCFA has identified as being a clinically effective use of PET, and does provide reimbursement for that.

DR. BROOK: May I ask a clarification? What proportion of indeterminate nodules would fall, in the Medicare population, in the low probability number that you just dealt with?

DR. FLAMM: Your question is fair. I don't think I can answer it. They're not 30 years old, I can tell you that much, but there may be nonsmokers, but you're right.

DR. BROOK: Out of all the indeterminates, do we know even what proportion of the indeterminates CAT scan and MRI fall into, the .01 and .02? Does anyone have any answers to this kind of simple epidemiology, has anyone ever done prior probability studies on these solitary nodules, or is this all an academic aid?

DR. FLAMM: No, I think that that sort of

.00176

information is relevant.
DR. GARBER: Well, there are a series of studies that look at the pretest probability that it's cancerous, and age is an important positive risk factor that it's cancer, but there are other things, the shape of the lesion, the size, the speed with which it's changed when you have serial chest x-ray.
DR. BROOK: Alan, I'm asking a different question. I'm asking, of a thousand indeterminate pulmonary nodules that come out of it, is there anyone who can actually place them on that prior probability, or do we know anything, do we have any model to place them there.
SPEAKER: (Inaudible.)
DR. BROOK: Actually using radiology today?
SPEAKER: Yes, there are, at least clinically.
DR. FLAMM: There may also be patients who are very risk averse and don't want to have a biopsy, and may choose to base it -- if they have a 25 or so percent prior probability, they may be happy with a 98 percent, given the risks for biopsy in that setting. Another patient with COPD, they may not be low risk, but you know what I mean.

.00177

I'm going to briefly touch on the second indication, which is staging mediastinal lymph nodes in non-small cell lung cancer, and here we're dealing with preoperative patients who are deemed to be operative candidates potentially, and the patients are generally referred for mediastinoscopy or other means of biopsy based on the results of CT, and the goal is to avoid that pre-op biopsy or mediastinoscopy step if possible, in selecting patients for surgery.
Without going through all of the evidence, the studies were of good quality and did show that PET was both more sensitive and more specific than CT, and when PET and CT are used together, a decision analysis conducted by Dr. Gambhir did show that the mediastinoscopy biopsy step can be avoided safely with an improvement in overall health outcomes when

both tests are negative, and that was a nice display of that.
I'm going to spend a little more time on staging in the colorectal section. Briefly, the use of PET for monitoring patients after treatment for lung CA is another indication. CT is the standard monitoring test, and reference or truth standard is histologic diagnosis, but we would be comparing it to
.00178
CT in this circumstance as a potential replacement. When we evaluated this in 1997, only four studies met our entry criteria, and looking at these four studies there are a lot of methodological issues. I think if you're looking to establish the comparative accuracy of CT and PET, you need to have the study compared to another test and only one of these studies did. A blinded assessment is helpful when you're looking for relative diagnostic accuracy; it's not fair to have one test result available for the other. And reference standards were not completely well utilized in these studies. So unfortunately, the multiple methodological limitations of the available studies did not at that time permit conclusions about the ability of PET compared to CT to detect recurrences after treatment for lung cancer. One down, two to go.
Colorectal cancer staging. Looking at the staging of hepatic and extrahepatic metastases in patients who appear to have evidence of resectable disease, so without PET, this is a group of patients who would proceed with surgical resection and the goal of PET imaging is to identify patients who have nonresectable disease and who could be spared the morbidity of a surgical procedure that's not going to
.00179
cure them or not going to provide them an improvement in health outcomes. The reference standard here is again biopsy confirmed staging, and due care that this be correct. It's intended as an adjunctive test in the diagnostic evaluation and what we're really asking is to know how this compares to staging without PET, so it's not a replacement for CT, but looking at its own added value is the major question.

Looking at the body of evidence, there were eight studies that looked at the accuracy of staging hepatic metastases, four studies on extrahepatic metastases, and 11 studies in 680 patients that looked on the effect on management, which is nice.
When you look at the literature on detecting and staging hepatic metastases, there is some variation of study design, analysis and quality, but PET is generally reported to be more accurate than CT. The literature looking at detection of extrahepatic metastases also suggests that PET is more accurate, and certainly at least as accurate as CT, so we are getting some added diagnostic information above what's available with CT.
Looking at the studies that address change in patient management, the best study available then
.00180
was by Flaman, published in 1999, and it included 172 patients specifically with a solitary liver metastasis, and that's a good indication for surgical resection. But PET did alter management in 8 percent of those patients. PET results were discordant with what was available by conventional staging strategy 10 percent of the time and among the disagreements, PET was correct over 85 percent of the time. PET more frequently upstaged disease and ruled out surgery.
The remainder of the 11 studies are fairly supporting of the findings of Flaman, and the discordant PET results are usually correct in the majority of the cases. In other studies, PET altered management between 7 and 68 percent of the time and the unweighted average was around 20 percent. PET ruled out surgery about 12 percent of the time and prompted surgery about 8 percent of the time, so it can do both, interestingly.
So in conclusion, PET does appear to be clinically effective for staging colorectal cancer in patients who have clinical evidence of resectable disease.
Looking at another indication within colorectal cancer, the differential diagnosis setting

.00181
between local recurrence and scar tissue. In a setting without PET, we're going to assume that these patients would ordinarily be referred for biopsy diagnosis, so the use of PET is potentially to avoid having to do the biopsy, an uncomfortable procedure. The reference standard then is biopsy diagnosis; we're not really comparing this to conventional imagining, and PET is an adjunctive test. Looking at the body of evidence there were six studies, including 198 patients. Four were clearly prospective, none were clearly blinded, but we will assume that they were blinded to the biopsy diagnosis findings; it's often not well reported in the studies. Looking at the diagnostic accuracy overall there was about 96 percent sensitivity with a range of reported estimates of 92 to 100 percent, and 98 percent specificity. These are really quite high numbers, but it's important to consider in this population, we have a prevalence of recurrence, tumor recurrence of 69 percent, with a range of 61 to 86 percent, so we are dealing with a majority of patients who really do have a recurrence. So when you look at the post-test probabilities, even given this very accurate test, at 69 percent pretest probability, you still have a 92 percent chance that
.00182
you don't have tumor, but there is still an 8 percent chance that you do, and the question then becomes, do you risk that 8 percent of missing somebody who has a local recurrence and relying on a negative PET test to avoid the biopsy or not. That's the judgment that needs to be made in thinking about the health outcomes effect. So since the probability of tumor recurrence was relatively high, in this range, it seems unlikely that patient and physician would forego biopsy diagnosis and risk delay for 8 percent of the patients.
Finally in colorectal cancer, looking at the indication of detecting a primary lesion, as of April 1997 there were no studies in the literature identified that met our minimum eligibility requirements. I haven't updated it since them but

that was the status then.
Okay, we're almost done. I'm going to briefly touch on dementia. This was an assessment that was written and published in '96 and does not include the recent studies that were alluded to in a previous presentation. I am merely presenting this as our analytic approach to the assessment and what the status was then, and that needs to be kept in mind.

.00183

So we're looking at once you've done the whole physical examination, all the neuropsych testing, everything, and you're still unsure about the diagnosis, how helpful would PET be in that circumstance. Again, it's considered an adjunct, and there are a variety of potential reference standards that might be considered reasonable in the studies. These were used, histological analysis at autopsy or biopsy, long-term follow-up with correlation or response to treatment. None of these is, you know, maybe a perfect reference standard, but those are representative in the literature. One intermediate outcome is the diagnostic accuracy of PET, but ultimately the interest is in how the quality of life is affected by using PET.
As we talked earlier, the effectiveness of treatment is mediated presumably through slowed progression of dementia. As of February 1996, seven studies were included in the assessment, including a total of 319 patients. All of these studies performed PET after the clinical diagnosis of dementia had been made, and that's an important difference, that this is not the group of clinical patients where we are unsure about the diagnosis. These were sort of more early technical efficacy type

.00184

studies primarily.
Six out of seven studies performed PET in a group of patients that clearly had Alzheimer's disease and a group of patients that were clearly the control subjects; that's not the optimal study population for defining sensitivity and specificity of performance characteristics of the test, as it

would be in a set of unknowns.
Only one study, Salmon in 1994, did provide sensitivity and specificity estimates in a mixed population with dementia of varying etiologies, and that study reported 96 percent sensitivity and 61 percent specificity, and our review of diagnostic accuracy of the clinical evaluation ranges in the 65 to 85 percent, and specificity 80 to 90 percent range. The available studies were really not sufficient to estimate the diagnostic performance of PET, largely due to the population that was studied, and they didn't use blinded observers, they knew what the diagnosis, and it was really just how well does PET show this classic appearance, perhaps maybe was more what was trying to be demonstrated in those studies.
In closing, our approach to technology assessment is indication specific. Analysis of
.00185
indirect evidence is frequently required, and clinical effectiveness in one indication may be difficult to generalize to other indications because of the complexities of the clinical context and the differences in diagnostic performance across some settings.
I will mention just one thing that came up in a side conversation, that isn't a topic on the table today, but prostate cancer is one setting where some of the published studies show that it has a very poor discriminating power in diagnosing prostate cancer versus benign prostatic hypertrophy in a study published in 1996 by Effert in 64 patients, established that in a prospective fashion.
So it's really hard to generalize; some tumors despite the biologic reason underlying things, don't display the same good imaging properties as others. So PET may be useful in some settings, not useful in others, and indeterminate depending on what we know from the available studies. It's a complex process analyses, and I'll just stop there. I hope that I have illuminated our thinking a little bit, but I may have raised more questions, and I'm happy to answer any questions you have.

DR. SOX: Thank you very much, Dr. Flamm.

Any questions or comments? Leslie?
DR. FRANCIS: Just one question. The VA studies that we were provided mention a European prospective study of Alzheimer's patients and PET. Do you know anything about that? It's mentioned in '96 and again in '98, but nobody had any results at that point.
DR. FLAMM: Well, our assessment was published back in '96, so I think that predated our review of that literature, and given the timing and constraints for this meeting, I wasn't able to update our work.
DR. SOX: Randel?
MS. RICHNER: You mentioned that six of the seven studies on the Alzheimer's patients had a definitive diagnosis before they were evaluated using PET, and how were those diagnoses determined?
DR. FLAMM: Definitive on clinical grounds.
MS. RICHNER: Clinical examination?
DR. FLAMM: Clinical examination and other tests, perhaps neuropsych tests. I don't remember the details specifically of all the different protocol requirements, but I think that they felt comfortable that these were what they would call

clinically Alzheimer's patients and how does PET look in the patients was more the thrust of those studies.
DR. MURRAY: The one study that was blinded, is that the Salmon study that had the low specificity.
DR. FLAMM: I think that was.
DR. SOX: Okay. We will move on then to hear the VA technology assessment presentation, and the speakers are Elizabeth Adams and Karen Flynn.
MR. COYNE: Hal, can I make a brief announcement while the speakers are setting up?
DR. SOX: Please do.
MR. COYNE: Thank you. Some of you may know, there was an interesting article in the Post this morning concerning HCFA consideration of PET,

the Washington Post. We've made copies for the EC and guests. If anyone else is interested, we do have some copies on the table outside the registration table, which you can pick up. Thank you, Mr. Chair.
DR. SOX: Thank you. I just want to remind the speakers that we are currently running about a half hour behind schedule, and so I want you to adhere strictly to the 20-minute limit, to give a little time for questions after your presentation.
MS. ADAMS: Dr. Hollian, who could not be
.00188
with us today, had asked that we give our presentation to the board, and we're very happy that HCFA took him up on his offer and we're very happy to be here today. The VA has a long history involving PET scanning. For those of you who don't know, the VA is a very large healthcare system with over a hundred hospitals and facilities. We share the ownership and operation of ten PET facilities within the system and among most of the facilities, clinical and research studies are performed.
In the early 1990s the VA had received a request to build additional PET facilities but there was a moratorium in place on adding PET capacity to the system until demonstration of its clinical efficacy had been determined. In 1993 the then acting undersecretary for health requested health services research and development service for an evaluation of PET. He first wanted to know how PET was being used in VA. To that end we conducted site visits and surveys. He also wanted to know if VA should add more PET centers. One rationale for adding PET capacity might be to make clinically useful PET studies available to veterans throughout the system. To that end we undertook a systematic review of the literature.
.00189
The systematic review was designed not just to tally the volume of literature or the number of subjects studied for a particular indication, but to identify and synthesize the highest quality results from research to answer the question of PET's clinical utility to the veteran population. We

convened an advisory board to help focus the process. They selected six clinical indications for PET that were of greatest interest to veterans, and they helped identify criteria for including studies in the review which you will see in a moment. The advisory board included members of the technology assessment and health services research communities, as well as several members from the clinical PET community. They unanimously approved the findings and recommendations in the report, which was submitted to the under secretary at the end of '96. Findings from the VA report have been presented in a number of venues including annual meetings of the International Society for Health Technology Assessment. Karen Flynn participated in the last HCFA technology advisory committee meeting in '97 to discuss PET.

VA belongs to a group called the International Network of Agencies for Health

.00190

Technology Assessment, which undertook a joint project on clinical PET in recognition of a growing interest in clinical PET in a number of health systems around the world. The VA assessment was included in an evidence synthesis in that project, as well as assessments from Blue Cross and two other agencies. The report was submitted in 1999 and is available on the web.

The under secretary agreed with our findings and agreed to implement the report recommendations. A group was convened to initiate a registry for VA PET facilities. That was to bring together all of the VA PET facilities into standardized data collection. The under secretary also commissioned annual updates of the systematic review. So far the registry's in place, but the registry data have not yet led to any changes in VA policy. The technology assessment program completed its first review update in '98 and two other reports from the technology assessment organizations which also applied VA methodology to their systematic reviews, served as updates for '99 and 2000.

For today's meeting, HCFA asked that we

present our findings from our systematic reviews for four specific disease indications. The findings from

.00191

the VA '96 systematic review will be presented today, and you can get more details on that, it is available to the public on the web. The approach we used didn't come out of a hat. We consulted well accepted methods in the literature on evaluating diagnostic tests to construct our review protocol and quality criteria, recognizing that the validity of patient centered research depends on an appropriate match between the research question and the methods used to address it and on the way in which the study was carried out.

We used a formal kind comprehensive search strategy to insure the broadest possible retrieval in each disease specific area. We relied on National Library of Medicine databases and multiple combinations of free text in their subheadings.

Rapid improvement in technical performance of PET scanning in the 1980s supported restricting the search period to 1985 and later for Alzheimer's disease, and for 1991 and later for oncology.

For articles to be included in the review, they must have been published in peer reviewed journals, in the English language, and they must have recorded primary data with at least 12 human subjects with a disease of interest, using FDG as the tracer,

.00192

and these were defined by our advisory board. We excluded articles that didn't contain sufficient details necessary for study appraisal such as you see there.

The last bullet requires a little additional explanation. For many new technologies of limited availability, such as PET, it is not uncommon for research reported in the literature to be confined to a few institutions. To avoid what they call desegregation or redundancy, or double or multiple counting subjects in the study base, we excluded articles that were duplicated in the literature or were superseded by another study from that same institution if it was done for the same

purpose. This allowed to us to gauge a better estimate of the true study base represented in the body of literature and a more accurate estimate of the diagnostic accuracy.
In our search criteria, which is located at the top row, these were screened for review, and the bottom row indicates the number of studies that met criteria for inclusion in the review.
DR. FLYNN: To conduct our reviews we designed a systematic review protocol whose foundations you have heard a great deal today. We
.00193
started with a very broad global overview of diagnostic literature which relied on an article published in 1991 by a pair of researchers who designed an efficacy hierarchy which goes from, the lowest level is technical performance, painting pretty pictures, and the highest level is considered societal impact which presupposes a cost utility analysis based on usually randomized clinical trials which of course are pretty rare. The efficacy hierarchy however, does not give us any real clue to what the quality of the studies at any particular level are, so we went to the evidence based medicine literature for a set of simple criteria to gauge how accurate the estimates of accuracy mighty be from the studies we were reviewing.
And even beyond that, for -- we find it was quite helpful to assign letter grades just like one was in school. And these again, we did not pull out of a vacuum but from the literature. And this just gives you an idea of the overall grading, the volume of grades and the scope of distribution of them. We were not unduly impressed by the literature that was available in 1996. It may be better now, but Liz has basically taken over primary authorship of our PET involvement, so I am a little bit distant
.00194
from it.
And this is your usual grading scheme for evidence, about a causal link between an intervention and health care outcomes. We realized as we worked through the literature that so many of the PET

protocols were so variable and there was so much heterogeneity in the studies that we didn't think that meta-analyses were warranted, which took us off the hook for them. Again, please remember that what we're talking about represents a snapshot of the literature taken in 1996, it may be better now, and people who have commented today have indicated that it might be significantly better.

What in fact we saw was much of what Dr. Brook talked about, an awful lot of work being done but no systematic approach to analyzing what was happening, both within VA and in the world at large. I think Alzheimer's disease has been pretty well beaten to death today. We don't have a great deal different to add except a reminder that this is 1996, and that the only really definitive diagnosis of Alzheimer's disease is by autopsy material and if you don't follow your subjects that far, you're dealing with a presumptive diagnosis, so it's a very imperfect gold standard.

.00195

I've lost track of my slides, but you can read them for yourself and I'm sure you will do that. Obviously from this slide we were pretty unimpressed with what was available in 1996 and as a result, the VA moratorium was continued. And also, a reminder; I believe all of you have copies of our report or the web link for it, and we also gave you a handout that details the methods a bit more than we've given you here.

DR. SOX: Thank you very much. There is time for questions and comment. Maybe I could start of by asking -- is there more?

DR. FLYNN: No, I think we're just about done.

DR. SOX: I would like to ask both of you and Dr. Flamm, since both of your technology assessments for Alzheimer's disease stopped in 1996, are either of your two groups aware of any high quality published studies that would --

DR. FLYNN: I'm sorry, there was one more slide that I forgot. Our protocol is being used by the International Network of Agencies for Health

Technology Assessment for several other assessments around the world, for instance one in Australia this year and a previous one in the UK. I believe you

.00196

also have copies of those, but I can't answer the question right now about what's going on with Alzheimer's disease. There was a much better designed study in progress in Europe in 1996 and I don't know if that's published yet.

MS. RICHNER: I also noticed in your report that there were several active NIH studies going on, as well as in Appendix 3, there were many other studies, and have those been discontinued or are those continuing on? They look like they're registry type studies of sorts.

DR. FLYNN: Do you mean for Alzheimer's?

MS. RICHNER: For Alzheimer's, yeah.

DR. FLYNN: I honestly don't know. As I said, I've stepped back from this project a bit.

DR. BROOK: Am I to understand from this slide that the '99 assessment by the NHS and the 2000 by Australia supports the conclusions that you just said?

DR. FLYNN: Yes.

DR. BROOK: So that again, and that these -- do you know anything about what they've done in terms of, do they believe the efficacy is not there yet for PET, is that what they have said in both the NHS and Australia?

.00197

DR. FLYNN: Yes.

DR. SOX: In '99 and 2000.

DR. FLYNN: Right.

DR. SOX: To your knowledge, were those assessments up to date, that is to say, including studies published within a year of their publication?

DR. FLYNN: Yes, they were.

MS. ADAMS: The Australian report went up, their search went up through January of 2000. And they looked at -- what they did was update our report from '96 and in some cases expanded on other indications, and not just the ones that we reviewed.

DR. SOX: Did either comment on the

gradient of study quality as we got closer to the present, that is, better studies in the last three or four years than back in the early '90s or late '80s?
MS. ADAMS: I commented in the '98 report in the lung cancer staging literature, in '96 we weren't seeing a lot of blinding and frankly, the studies were not written or reported very well, so it wasn't always easy to tell just how much blinding was conducted. We were seeing some better quality in terms of the writing and at least some evidence of blinding interpretation by the '98 report.
DR. SOX: Bob, I think you were first.

.00198

DR. MURRAY: In your December '98 technology assessment you refer to an ongoing European multicenter PET study. Has that been completed or have any results been published from that study?
MS. ADAMS: I can't comment on that, I don't know. We haven't looked at it since, or yet.
DR. SOX: Frank.
DR. PAPATHEOFANIS: Just a comment on the VA's ongoing interest in PET. My chair did a monitoring board for a prospective multicenter cooperative trial on lung cancer for PET that the VA has sponsored. It's in year two of a seven-year run. So I know there are several other major prospective multi-VA trials that are ongoing, so the VA still has an investment in this.
DR. FLYNN: We haven't actually thrown any of the scanners away; we just haven't bought new ones.
DR. PAPATHEOFANIS: Don't throw anything away.
DR. SOX: Are there other questions? I guess I'm hearing from you that your assessment and that of the Australians and the people in the UK is that right now the study quality is not adequate to

.00199

be very certain about test performance for PET.
DR. FLYNN: That's my assessment, yes. It's not necessarily that it's not clinically a good thing to do, but we just really don't know yet.

DR. SOX: I'm curious as to what appears to be somewhat of a discrepancy between your presentation and Dr. Flamm's. How is it -- my reading of the Blue Cross/Blue Shield reports were that they were somewhat more favorable towards study quality, and I wonder if you can provide any explanation for why there's such a discrepancy between your conclusions and what's up here.

DR. FLAMM: I think our technology assessment reports do lay out the quality criteria that we looked at the evidence on, and perhaps we were more flexible in terms of what is enough here and enough there, and these are always value judgments in terms of what's sufficiently free of bias to gauge where performance is. So, without sitting down side by side, we did this, you did that, I thought we both came up with approximately the same amount of studies, both still looking at the same body of evidence, probably.

DR. SOX: So you don't use standardized grading criteria such as --

.00200

DR. FLAMM: We use the same categories but whether we have required all of them to be present, whether we've looked at the best studies and did the other studies sort of go along with that, do we feel that there was such a bias that we couldn't make any sense of the results. There is no gold standard in quality evaluation either, even though these are accepted standards for looking at evidence.

DR. FLYNN: I think too that one of the vantage points that's important to remember is that VA has 10 scanners, and there's an awful lot of activity but nothing very systematic happening, and not much good research coming out. So, our approach was to lever these guys into doing some better work.

DR. SOX: Any other questions? Dr. Valk, just a quick comment please; we need to move on.

DR. VALK: I will make it quick. I have fully reviewed the VA report from '96 and its subsequent follow-up. Essentially the problem was that this review has confused diagnostic and therapeutic evaluations completely. In fact, the

criteria that was used for grading each paper were taken from a paper with the criteria that had been developed for evaluation of treatment efficacy published in a paper by Cook et al., entitled Rules
.00201
of Evidence in Clinical Recommendations on the Use of Antithrombotic Agents. Those were the rules that were used for evaluating the articles on PET. I find that totally astounding and I don't see any reason to expound on it any further.
DR. SOX: Would you care to rebut that or try to?
DR. FLYNN: I don't think there is much discussion in the technology assessment committee that the rules of evidence apply to almost everything under certain circumstances, and there are useful jargon if you like, for talking to each other and for recording what we founds. We did not actually assign the letter grades that I was talking about in any meaningful way that translated into a quantitative score. We were trying to achieve some sort of documentation of the quality of the literature a little more in a qualitative way than anything else.
DR. SOX: Okay. Did you want to comment? I think we need to move on.
MS. ADAMS: The grading scheme that we used where you saw A, B, C, D, were applied to studies of diagnostic accuracy and then the next level down, the ones that tried to estimate changes in diagnostic certainty. The ones that get to
.00202
outcome, where there is therapeutic impact, patient impact further down, those are where we applied the causal link table that you saw, and that is settled criteria.
DR. SOX: But when you comment on the study quality for measuring test performance, it's applying criteria that have been developed for evaluating studies of test performance; is that correct? Did I hear you correctly?
DR. FLYNN: Yes.
DR. SOX: Not randomized trials, which would be inappropriate.

MS. ADAMS: No. The Kent and Larson articles I believe came from evaluations of MRI, so we didn't even look at diagnostic imaging evaluations.
DR. SOX: Okay. Quick comment and then we really must move on.
DR. SMALL: There is a study of 284 patients that has not yet been published, the manuscript is in print, and I could make that available to the committee members if they wish, as quick as tomorrow.
DR. SOX: Thank you. Let's move on.
DR. JOHNSON: One final question.

.00203

DR. SOX: Yes, Joe.
DR. JOHNSON: A few comments back, I didn't hear the end of your comment, I thought I did, but I want to be clear. You stated that it's not, your conclusion was that it was not necessarily clinically the right thing to do, but that the standards didn't measure up academically on the paper? That's the part on the final comment that I didn't hear.
DR. FLYNN: Well, what I was trying --
DR. JOHNSON: That clinically it may be the appropriate thing to do, but your review of the literature --
DR. FLYNN: The currently available research does not give a clear answer on that point, in other words, we really don't know yet, and the research hasn't been good enough to support a firm answer.
DR. SOX: Thank you very much for your presentation. The next presenter will be Dr. Joseph Lau, who runs the evidence based practice center at the New England Medical Center, and he evaluated materials submitted by the proposers for coverage of PET scan.
DR. LAU: Good afternoon. I'm Joseph Lau.

.00204

I'm the director of the New England Medical Center evidence based practice center, one of the 12 designated by the Agency for Health Care Research and

Quality to conduct evidence reports under contract for the Government. We were asked to evaluate the PET report submitted by the PET community. They submitted a report which I believe you all have, t HCFA in their request for a broad based reimbursement for PET.
Their report stated to have used the data from 476 articles or abstracts that represented over 19,000 total patients studied with FDG PET. HCFA then requested an evaluation of the submitted data by NEMC and due to the time constraint, we only had six weeks to do so, it was decided that the evaluation would be limited to selected areas as shown in this slide.
First, we were to replicate the literature search from 1995 to the present, and we were to list all articles submitted, excluding abstracts, review articles and case reports, and we were to list all articles not cited in the submitted material.
The second task, we were asked to conduct a literature search for 1990 to 1995 and list all potentially relevant articles found.

.00205

The third task was to determine what proportion of the literature was submitted and whether articles submitted were representative of the body of the literature.
And the last task was to identify the key strengths and weaknesses of the data table submitted in the PET report, including quantitative errors or misrepresentation in the submitted material.
It is important to note that our tasks were not -- we were not asked to perform a de novo evaluation of the original PET studies. We were asked to evaluate this report. We applied commonly accepted standards of conducting systematic reviews to evaluate the PET report. Strictly speaking, we were unable to replicate the specific results of the PET report because replication requires knowledge of the exact definitions and processes used in the original work. Commonly accepted standards of systematic review required a well focused and clearly defined questions and associated terms.

There was some discussion this morning by Bob Brook about the narrow focus question versus broad issues.
And study questions were not explicitly formulated in the PET report, nor were requirements
.00206
of the reference standards and the test specified for each of the conditions. In our attempt to replicate the PET report ourselves, we had to infer from limited descriptions and extrapolation of the data presented in the tables. However, our assumptions about what was done in the PET report were often violated by discrepancies and irregularities we found within the data tables. The PET report appeared to have used a very broad definition that was inconsistently applied to all the conditions, thus making it difficult to determine the inclusion exclusion criteria.
And for example, the exclusion criteria stated in the PET report included less than or equal to five patients studied, lack of clear methodology, results reported incomplete or inconsistent, or not easily convertible into data for a spreadsheet.
However, the criteria of the lack of clear methodology was itself unclear. Many studies in the table did not report test performance data, or reported only incomplete data. It appeared that the PET report did not apply this exclusion criteria consistently.
So we, in this slide, methods would be applied. We therefore had to conduct our own
.00207
literature search based on our best belief on how this should be conducted on the Medline and Biosis Previews, the same two databases used in the report, for each of the clinical conditions listed in the PET report. And we then screened the search results and identified potentially relevant studies that evaluated test performance. We then compared our search results with those listed in the report to identify potentially relevant studies missing in the PET report. And finally, we critiqued the data tables in the PET report in order to highlight key

strengths and weaknesses, such as the presence or absence of sensitivity and specificity data as well as commenting on the methodology of combining the data.

Here are some of our basic criteria for defining what are suitable to assess test performance. This is what we considered as minimum criteria. According to what HCFA has asked us to do, we allowed only published full articles based on original research; abstracts, review articles, and case reports were excluded, and for diagnostic purposes we looked at the enrolled patients with and without diseases. Ideally all the studies should include patients prospectively selected in the

.00208

original articles, but it was often not the case and some patients were selected prospectively or retrospectively, and in sensitivity and specificity results, studies that report only sensitivity were excluded, and my colleague Dr. Balk will present the results.

DR. BALK: I'm just going to give an overview of the results we had. Of the 476 articles that were used, that were reported in the PET report, there were a number as shown here that didn't meet our criteria. 94 of them were abstracts from a single issue of one journal; there were 59 other abstracts. 20 of the articles were review articles or from consensus conferences. There were 19 methodological articles such as meta-analyses and decision analysis that did not present original data, and a variety of other articles that didn't meet the criteria either set by HCFA or set by the authors of the report.

This table summarizes the numbers both presented in the report and our revisions of those numbers. We have all the conditions that were covered in the report and in the second column here, the total number of unique full articles that we found, those are the first column of numbers, and

.00209

then in the parentheses are the number of articles reported in the PET report. And in the last column

are the number of patients, both the number of unique patients that we found and the number of patients reported. So as you can see in general, the number of both articles and patients that we found was considerably smaller than the number reported. This number is similar to what was mentioned earlier. The report had 23,000 patients but from our analysis there were only 5,000 unique patients that were in research articles, original research articles, so about a quarter of the number. And again, 104 articles total, as opposed to 476. And let me just point out that for a number of conditions, we found no evidence whatsoever, prostate cancer, venous cancer, zero articles. Thyroid cancer, unknown primary features, there were no original articles that met the criteria.

As Dr. Lau mentioned, we did our screen in Biosis Previews. We split it into two categories. The 1993 to 2000 essentially overlaps the period of time that the report, the PET report covered, and this is the earlier period from 1990 to '92. So we found, just in the last column, '93 to 2000, 3,500

.00210

potential abstracts of interest from Biosis. After screening them, only 77 of these articles met criteria and again, many of the conditions had no articles.

I have similar data from Medline. This time to conform with what was in the PET report, we used 1995 to 2000 to be the period of time represented in the PET report from 1990 to '94 as the earlier period. In this search we found about 2,500 abstracts between '95 and 2000; only about 340 of those met criteria. This number plus the Biosis number of 77 here, would be the number of articles that would need to be reviewed in full prior to being included in any full analysis.

These are some of the critiques we had of the statistical methods of summarizing the data. Our first point is that there was multiple counting of subjects. In their table they have a column of total use patients, and this heading listed, and was listed

in each of the tables, and was defined as the total number of patients actually studied. The PET report frequently listed the same studies multiple times, sometimes multiple times under the same purpose and sometimes multiple times under different purposes. Thus, the same studies, original articles in the PET

.00211

report may have contributed subjects to the total numerous times, resulting in exaggeration of the number of subjects evaluated. I will be showing examples of this.
We had some issues with the method used to combine test performance. The test performance data in the PET report were combined using a weighted average of the sensitivity and specificity independently of each other. And there are some issues with that that I'm not going to go into at this point. The same weight was applied equally to combine the sensitivity and specificity values across studies even though there were a different number, in each study there were different numbers of patients contributing data to sensitivity and to specificity values.
Some of the studies listed in the PET report tables provided only test sensitivity; there were no patient studies that did not have the disease. Without the corresponding specificity value, the sensitivity result is not very meaningful we believe, as any test can be made to have virtually 100 percent sensitivity or conversely, 100 percent specificity. In the PET report, singularly listed test sensitivity of 100 percent was found 46 times in

.00212

the cancer tables, and these were combined with other studies that reported both sensitivity and specificity results.
The PET report had a column discussing management effect, and I'm not going to go into detail here, but it essentially was, we thought it was, this management effect was applied inconsistently and it was questionable what value it had.
I'm going to show a couple sections of

some of the tables that were reported, if I can get this all on the screen. This is the first part of the table for lung cancer studies. As you can see, there are a number of studies listed; however, only one of them is a research article, the rest are abstracts. For this research article by Lowe in 1998, there are multiple purposes of the study, and these are listed here. For example, the PET scans were analyzed both by visual analysis and by using the SEP data. What, the way that the total number of patients was calculated was that this column of numbers here was simply added up. Thus, there were 89 patients in this Lowe article; however, it contributed 89 twice and 34 patients twice, so this is an example of the multiple counting.

.00213

They actually had two separate columns for numbers of patients, the total number of patients and the total use patients and as you can see here, they were somewhat inconsistent in which column the patients fell into, the total number or the total number used. And again, these totals are just simply additions of the columns. This was likewise carried through to the combination of sensitivity and specificity of the PET scans -- I'm sorry this pen isn't showing up very well -- where the multiple duplications were just simply averaged together. And the last column here is the gold standard used. I just want to point out that there was some inconsistency here. For example, in the Lowe article, where there is no gold standard reported, in reality histology was done. This is a part continuation of the table on lung cancer. Again, an example of one study contributing the same patients numerous times. And there are also a couple of studies here, the Saunders study and Marom, where there were, the sensitivity was reported but there were not data available on specificity, and on both of these studies, as was typical, the sensitivity was 100 percent. Of note, this table includes the

.00214

meta-analysis that was mentioned earlier, so this

clearly doesn't meet the entry criteria as it's not original data. To compound the problem more, there are some of the studies in this table and the continuation of this table that are actually, those patients are already included in the meta-analysis. In addition, this 2,200 patients, use patients here from the meta-analysis, actually contributed no data to PET scans; those 2,200 patients had PET scans done.

So this is a section of the lymphoma table, a couple of points here. Here this Bangerter article is repeated twice and with exactly the same information across the row and with the patients and test diagnostic accuracy being counted twice. There is an article here at the bottom, Stump, which points out a problem of actually many of the original studies themselves. This study looked at 50 patients, which actually works out well here, 35 plus 15. However, the sensitivity and specificity were actually derived from 71 scans, where the multiple scans in patients were counted as being independent of each other.

And one last example here from gastroesophageal cancer, Flanagan up here under their

.00215

diagnosis topic, subtopic, and Flanagan here under staging, are both looking at primary tumors presenting exactly the same data with a duplication under different subcategories. And another example here, this study, Luketich, which had a hundred patients derived, the sensitivity and specificity from 276 sites of distant metastases.

We have a few more articles here that had missing specificity with either 100 percent or very high sensitivity. As an example of the issue I raised earlier with this management effect, you can see that the management effect was only listed occasionally, and there was no explanation in the document as to why that was, that information was given only for some articles.

So, we had a number of problems with the PET report. I'm not going to read through them all, I just wanted to highlight a few of them. As Dr. Lau

mentioned, the search strategy was poorly defined, the report included very large number of abstracts and very small studies, the report included studies that had only incomplete test performance results, specifically only sensitivity.
Some issues with the reporting of the data. As I mentioned, there was multiple counting of
.00216
the same study patients, the report used data from the total number of scans when multiple nonindependent scans were performed, and also from total number of lesions or sites from fewer patients.
The report also repeated data from the same articles an analyses in multiple categories.
Some further issues with the reporting, they included multiple outcomes in the same subcategories, for example diagnosis and staging and recurrence were all reported as being diagnosis articles, or staging or recurrence. And also, I didn't give an example of this, but a number of tumors were misclassified; a specific example was that intracerebral metastases were classified as primary brain tumors.
Some problems with the synthesis of data.
We believe that an incorrect meta-analysis methodology was used to combine the sensitivity and specificity data. There were many numerical errors in reporting the data. There were mixed different methods of reading PET scans combined together, different test positivity criteria combined, different mixed sites combined. And they also, we believe, inappropriately combined all the cancers together into a single test performance value, which
.00217
was mentioned earlier.
Some other problems with the synthesis, there was lack of evaluation of methodological quality of the individual studies, and the definition of management effect was vagus and we believe mostly meaningless.
And finally, overall there was an overinflated number of studies stated to have been used in the PET report due to the inclusion of many

abstracts and other inappropriate nonoriginal articles. There was an overinflated number of studies due to the use of inappropriate citations. There was an overinflated number of patients used in the report due to multiple counting, and the use of abstracts. And there was an inappropriate extrapolation and an interpretation of the results such as the sensitivity values, and a large number of potentially relevant studies that we had screened appeared not to have been reviewed, although we were unable to do enough analysis to make that statement definitively.

DR. SOX: I'm eager to move us to have a real discussion of the specific topic, but if there are specific questions that anybody would like to raise for Dr. Balk, let's do it now. Thank you very

.00218

much for that. We're now going to hear from Sam Gambhir.

DR. FERGUSON: I have one, Hal.

DR. SOX: H, I'm sorry. John?

DR. FERGUSON: Monte Erlichman gave me a couple of things referenced in the NHS study and the Australian study and they are very brief, and I think that they provide a little bit of input regarding those two things.

DR. SOX: John, would it be best to talk about those now, or when we get into specifics?

DR. FERGUSON: I just wanted to let you know that we have them here.

DR. SOX: Okay. Why don't you bring that up when we get into the discussion of specific topics. Thank you. Dr. Gambhir, before you start, we're eager to get discussion, so please don't go over 15 minutes or I will be forced to ask you to stop.

(Inaudible comments from floor.)

DR. SOX: I'm sorry. Dr. Gambhir was scheduled from 1:30 to 1:45.

SPEAKER: I know, and the previous presentation was a 15-minute schedule and it went half an hour.

.00219

(Inaudible discussion.)
DR. GAMBHIR: Anyway, I'll do my best to be as time efficient as possible.
DR. SOX: Do your best. Just realize, we're trying to get on with what you want us to be here for.
DR. GAMBHIR: So, I'm Sam Gambhir, I'm from UCLA, and I am here to try to defend the broad coverage statement that was criticized and critiqued by the last set of reviewers. I'm also here to try to put together a few of the things that we've heard throughout the day today.
Just to give you a background so as to tell you a little bit about how seriously I take this work, I run a decision analysis laboratory at UCLA, I teach decision analysis statistics and modeling, my doctorate's in mathematics, and I have an M.D. and training in nuclear medicine. I've published numerous formal meta-analyses and numerous cost effectiveness articles, both in PET and in non-PET imaging, and I read nuclear medicine scans including PET scans one out of every four weeks. In addition, I actually work to help develop new tracers to image oncological processes. The reason I mention all this is that the kind of things I'm going to show you are
.00220
not attempts for me to just look at these things casually, I take them very seriously. I also try to remain as unbiased as possible. Everything I'm going to show you is done without funding from industry, it's done through the help of undergraduate students and graduate students in the laboratory, and it is not in any way influenced by funds coming from a potential party that may have an interest in promoting PET.
To give you an example of how I would like to do the kinds of work that we tried to do in that broad coverage document, I want to show you examples of just five articles in the last two or three years from my group. These are formal cost effectiveness articles and meta-analyses that take a look indication by indication for the use of FDG PET, go through and analyze the literature in detail, then go

through and critique each article through a series of 18 subpoints for their validity, quality, all biases. Then we formally pool the data using ROC analysis, et cetera, and then we go on and do a formal cost effectiveness decision tree model. So it's done in a very systematic rigorous way, it's not an ad hoc way of reviewing or analyzing, understanding the literature.

.00221

And these results are published in collaboration with surgeons, with oncologists and with imaging physicians, so they really require a large base of expertise that we provide, and they are done in journals that are considered fairly broadly read as opposed to specialty journals in just imaging. So just to tell you, before I get to the broad coverage document and its goals, what we do and have done, we usually go through and compare in a very systematic way the incremental cost effectiveness ratio for an FDG PET based strategy versus a conventional algorithm. We look at all costs, both the cost of the studies, the cost of downstream tests, the cost of complications, as well as issues of life expectancy and when possible, quality of life. So all our decision models are formally rooted in decision tree and public health care policy.

We in fact go through and compare with regards to hypothetical strategies like was mentioned earlier today. For example, in recurrent colorectal cancer we look at and have looked at CT alone versus CT plus PET, versus just observing the patient, and trying to really understand, what are all the subtleties to management that dictate the outcomes

.00222

that we need to carefully define and understand. We formally model all pretest likelihoods, the propagation of probabilities down these decision trees to arrive at the exact outcomes for a given pathway due to the diagnostic test being introduced. As one example of this, I just want to show you the latest work we've done in recurrent colorectal cancer. The big issue here is why bother

doing anything at all for patients with recurrent colorectal cancer, unless there is some difference down the road. So the big issue is in terms of life expectancy; there is a five-year survival difference if you operate on patients with hepatic only mets versus if you don't operate. So this is the basic life expectancy data that moves us in the direction of trying to identify those patients that are in fact really operable candidates.

So we went through and again, using undergraduate and medical student help, this is Dr. Hubern, who is now a medical intern in Germany, went through and analyzed the literature systematically to do a formal literature review, looking at each and every article published in the area of recurrent colorectal canter, we try to define all the weaknesses and strengths of each article, we

.00223

try to pool in different ways so that we can understand what in fact are the limitations of our pooling process. We fully define confidence intervals.

And in this case what you're looking at is for whole body, you're looking at a sensitivity of 97 percent across a total number of patients of around 281, and a combined specificity of 76 percent, with the confidence intervals as shown here. Ideally of course, for every indication we would like to be able to do this kind of formal meta-analysis of the data and then publish it in a timely fashion, but because of limited resources, we just can't do it fast enough.

We also look at management data because in fact, it's not just these sensitivities and specificities that count, it's in fact how that leads to change in patient management which hopefully then correlates in some way to the formal cost effectiveness ratios that I mentioned. In the case of recurrent colorectal cancer, again in this article just published two months ago, we've shown a pooled management change with a confidence interval of 25 to 34 percent, and a mean management change of about 30 percent for patients with recurrent colorectal cancer

.00224
who had an FDG PET scan in addition to their conventional workup.
We take these values, and then formally go back and also meta-analyze the rest of the literature, that is CT's accuracy in this case, the accuracy of biopsy, the morbidity mortality rates of the various procedures. This is all published in that same article, and what we find are the values listed here. And the key thing for you to note here is that CT sensitivity outside the liver is about 76 percent, specificity of 56, in contrast to FDG PET of 96 and 76. And now this is not just picking one or two articles out of the blue, this is actually the formal meta-analysis linked in to both analysis for FDG PET and then going back and looking as well for the other issues.
The cost of the various procedures of course are well understood. We use Medicare reimbursement costs for most of our models, we don't model indirect costs currently. FDG PET in these cases being $2,000, surgery, which we're trying to avoid for patients that have extrahepatic mets being $22,000, and CT in this case being around $800.
This particular set of results is just accepted and going to be published in Annals of
.00225
Surgery, along with two other articles that are looking at the role of FDG PET in colorectal cancer.
It took us about one and a half years to formally build this decision model, to account for all the variations through the sensitivity analyses.
What we show in this model is actually if you add a PET to a CT, that is the conventional strategy, you actually increase cost slightly at the gain of life expectancy. Now this is not the life expectancy gain for one little individual of .03 years; this says for the whole population there is this gain in life expectancy. And it's this gain in life expectancy that comes in this case at an additional price. We calculate the formal incremental cost effectiveness ratios; as you know, for the health economists, the number we like to look

for is $50,000 per year of life saved, and PET clearly falls below that. If you in fact penalize all the PET parameters, the sensitivity, specificity, cost of PET, you still end up with a cost effective PET usage and in fact, the actual number of patients we predict in the US that will have management change in this case is about 170 patients will avoid unnecessary surgery by adding a PET study. We have published similar models in

.00226

solitary pulmonary nodule management, non-small cell lung cancer staging; in those cases you save costs and you gain life expectancy. In this case, you actually increase cost somewhat at the gain of life expectancy. So we weigh all these things in these decision models, and we try our best to constantly update these models.

So, I'm showing you all this because I want you to know, there is that kind of data available, at least in lung cancer and colorectal, but not across the wide spectrum of possibilities in FDG PET imaging. So what was given to us was this task of how do we figure out now within a reasonable period of time, whether broad coverage of FDG PET is even a possibility in terms of what the literature shows.

So our goals in that broad based document were not to do the kind of things that I just showed you that we usually do, which is the meta-analysis and decision analysis modeling; our goal is to perform a literature search, to have a broad overview of the use of PET across all applications. It was meant to be a library, or a collection of all the articles with the details that each article provides of the actual use of FDG PET and the limitations. We

.00227

were not, like I say, going to do a formal meta-analysis; that was not the task, or a cost effectiveness analysis. As a matter of fact, the more we put in the more requests we got, well, can't you tabulate the data in some way, can't you give us some overall summary measures, and that's what led to the kinds of data analyses that I'll get into in a

minute.
So please keep in mind this goal. And this goal was not dictated or one that I made up, or the PET community made up; this was a goal that was agreed on in the earlier meetings between Drs. Kang, Phelps, Coleman and others when they tried to understand, well, how could we try to get a broad brush stroke analysis of the PET literature.
We in fact decided the following: Unlike the reviewers of our proposal, we decided we would physically retrieve every single article. You cannot do these analyses or critique them without physically getting each and every article. Even then, it is a logistical nightmare to actually read the article and understand the limitations. But we said in the six-month time frame we had, we will physically retrieve them -- and by the way, there are still articles that we haven't been able to physically

.00228

retrieve because of limitations in libraries being able to interloan some of these articles out. Some of them are in unusual journals.
We did search different time period periods based on the application, mainly because in PET, the first applications came out in neurology, subsequently in cardiology, and then in oncology, so our oncology literature does not go as far back as cardiology and neurology.
We did find that in fact key words, subject and title searches lead to a different set of articles. This is inevitable for every literature search mechanism you can employ. As a matter of fact, there's now an article that's come out four weeks ago showing that based on the key word you specify, that the range of FDG PET literature you'll pull up can vary as much as 40 percent, because the key words people are using in their articles and in their title and subject headings is extremely varied.
So it's physically impossible to guarantee you're pulling up every article, and I'll try to explain limitations and how we're trying to work around that even right now.
Our inclusion criteria were based on not

trying to do a meta-analysis, but trying to cover all
.00229
the literature that we could out there. We were told to include abstracts, we were actually encouraged to do that because if you don't, for the newly emerging applications, thyroid cancer, prostate cancer, musculoskeletal cancers, you won't get any published research articles. So we said it's important to include abstracts, we were encouraged in fact to do just that by HCFA. We weren't trying to hide which are abstracts and research articles. As a matter of fact, in the report we clearly outline what is an abstract and what is a research article, because we want later to be able to do subset analyses to see what differences there are between the two when in fact that becomes an important question.
The other thing to keep in mind is, in our field, abstracts are not just published in journals that are not reviewed. Those abstracts from the June 2000 meeting, are from peer reviewed submissions of abstracts and they are from presented data. We purposely kept the June 2000 as one of the guidelines, because we knew a lot of new data would originate at that time, and that's why it was included. We also did have to use clinical judgment. A lot of the errors you were hearing from the last reviewer won't make any sense once you
.00230
actually sit down and look at a lot of these articles. You have to have read the article to understand what in fact are some of the subtle issues in the spreadsheets that I'm going to talk about. That's why in a new refined version of the report, we've now actually put footnotes for each row of the spreadsheet, so that people can't accuse us of trying to hide any kind of data. In fact what we're trying to do is be as open as possible about our criteria, be as open as possible about the limitations, and in fact constantly strive to improve the criteria and the inclusion of new articles.
We did try to exclude review articles. I apologize if there's two or three meta-analyses that mistakenly ended up in the lung cancer section;

that's going to happen in any large issue like this. I don't know how much that would affect the final result, we didn't hear that, but in fact we did have some mistakes of that kind, which we are trying to correct.
For us, we did exclude studies less than five patients because in our literature, as you've heard mention, 20 or so patients is a lot of patients in a study. It doesn't sound like much but in an imaging world that's a lot of patients. So, less

.00231

than five we considered to be excludable, greater than five to us is worth putting in our summary of data available.
For the non-English criteria, the abstract if it's in English, is useful. So we include the article even if the article is in German, because the abstract is in English. And to tell you the truth, I would include the whole article if I could have had time to translate the German, because I think a lot of good data is originating let's say from Germany.
So the key here is, there was joint agreement to include both abstract and research articles. A large bulk of the discrepancies you're hearing pointed out by the last reviewers is because they're saying we shouldn't have included abstracts. As a matter of fact, HCFA agreed that we should. Clearly we marked the abstract versus research articles, and our goal actually was to be less biased by being more encompassing of data available from the literature.
A lot of these abstracts are from community based physicians trying to do the studies. They're in fact reporting lower sensitivities and specificities than we would see at academic centers. We would love to have run ROC analysis; not possible

.00232

with the type of data presented. We'd love to have done two-by-two pooling of the data; not possible.
So why did we choose weighted averages? Because not all of the abstracts and articles are reporting, as you read them, the formal true negative, true positive, false negatives, false

positives, for us to do a formal pooling. So we could in that case, not include any summary data, we could just list the articles for you and say go ahead and just look, or we could do what we tried to do, we'll just say let's get a flavor for what these articles are saying by at least looking at a weighted average, and that's what we attempted to do. There's this whole issue being brought up about overlapping patients and somehow we're trying to increase the numbers of patients. That's just a misunderstanding. What we're trying to do is trying to show that for each article, you can look at diagnosis, diagnosis and staging, recurrence or monitoring therapy, there is differences within each article in the goal and it was assessed in the same way. This was true in PET across not only the major types of cancers, but across these clinical categories. Well we report are not only overlap, or not on the total number of patients, but the number

.00233

of overlapped patients, that is, how many times we double counted. To the best of our abilities, we want to make sure, and as you will see in the spreadsheets, we are making clear what the overlap is.

So, the number of multiple counted subjects or overlapped patients in the expanded document, we have continued to look for more papers, even since the submission of the original thing, is about 3,844; the number of actual total patients is 24,395; together these come to 28,239. It's not that this number is made up of half or 80 percent overlapping patients. As a matter of fact, as you can see here, it's about one-seventh, one-eighth of the total number we're reporting, and we clearly report what those are.

After conducting our literature search we have been able to retrieve now 813 articles and abstracts, and I mean physically get these and read these, not just look at their numbers and try to put them in a spreadsheet. Of these 813, we used 549 article abstract listings within the spreadsheets, of which 66 are repeated across categories for which

they are relevant. So we've got about 483, and our unused reference library contains the remaining
.00234
articles.
We also thought it would be fair to show you the articles that we read but we're not including because in fact either they're talking about ways of improving detection, or ways of looking at PET scans, it doesn't help us fill in those spreadsheets.
So, I want to make this next slide very clear. Our original search, we estimated a sensitivity specificity of 84 percent and 87 percent, and a management change of 33 percent in about 18,198 patients. We've continued. I knew even from the time that report went out that we would continue to find more articles that we could physically retrieve. And in fact, in our expanded search, with an additional 167 articles -- most of these by the way are now articles, 70 percent versus 30 percent abstracts -- the overall sensitivity and specificity is still 84 and 88, and the overall management change is now 32, and this is now in 24,395 patients. I don't think we're diverging away or undersampling the real literature out there. I think in fact, we're converging toward numbers that are in the mid-80s, including abstracts that are in fact done from community practices.
We've also, by the way, looked at the real
.00235
other question. No one's addressed yet, well, if you start putting the other technologies under this same scrutiny, how well do these other technologies do. In fact, luckily, in our articles, a lot of the studies have gone on and looked at patient analyses in both CT and PET in the same patients. As a matter of fact, 8,000 of the 24,000 or so patients did this. That's one-third of all the studies we reported. In those one-third, you're seeing that PET had a sensitivity specificity as at least assessed through this rather crude weighted analysis, of 85 and 89, but the same weighted analysis leads to a sensitivity and specificity of CT 66 and 76. These are not trivial differences, these are significantly real and

I'm sure even with additional articles and additional pooling of data, will continue to bear this kind of stuff out.

So, our conclusions. No literature search strategy is all encompassing. We even, like I said, now are seeing articles that are addressing how to find more FDG PET literature. Approaches we used tried to utilize as much as the data as possible available from the literature, not to try to exclude data like we would in a formal meta-analysis and cost effectiveness analysis. Finally, our expanded search

.00236

shows near identical results to the original search and in fact, that convinces me more so that we're converging towards a real answer. And FDG PET does significantly outperform CT.

So I will end with one last thought, and that is that I was coming here under the impression that we would focus on broad coverage, that colorectal, which we have decision models for, lung, which we have decision models for, to me, those are givens, they're clear. As a matter of fact, if you go to those articles, you will see that decision models bear out support of PET in those applications. I thought the focus would be how do we jump from those givens to broad coverage. And I would throw out that what you have to keep in mind, especially in cancer, is that when we look back 30, 40, 50 years from now, cancer will not be viewed as an organ specific entity. We won't be looking at breast cancer, lung cancer, colorectal cancer. We will be looking at molecular pathways that unify cancers across different occurrences in the body. Memorial Sloan Kettering under the direction of Dr. Varmas, has already started to restructure the entire institution not to be organ based in its approaches, but to be molecular based.

.00237

PET is a molecular technology, and you've got to get past the thinking that you need to prove for each application a given set of numbers. You've got to go back and say all the cancers share molecular abnormalities, and we in fact are tracking that

molecular abnormality with FDG. So I'll end with that.

DR. SOX: Thank you very much, Dr. Gambhir. Does anybody wish to comment or ask questions? Leslie?

DR. FRANCIS: I would just like to ask you, most of what you just said was directed to cancer, and you in the report here, management change data for patients not directly available from the literature and the decision model not applicable to this management problem for patients with dementia. And I'd just like to ask you to comment on whether you think there is really -- I mean, all of the studies you had were cancer studies and so on -- whether you think there's anything at all out there about management in patients with dementia.

DR. GAMBHIR: For dementia, I'm glad you asked that, because there is of course the literature, although some of it is still not published and just about to be published on the

.00238

actual accuracy rates. There's not a formal meta-analysis or decision model. As of four months ago, we started the construction of a formal cost effectiveness model actually in collaboration with Dr. Gary Small, who presented earlier, and others, to actually model the entire management process in dementia, including diagnostic imaging. There's one or two articles that have appeared previously in the management of dementia and the cost effectiveness, but they have failed to incorporate diagnostic modalities into their algorithms, so we are now trying to increase the utility of those algorithms by updating the management component through these diagnostic tests. But no, there isn't a preexisting decision model for dementia.

And again, keep in mind, these decision models take one and a half, two years to build. These are not gather the literature and plug in a little decision tree. To understand all the subtleties of clinical management requires a combination of expertise and especially without any

real funding, unlike drug companies who have an interest to see the drugs rapidly improved, and there's a lot of money, for these it's our own
.00239
attempts to merge this data, and that's why I don't have decision models for all these categories already ready for you. And I would even add that to get those ready would take 20, 30 years.
DR. SOX: Well, it's time to move on.
Thank you very much.
I just want to remind those of you who came late or those of you who missed or forgotten the earlier remarks about why we're here, from HCFA's point of view and I think from the panel's point of view, the most important thing we can accomplish today is to give a good workout to some guidelines for evaluating diagnostic tests among which is PET, and secondly, to advise HCFA on the quality of evidence for several selected examples. But the main thing to do is try out these guidelines and see if they work. To do that, we're going to have to have a discussion among the panel and we are about ready to launch into that.
Because some people have to leave early, I'm going to restructure the agenda in order to allow as much discussion among the panel to occur before we start to lose folks. So the plan first of all is to have Sean sort of frame this discussion around what HCFA's needs are. Then we're going to discuss a
.00240
couple applications of colorectal cancer, and use of PET. Then we're going to give a chance for some public comment. And then we're going to form a consensus about colorectal applications. Then we'll move on to talk about Alzheimer's disease and we will see what time it is by then. So Sean, do you want to sort of get us pointed in the right direction here?
DR. TUNIS: Yeah. First, let me just check. I don't want this change in the schedule to prevent anyone who's scheduled for a public comment to not be able to do that comment. So if there's people who scheduled for public comment who have to leave within the next hour, we would take their

comment before this panel discussion. But, we do feel it's important to have an opportunity for the panel to start to digest what they have had heard here.

In terms of framework, just as a little backdrop, as many of you know, the coverage function within the Health Care Financing Administration has been trying to move towards a more clinical effectiveness and evidence based approach to coverage policy, doing it in the open, and using empirical evidence to try to be consistent about what is and isn't paid for. As part of that, obviously, we are

.00241

faced with the question of how do we apply this to diagnostic technologies, and particularly in this case, we have the request for the broad coverage request that Sam just talked about. And what we felt, particularly on short notice, what we would be able to do at this meeting is try to apply an evidence based framework around diagnostic testing, to some applications of PET, and to see how far that gets us in terms of being able to think through how to make coverage decisions related to diagnostic technologies.

So that's, in that spirit, Alan and Hal had drafted this framework and we decided to focus on a couple of essentially case studies to try to apply that framework and for today's purposes, the case studies were lung cancer, colorectal cancer and Alzheimer's disease. We understand that there is already Medicare coverage and you know, based on good evidence, for lung cancer and some applications in colorectal cancer, though not all. However, whether or not these uses are covered, the framework can still be given some exercise, so that's what we're going to proceed to do now is open that discussion, try to apply this framework and as part of that discussion, as part of trying to apply this

.00242

framework, the whole issue of extrapolating from empirical evidence in one condition to making judgments about clinical utility in other conditions will necessarily be part of that conversation. So

that's where we're trying to go now, and I'll hand it back over to Hal.
DR. SOX: The approach that I would like to take trying to keep us using our framework, is to go through a summary of the data relying mostly on the Blue Cross/Blue Shield assessment, but trying to put it into our framework, and to sort of have an opportunity to discuss each step in the framework. So I will be doing a presentation with transparencies that may to some degree overlap some of the material you have also already heard from Dr. Flamm. But the purpose will be to try to sort of keep us on course in using an evidence based approach. Is that agreeable to everybody? Is everybody comfortable with that approach?
Sean reminds me that before we jump into that, I think we ought to give the panel members a response to comment on the past two hours what they've heard from Blue Cross/Blue Shield, from the VA, from Dr. Lau and his colleague, and also from Dr. Gambhir, so if there are any reactions or

.00243

anything that people would like to say about this cornucopia of information that we've heard, this would be a good time to say that. Yes?
DR. FEIGAL: Yeah, I will say something.
I think it's helpful to get useful technology assessments from a credible group of individuals that have clear-cut criteria and it's up-to-date literature that they're looking at. I think that can be very helpful. I think what we've also learned is you have to be careful about the questions you ask to some of your consultants. I think there's some tendency to look at the trees instead of the forest issue, and I think, you know, Sean and I have talked off line about the helpfulness of some of that type of information. I think trying to look at the broad picture, trying to look at the preponderance of data without getting into each individual study and whether it was 51 patients or 52 patients, that kind of information I don't find extremely helpful.
But I think more of the broad overview with particularly the Blue Cross/Blue Shield TEC

assessments, those type of assessments I found useful.
DR. SOX: Thank you. And Manuel?
DR. CERQUERIA: Well, I would sort of like
.00244
to comment that certainly the data that we've heard has been for the most part supportive of the indications for PET, which I think is pretty amazing that that's come through all of this. I think we've also heard that the criteria that you use is going to determine what you pull up, and the VA used one criteria, Blue Cross/Blue Shield used a second criteria, I think the UCLA group did a different criteria, which was basically what HCFA asked to provide them. So the methodology, I think, needs to be a little bit more specific in what, if you're going to do a meta-analysis from the literature, you have to -- you know, you've defined a process, but you need to define what kind of data you're going to put into it, or how you're going to select it. I think that would certainly be useful for people in the future that are going to present, and I think it would be useful for the panels as well as the Executive Committee, to decide how they're going to make their decisions.
So -- and you know, I think out of fairness to the submission, they didn't know that the submission was going to be, you know, handled in this particular way. They didn't know what the criteria were that they were going to be held to. So I think
.00245
the fact that their data, you know, I think is supportive of the indications is very commendable, but I think somewhat unfair to the way they've been asked to submit.
DR. SOX: Leslie?
DR. FRANCIS: I was puzzled by the questions asked the New England Medical Center group, because it seemed to me that the question that I really want to know the answer to is not, is there some excess stuff in here, but is there any good stuff?
MS. RICHNER: Exactly.

DR. TUNIS: Let me just to -- first of all, on the New England Medical Center critique, what we at HCFA felt we were facing was what looked to us or at least what we were trying to figure out is, can this be looked at as, you know, 22 or 26 separate requests for coverage, or can we look at this as a broad coverage request. And so we did a significant amount of, and committed a significant amount of internal staff to reviewing the information that was submitted.

On a parallel track, to make sure we didn't get, you know, too afoul of our 90-day time line, we felt okay, we could use some help with this,

.00246

there's these evidence based practice centers. And essentially the question we asked them boiled down to, can the submission be evaluated as sort of a typical meta-analysis systematic review? In other words, can we base our judgments directly on this as that kind of document? And the -- you know, and so -- first of all, any flaw in the New England Medical Center report, if you will, or any critique of it, really is on the shoulders of HCFA, because they gave us the answer we asked for, so that should be clear. And you know, no one should think that by itself, the New England Medical Center report gets substituted for HCFA's response to this coverage request. It is a piece of information, you know, once we requested it in a sort of -- I think I'm not disagreeing with anything you all have said, or even what Dr. Gambhir has said. I just want to frame it squarely that it's sort of HCFA's doing, HCFA's question, we needed extra help, we needed extra staffing, and that's why we put it out there.

DR. SOX: Well, before we start the discussion of specific topics, Sean, do you want to comment on the issue of voting versus consensus of the group? How do you want us to proceed?

DR. TUNIS: Yeah. I guess the only -- in

.00247

thinking about, you know, based on some of this discussion this morning where we were playing with the questions about the framing of the questions to

the panel, particularly the form of a question that says is the evidence sufficient to conclude X. And we talked a little bit about how there is more of degradational qualities of evidence as opposed to some magical line that occurs where there's a yes and a no.

So to the extent that we can get the Executive Committee's consensus or vote on a somewhat more qualitative judgment about the quality of the evidence and you know, I have pitched the idea of potentially subdividing it into groups like inconclusive, suggestive, or conclusive, but that's just one way of framing it. You know, it sort of actually turns a bimodal question into a trimodal question, to be honest. But the notion is, it may not be that useful here to give us a yes/no, evidence is sufficient, evidence is not sufficient, but try more to come to a consensus about how we can apply this framework to colorectal cancer, if that's the exercise we're going to go through, but by showing us how you do that, also giving an illustration about how we should be applying this framework to the other

.00248

requested applications.

And again to emphasize, the alternative to that being how we would need to modify this framework to address the issue of broad coverage.

DR. SOX: So, perhaps I'll need a motion from someone at the end of my discussion, and Sean suggested perhaps we think about the categories of evidence as inconclusive, suggestive, and sufficient, as representing sort of a spectrum of evidence. So at the end of discussion, I would like a motion that we can kind of talk about it, and I think we actually would prefer to avoid the formalities of a vote, if only because they slow us down so much, and we'll just try to get a sense of the group on their response to the motion.

DR. GARBER: Can I just make a suggestion, Hal?

DR. SOX: Sure.

DR. GARBER: I actually am sympathetic to the desire to have three categories, but I hope that

Sean will think very carefully about which words he wants to use to describe those categories. Suggestive, for instance, is something that you could apply to almost everything, and if you could give us an idea of what sorts of categories would be helpful,

.00249

other than the fact that it should be tripartheid rather than binary, I think that would help us.

DR. SOX: So, do you want to think about that and get back to us when we get closer to the point of taking a vote?

MS. CONRAD: While you're setting up, let me read an obligatory statement. For today's committee meeting, voting members present are Robert Brook, Leslie Francis, John Ferguson, Robert Murray, Alan Garber, Michael Maves, Frank Papatheofanis, Ronald Davis, Joe Johnson. A quorum is present and no one has been recused because of conflicts of interest. Thank you.

DR. TUNIS: Actually, Dr. Brook is here in spirit but not in body, as you've noticed, so he's not counting towards the quorum.

DR. SOX: Could I have the laser pointer? So, the first colorectal cancer topic we were going to talk about is the question, does an indurated area near the original incision represent a post-operative scar that's just a bit exuberant, or does it represent a local occurrence? If it were scar tissue, presumable you wouldn't intervene; if it was a recurrence, you would reexplore the patient with a hope of a curative procedure. The

.00250

alternatives certainly include doing a biopsy of the area, which is invasive and uncomfortable, or doing a test that can reduce the probability that an indurated area represents recurrent cancer. And perhaps if that test were negative, to simply watch the patient, and if it were positive, to do a biopsy. So one of the questions for us to think about in trying to decide on whether the test could alter clinical outcomes is how low would the probability of recurrence have to be in order to defer biopsy? Would we defer biopsy only if the

probability of recurrence was 1 percent, or would we perhaps be willing to defer it when the probability was 10 percent or so?
So, following now after posing the question, following our framework, the first question is, is the evidence adequate to determine that something about the use of PET scan performance -- trying to reframe it the way Dr. Brook suggested. So then the question is, are there high quality studies of the performance of PET scanning in detecting local recurrence of colorectal cancer?
And I relied upon the Blue Cross/Blue Shield evidence report when I put this together, and they did not describe the diagnostic reference

.00251

standard, so we really don't know whether that represented biopsy or surgical exploration with histology; that's an unknown. And if anybody knows that evidence, that particular piece of information, it would be helpful for the panel to know that.
In five of the out of the six studies, the patients were patients with suspected local recurrence, which is the appropriate study population. So, it seems like a reasonable study population. None of the six studies evaluated the PET scan with observers who were blinded to other clinical data, which would tend to cause an overestimation of sensitivity and to underestimate specificity. And finally, four of the six studies were prospective.
So let me stop with this sort of first step and ask what people's take is on this evidence, whether it represents good quality evidence or marginal evidence, or what. Alan?
DR. GARBER: I will take a stab at it. I think it falls short of ideal but it's enough to convince me that it's adequate to make a decision that it increased accuracy.
DR. SOX: And could you explain your reasoning for the rest of us?

.00252

DR. GARBER: Well, the blinding is an important defect. The lack of a reference standard I

discounted somewhat, because I suspected that they probably always had histology in some form, and I didn't think the blinding was sufficient.
DR. SOX: Well, VA group, do you remember what diagnostic reference standard they used for --
DR. GARBER: Carole's right here.
DR. SOX: Oh, I'm sorry. Carole?
DR. FLAMM: It was biopsy.
DR. SOX: So there was a satisfactory gold standard which -- is -- does everybody feel comfortable with Alan's assessment of that? Okay. The next question, which may or may not be entirely pertinent because as we will see in a minute, PET scan performs better than CT, does PET accurately identify CT negative patients who have colorectal cancer? In other words, does PET complement CT? And the studies show that PET scan had a higher sensitivity and specificity than the comparison test in four out of four of the comparative studies, which is pretty strong prima facie evidence that it picks up patients that are negative to the comparison study. However, zero out of the six studies provided direct information about

the ability of PET to pick up patients that were negative on the comparison tests.
So, the next point is, does an indurated area, just to rephrase the question, does an indurated area near the original incision represent scar tissue or a local recurrence? Query, does a negative PET scan lower the probability of recurrent cancer enough to alter the decision to biopsy? The pretest probability of recurring CRC in an indurated area is high, 70 percent basically, and presumably one would either operate or biopsy if the probability is that high. The question is, does PET scan lower that so that you would in fact decide not to biopsy? The pooled sensitivity of PET scan is 96 percent and the specificity was 98 percent. Test performance doesn't get much better than that, but notice that as was pointed out by Dr. Flamm, the pretest probability is pretty high.
DR. FERGUSON: Question. Hal?

DR. SOX: Yes, John.
DR. FERGUSON: Are we talking about recurrence in the scar tissue on the skin or in the bowel?
DR. SOX: I think it's underneath; it's the bowel, I believe.

.00254

DR. FLAMM: It must be, yeah.
DR. FERGUSON: Okay. So one has to open up somebody in order to biopsy, okay.
DR. SOX: Thank you. So far we've said that evidence about test performance is good enough for us. There's some prima facie evidence that PET scan picks up patients that would be negative by some other test, and then the next question is, how much does a negative PET scan alter clinical management? The first step in evaluating that question is to calculate post-test probability of recurrent cancer given a negative PET scan, and that is found here. This curve represents the probability of CRC given a negative PET scan for various values of the pretest probability. And the pretest probability, average pretest probability is about .7, which corresponds to a post-test probability of about .8.
So the next step in our reasoning then would be, is the probability of recurrence of .08 low enough so that we would undergo watchful waiting rather than biopsying a patient? Any questions so far or comments so far?
So the way I thought we could frame that question to try to get at this question of alteration, or effect of the test on management

.00255

strategy, is the following: Since the pretest probability of recurrence is .69, the post-test probability after negative PET scan is 8 percent, if recurrent cancer is present despite a negative PET scan, the patient will forego or at least delay reoperation, which has a 20 percent chance of curing the patient. So that's some measure of the health effects of a correct decision about whether to explore or to do watchful waiting, and those are pretty high stakes, I think we'd agree, at least I

think they are.
So, we could frame the question this way.
Would you biopsy the patient, the indurated area, if there was a 70 percent chance of recurrent cancer?
Well, would you biopsy if there was one chance in 12, or an 8 percent chance that it represented recurrent cancer? And if you would biopsy the patient at both of these probabilities of disease, then you could argue the PET scan really hasn't altered your management, and should you do it. Or, would you just observe the patient and if you would just observe the patient, then PET scanning would affect your management and it would be appropriate to do. So I don't know whether I framed that correctly, but it's out on the table for discussion. Leslie.

.00256

DR. FRANCIS: How about if you were the patient, would you want to have the information about the differential chance in order to decide depending on how invasive the biopsy is? That is, if I were a patient, I might want to know that data breakdown, particularly if the location of the scar or suspected recurrence was one where the biopsy would be quite invasive.
DR. SOX: So, I --
DR. FRANCIS: The way I'm putting the question differently --
DR. SOX: Would this information be helpful to you, to know that it was an 8 percent probability?
DR. FRANCIS: Well, yeah. I think the question is, when you say your management, would this be information that a patient might want to take into account in making a choice about whether or not to have the biopsy, particularly given the fact that some biopsies might be quite invasive and others not.
DR. SOX: Well, of course, we can tell the patient what the probability of recurrence is given a negative PET scan without doing the PET scan, right? So, in other words, if we do this test and it's negative, the probability of your having a recurrence

.00257

is only 8 percent; given that information, would you

want us to go ahead with the biopsy anyway, or would you prefer to just kind of watchful wait?
DR. FEIGAL: But you might pick up recurrence.
DR. SOX: Pardon me?
DR. FEIGAL: Without doing the PET scan. Yes, you might tell the patient, if it's negative, you have an 8 percent chance of recurrence. What you can't tell them is what that PET scan will show. It might pick up the recurrence. You can't predict that.
DR. SOX: Of course there's a 70 percent probability of having recurrence even before doing the PET scan. Alan?
DR. GARBER: I think Leslie is talking about what some of the critical unknowns are here; one is what is the risk of doing the biopsy, what are the down sides? The other thing is, what are the consequences of watchful waiting if in fact a tumor recurrence is present. Now, if you were just to take things at face value and say you really miss it if you -- in other words, watchful waiting is a very dangerous strategy if there is actually a recurring cancer present, then you're multiplying the 8 percent

by one-fifth, which means you would have a close to 2 percent chance of just missing something that would otherwise be cured. And I would contend that it's likely that no matter how inaccessible the location is, under those circumstances, you would always biopsy.
And in fact, that was the discussion that we heard from oncologists before on this very subject. But the issue is, do we really know anything about what happens with watchful waiting with recurrence and the last time I heard this discussed, there wasn't really much information on that subject. But I would guess from my discussions with patients, I agree with the conclusions of the Blue Cross/Blue Shield report that most patients would want this biopsy even if the PET scan were negative. In other words, you would biopsy regardless of the results of the test.

DR. SOX: So you're saying you think most patients would take a one in 50 chance of picking up a potentially --
DR. GARBER: They would not be willing to tolerate the one in 50 chance if they had the PET scan.
DR. SOX: Of missing an opportunity for a
.00259
cure?
DR. GARBER: Right.
DR. SOX: Kathy?
DR. HELZSOUER: I think there's a body of literature to support that level as low as one percent or even less than that, that people will go for that chance for a cure, so the margin is very small, and with very little tolerance in oncology patients to miss that chance for a cure.
DR. SOX: Well, that's really important input and I guess if I hear you correctly, you would argue the PET scan is not going to make much difference in this instance. Positive or negative, the patient is still going to want to go for biopsy. That's your clinical opinion as an oncologist?
DR. HELZSOUER: Yes.
DR. SOX: Let's see. John, then Sean.
DR. FERGUSON: This same questionable PET scan, should we do it or should we not do it, might also at the same time as telling whether this scar is a recurrence or a scar, might also show that there is something elsewhere in the body, and therefore, the equation is changed by that very same PET scan, so I think it becomes a little more complicated, at least to me it does. If I say well, we have CT evidence
.00260
that we've got a recurrence in the scar or that there is something there, and maybe we should biopsy it, and then we say well, from everything else we've seen today, gee, the PET scan might tell us if this is tumor or not, and might also tell us about distant metastasis and liver involvement.
DR. SOX: Sean.
DR. TUNIS: I guess this is a question I actually want to direct to Dr. Phelps, who I guess

stepped out, but Sam, if you can answer, it's basically, it's sort of the question that precedes this discussion, which is, it seems to me at least that without the reliable information about the sensitivity or the specificity of the PET in this case, the empirical evidence, it would be hard to go on and have the discussion about the clinical utility in any particular patient's case, whether it's for, you know, reassurance purposes or for decisions about biopsy, et cetera. And when I, you know, what I'm posing to you is let's say we didn't have that empirical evidence in this case, colorectal cancer. We happen to, but we know we don't have it for some other cancers. How does one have an intelligent discussion about clinical utility without the empirical evidence, and particularly, how does your

.00261

whole argument about this is a molecular approach as opposed to an anatomic approach help us with that, because that seems --

DR. GAMBHIR: I think that's a very important question, and part of the way you can answer that question when you're lacking the exact sensitivity specificity for a given application within let's say colorectal cancer, you can look at the sensitivities and specificities of the other applications within that disease category, as estimates of what you would probably observe. This is this whole issue.

And why is that by the way? The reason is what causes the sensitivity specificity problem, why it deviates from a hundred, has to do with the molecular reasons for the tumor and where in the body you're looking, that is, where is there background signal that confuses your interpretation, right? The specificity leads to false positives due to background signal and the sensitivity relates to what lesions are you capturing based on the molecular properties of the tracer localized.

So the way we usually answer this when we build decision models, and we don't have enough direct evidence for the sensitivity and specificity

.00262

for that specific case, is we look to the next closest relative, if you will, based on that region of the body or that type of cancer, or a similar cancer type. For example, colorectal will behave similar to, in terms of its FDG uptake, to let's say lung, and prostate will behave similar to pancreatic in terms of the amount of uptake. So there are lessons to be learned from other cancer types and keeping in mind the fundamental mechanisms, and that's what we would do.
That's where it is. I think what you're getting at, the deeper question is when you switch to these other categories where there isn't as much evidence, the broad coverage issue, what do you plug in for your sensitivity specificity, what do you plug in and what are your best guides for it. And what I'm arguing is, those best guides are obtained by looking at cancer as a continuum and looking at it based on molecular reasons and the location of the body.
DR. SOX: Thank you. Let's continue the panel discussion of this application a little bit longer and then I would like to go to the second application, and have as much discussion as we can before people leave. Bob?

.00263

DR. MURRAY: I'm a little uncomfortable with this discussion because the original question was, is the evidence adequate? We saw the sensitivity is 96 percent, the specificity is 98 percent, and the answer to the question is yes, the evidence is adequate. And now we've fallen into a discussion of how is that going to change the management, and that's a question for a psychologist, or a question for somebody who asked the research question and tracked patients, and looked at their responses. I don't think that's a question for this panel.
DR. SOX: Well, we're trying, to go back to the discussion we had in the first hour, we're trying to make decisions or make recommendations about diagnostic tests in the same framework as we do for other technologies, which is to try to frame it

in terms of health effects. And because diagnostic test studies only give you sensitivity and specificity, we get into what we've just done, which is to try to infer effects on management strategies and the effects of those strategies on health outcomes. You're right in a way. It does come down to trying to understand something about patient attitudes and preferences. But at least in this
.00264
instance where a negative test might lead to watchful waiting, we've heard from Kathy, who's an experienced oncologist, that most patients are willing to take a pretty low chance on a procedure that could give them a cure.
DR. MURRAY: There are many other aspects of this question that we have not discussed. We haven't talked at all about cancer staging, what was the original --
DR. SOX: But we're going on to cancer staging as soon as we're done with this discussion.
DR. MURRAY: Okay. The question of the age of the patient, of course, you know the Medicare population is going to be at much higher risk, my recommendation is that we note that the evidence is adequate and there are all of these other issues which we are not addressing or which we are only giving a, you know, taking a stab at.
DR. SOX: Yes, Manuel?
DR. CERQUERIA: I don't see many patients referred to me as a cardiologist who have cancer problems, but I certainly have a lot of patients who have cardiac problems who are in this situation, and making decisions about open biopsy, the chances of an 8 percent recurrence versus a 70 percent will
.00265
influence what I do in terms of the diagnostic evaluation of the patient, how aggressive we're going to be with intervention. So that, you know, as a consultant for a patient with this sort of problem, it would help me to make the decision in terms of how to manage them.
DR. SOX: Frank and then Leslie, and then I would like to suggest that we write down something

about what we think about the evidence for effect on test performance and also on clinical effect on health outcomes. Just write it down, we can come back to discussions of voting, but I want to get on to the second application. So with that, Frank?
DR. PAPATHEOFANIS: Sure. I just wanted to echo what Bob said, and I am not an oncologist, Kathy, and I respect your one percent tolerance of what patients may want done. And Bob Brook isn't here to talk about appropriateness, and so in lieu of that, what a patient may want obviously under any given circumstance and what is reasonable and appropriate sometimes are two different things. And as Bob said, with a 96 percent and 98 percent plus accuracies that we're seeing up there, that's pretty darned good.
The alternative would be every patient who

.00266

has a possibility of recurrence just gets a CT scan and a biopsy, you know, cancel all research in every other area because we're never going to be able to get any diagnostic test that's a hundred percent.
DR. HELZSOUER: I think there are two issues here, the accuracy of the test and then the interpretation of that test, and that's where the effect on health outcomes come in. If you're going to take that into consideration, that's what we're doing, so if it isn't really going to change the management at that point, then you have to ask yourself is it worth doing. And I think this is why it's hard to be very broad in the coverage when it comes to cancer, because despite the goal that you're going to have a molecular basis, you're going to tie all these cancers in as one type, they are very heterogeneous, the management is different, there are some cases where the diagnosis itself is not worth knowing if it's not going to change outcome, you don't want to live with that diagnosis, and these are all issues that are extremely important. But it does mean, I think, that we have to go site by site and question by question to look at it; it's not just a matter of accuracy, it's interpretation.
DR. FEIGAL: And I think you might be

.00267
overstating the case, if there's not a treatment option, it may not be worth knowing. I think that you really can't speak for all patients with that type of comment, and I think that for some -- you know, I think we need to think about the patient planning and decision making as well as the health care giver decision making on this. And I don't mean just feel good because you have a diagnosis, but I think that it gets into issues -- the PET scan, as we talked about, may not just show the site of local recurrence, it may show up other metastatic sites of disease, so I think that would be important information to have.
DR. HELZSOUER: And I'm not saying that it wouldn't, but I think that you can't say broadly that that is the case, that's the point.
DR. SOX: Should we go on to the second discussion? Remember, kind of write down your impression, or whether you think the evidence, the test performance is good, is of reasonable quality, and also your impression about whether the test would actually lead to important changes in health outcomes. Leslie?
DR. FRANCIS: On the health outcome point, I haven't really heard anyone respond to John's point
.00268
about discovering distant disease, and whether that's a likely management change.
DR. GARBER: We're talking about a -- there's a separate indication to look for metastatic disease apart from the indication of scarring, and what you are now raising is the question of incidental finding of distant disease, when the prominent feature was the scarring.
(Inaudible comment from speaker.)
DR. GARBER: What's that?
SPEAKER: It's not incidental; it's --
DR. GARBER: It's in a different population where it's being done for the purpose of finding out --
SPEAKER: It affects patient management.
DR. GARBER: Okay, agreed. But the point

is, it's an indication in a patient with a scar, rule out tumor, do you find other distant spreads. And we had a separate indication that we discussed about looking at spread of colorectal cancer and monitoring response to treatments. And one question for Carole, did these studies report the findings of spread elsewhere as a result of looking in this population of people with a scar?

DR. FRAMM: I think it's a good question,

.00269

but let me make one comment first. Here we are with an unknown soft tissue and we don't know whether it's a tumor or scar. I think the question you're begging is, once we've done the biopsy and we know it's scar, then a PET scan might be indicated for looking for multiple things because you have a potentially resectable local recurrence, and you're getting into a little bit of an analogous situation where we've seen in the other body of evidence that yes, PET can pick up hepatic and extrahepatic sites 20 percent of the time, 30 percent of the time in the population of patients who have an isolated liver recurrence. So why, because your recurrence is at the anastomosis, is that so very different from recurrence at the liver site? Okay, that was one little comment. But, you weren't going to like the answer to my question, that's why I did that first. I don't think that I can answer whether those studies overlapped in terms of -- because we would parse out that piece of information and put it off in the other part of the assessment where we looked at staging and extrahepatic mets, and I could go through and look at the names of the studies and see whether they did that, but I don't have a straightforward answer for you.

.00270

DR. FEIGAL: Yeah. My only comment, I think we're trying to neatly categorize things in a patient who doesn't neatly categorize their disease. And I understand the question you're trying to answer, but it may be that you get more information than you intended and then what do you do with it, does it actually change your management? And I guess

what you're saying is you don't have that information.
DR. SOX: And that's actually an important issue, because when you get -- when you do a PET scan of the whole body to look at the scar, and you see something down here for which there is no clinical evidence, the prior probability is low and therefore the post-test probability, even with a test as good as that, is going to be relatively low. You may find stuff that ultimately turns out not to be important clinically, but causes anxiety and more biopsies and the like. Mike, did you --
DR. VALK: Excuse me. I'm sorry to interrupt, but I think I have to at this point. The positive predictive value of a skeletal, focal skeletal lesion in some of the metastatic disease, even if the patient is completely asymptomatic, is very high. It's probably going to be 90 percent.

.00271

And so the worry here really shouldn't be do you cause them unnecessary anxiety. The issue is, you've almost certainly picked up an asymptomatic metastasis, and that's how you should manage it.
DR. SOX: Thank you. Mike?
DR. MAVES: I don't think this will help at all but you know, the other thing is, we're assuming that the biopsy you would get would be sort of all knowing and all telling, when in point of fact we understand, particularly in a scar tissue area, you may in fact have disease but not be able to obtain a positive biopsy. That happens as well, so I have a little trouble wrestling with the question that you put up here, Hal, on an on the ground basis. I mean, I've operated on people that had far less than a 69 percent chance of recurrent cancer even in some fairly inaccessible areas for all the reasons that we talked about. I've also understood that even getting negative biopsies in some of those situations may be just a limitation of histology, human technique, and sort of just human frailty. So it makes it, in my opinion, makes it a tough -- you know, my answer would be yes, it is an accurate test, but in the situation we've put up

here, I think it's a difficult one to say what's the
.00272
best way to address this, because there's enough uncertainty even on the biopsy side in this kind of instance that you might find yourself doing both to simply cut down that uncertainty, particularly if it's in an inaccessible area.
DR. SOX: Should we go on? Good discussion, time to move on.
So, the next question is, does PET scanning provide useful information about th extent of additional metastatic disease in patients in whom another imaging test shows a potentially resectable metastasis? The goal of testing is to improve the selection of surgical candidates so that preferably, nobody who has an unidentified metastasis gets exploratory surgery.
Our key questions then are, is the evidence adequate to determine that use of PET scan provides more diagnostic information which breaks down to these two questions: Are there high quality studies of the performance of PET scanning in detecting metastatic colorectal cancer, and does PET scan accurately identify patients who have additional metastases not detected by CT? And then subsequently, if the test improves accuracy, is the evidence adequate to conclude that the improved
.00273
accuracy will lead to better health outcomes, both by altering management decisions and by altering management decisions that affect patient health care outcomes, by identifying patients who could not benefit from surgery to resect a metastasis.
So, with artistic license here, I tried to frame the problem. Imagine that rectangle is a liver, and the dot represents a single metastasis detected by CT scan. And this represents the PET scan result which could show additional metastases, in which case you wouldn't want to try to resect this metastasis, or if it's negative, it would show no metastases.
Now, this is sort of a way of indicating the patient's true state, which in one case, the

patient's true state is yes, they had the CT detectable metastasis and yes, they also had the PET scan detectable additional metastases. This patient could avoid additional exploratory surgery and possibly an attempt at a partial hepatectomy. On the other hand, if the PET scan was a false positive, and these did not really represent metastases, then the patient would not get a potentially curable surgery. On the PET scan negative side, if it's a true negative, then the patient would go for a

.00274

potentially curable surgery. If the study is falsely negatively and the patient really does have metastases, then the patient would have surgery without really any hope of getting a cure out of it. So that's the problem we're dealing with. So, the first question then is, are there high quality studies about the performance of PET scanning in detecting metastatic colorectal cancer? Here maybe we can get some help from the VA folks and the Blue Cross/Blue Shield folks. The Blue Cross/Blue Shield evidence report does not contain information on the reference test or how patients were selected to get it. I have a note here that it was based on the VA, I guess analysis, that it was a mix of pathology and histological proof that they either had cancer or didn't, or else clinical follow-up, at which the patient eventually would show up as having metastatic disease or not. Do you want to comment on that?

DR. FLAMM: Those were the commonly represented reference standards in the literature. We (inaudible).

DR. SOX: So one question we could ask ourselves, is a mixture of histology on patients who get operation and clinical follow-up on patients who

.00275

don't get operation a reasonable reference standard? My take is it's a reasonable reference standard in the real word if the clinical follow-up is done carefully. The patient populations were appropriate; they were either patients with a suspected recurrence

of cancer or a solitary metastasis discovered at the time of initial staging. A few of the studies blinded those who read the PET scans; most of these studies did not blind them. So the first question I will ask the panel is, what's your take on this, is this reasonable studies of test performance? General nods. Anybody disagree? Good.

Then the next question we could address is, does PET scan accurately identify patients who have additional metastases, specifically does it detect patients whose metastatic disease would be missed by other imaging tests such as the CT scan that was done as a part of routine imaging. And there are several lines of evidence and they all indicate that PET scan does a very good job in this respect.

The best study which we've heard about before showed discordance between other imaging tests and PET in 40 patients, which was 10 percent of the

.00276

total patients who underwent the CT scan and the PET, so fairly frequent discordance. And in 35 of the 40 studies, PET scan in fact led to the correct diagnosis, presumably based on the reference standard test that we discussed just a moment ago. So this result indicates that PET is more accurate and adds information compared to the imaging test.

PET scan correctly upstaged 15 patients, who therefore didn't get an operation, because they had worse disease than was originally assumed, and PET scan correctly downstaged six patients, presumably by being negative on additional metastases, and they got a potentially curative operation. Valk et al. compared PET with CT at various sites; the study results indicated discordance between PET and CT in 40 percent, and PET was correct in 90 percent of the discordant results. What was the gold standard in that test, sir?

DR. VALK: It varied depending on the site. For the positives of course, the best gold standard you can get is histology. We did have histology except in a few patients who had multiple

lesions in whom surgery was not undertaken, and there we used progression on subsequent imaging studies.

.00277

If you are trying to validate a true negative, then of course a negative biopsy doesn't really do it for it, you may have just missed the lesion. And if you want a true negative validated, then you have to do follow-up and that's why we did at least 12 months follow-up on everyone who appeared to be negative by PET.

DR. SOX: Thank you. And just a last comment, Dr. Valk compared PET with CT at various sites, along with a reference standard as he just described. In every instance, sensitivity and specificity of PET was better than CT, although I note that for a few applications, the sensitivity wasn't terribly good, particularly in the abdomen, where a negative test wouldn't necessarily exclude metastases. But overall, it appears that PET definitely does add complementary information to the usual imaging tests. Anybody take issue with that?

DR. PAPATHEOFANIS: No issue, but in your numbers there, in 35 of 40 instances, it shouldn't be 80 percent, it should be 90 percent, PET was correct.

DR. SOX: Seven-eighths, you're right, thank you.

MS. ADAMS: Hal, could I ask a question?

DR. SOX: Please.

.00278

MS. ADAMS: When we are talking about PET scanning, are we talking about the dedicated PET scanners or are we talking about camera based PET scanners? There are a number of different hybrid models, modified systems. Are we, just the data that's presented is just dedicated scanners, a point of clarification.

DR. SOX: Does anybody have the answer to that question?

DR. FLAMM: I know that for the Blue Cross/Blue Shield assessments, we did restrict to only dedicated PET performance data. I think the question is still a good one, that maybe this audience is a PET audience with PET cameras, but

there certainly is the question of what's happening in practice with FDG imaging.
SPEAKER: Sam, do you want to talk about --
DR. GAMBHIR: Briefly, all colorectal data published, all research articles are on dedicated PET systems. If we go across all those articles in the HCFA requested report that we prepared, about 5 to 7 percent of the abstracts and articles combined are from what are called nondedicated PET. These are systems who are a little bit lower in cost and there

.00279

may be a slightly smaller sensitivity and specificity compared to the dedicated PET systems, but they are still a minor portion of the actual data we have available on accuracy.
DR. MURRAY: Could I ask a follow-up question? Of the PET scanners installed in the past year or two, what percentage are the camera based as opposed to the dedicated?
DR. PHELPS: Actually, that question is beginning to change rather rapidly, the answer to that question, because initially the camera based systems were devices developed for techniques that are all gone, so we've gone to thicker crystals to increase the sensitivity, reduce the noise by about a factor of four, so they have improved. There are also dedicated systems that are dual head systems that have a higher efficiency and equal resolution than anything you have seen here, so products are being developed with a clinical purpose. To go back to your direct question, probably about two-thirds of the systems now are in the category of dedicated, one-third to the nondedicated, but the growth is higher in nondedicated, but you have to be careful about what that means, because the performance of those are much higher than the initial cameras.

.00280

MS. RICHNER: What is the difference in price between a dedicated versus a camera based?
DR. PHELPS: Yeah, that's changing too. It used to be about five years ago, a PET scanner would cost about 2 to 2.3 or 4 million dollars.

Today the high ends are only about 1.3 million. And in fact dedicated systems, you can buy for 7 to $800,000 today. The cameras are around 500,000, and those are the high efficiency cameras.
DR. PAPATHEOFANIS: Can you put that in perspective with a CT scanner?
DR. PHELPS: Yeah. If you look at a CT, CTs are about 400,000 to about 800,000, some of course over a million dollars, and MRs are about 600 to 1.7, 1.8.
DR. MURRAY: In the VA 1998 follow-up assessment, there is, on page 2, there is a significant difference in the sensitivity, but you're telling me that what they were comparing to camera based are an earlier generation long gone?
DR. PHELPS: Right.
DR. MURRAY: Okay.
DR. SMALL: I just wanted to mention, the dementia studies I described were from dedicated scanners.

.00281

DR. SOX: Thank you. So now we need to turn to the question of, would information about additional metastases alter patient management, presumably by making a decision not to do hepatectomy or wedge resection. And any clinicians want to comment on that? I see Mike has left. My take would be yes.
DR FERGUSON: Yes.
DR. FEIGAL: Yes. You wouldn't operate on a patient with multiple metastases.
DR. SOX: And then the question would be, would that management strategy lead to improved health care outcomes and presumably for patients whose stage increases as a result of PET scanning, they could avoid the morbidity and mortality of surgery, and patients whose stage decreases as a result of PET stand could undergo a potentially curative procedure that they wouldn't have undergone had PET scan not been done.
So, any discussion about how you think the evidence shapes up in this particular application of PET scanning? What's the overall take? Anybody want

to step up to the plate?
DR. FERGUSON: I guess I just have a conundrum on the business of recurrence or a scar. I

.00282
can't get away from the fact that if -- I agree that if we knew that that was possibly a scar or a recurrent cancer and that was the only thing, that you would go in and patients would probably want it because there's a possibility of a cure. And if you knew -- if you did a PET scan and that was the only thing that showed on that PET scan, or even if nothing showed on the PET scan, you still might go in and try to remove that with the possibility that it was a false negative and that it was a cancer and you could cure this patient. And that same patient, you would have in the back of your mind a nagging thing, we didn't do the PET scan because it wasn't going to change our management. On the other hand, you just had a patient that day where you thought that was the only thing, and you did a PET scan and found other things, so there was a change in management. It's hard for me to escape, knowing that there is a 60, whatever, 70 percent chance of recurrence, that if I don't do a PET scan because I know I'm going to go in there and do that operation anyway, PET scan or not, that I might find something that would change my management. And that to me is, if I say I'm not going to do that PET scan because I'm going to go in there anyway, but I know the PET

.00283
scan might possibly change that, that's a conundrum for me, that makes this sort of a -- that brings these two situations very close together.
DR. SOX: Alan?
DR. GARBER: Well, I think as John and Ellen pointed out, there is that issue. In a patient with the scar rule out tumor, recurrence, whatever you want to call it, is that a high prevalence population for metastatic disease for distant metastases and unfortunately, we don't have the information, but it might very well be that that's the main reason to do a PET scan, rather than finding out whether this particular scar is indeed cancer.

And Carole handed me -- I hope I'm interpreting this correctly -- there is one study that looked both at the post-operative scar and distant metastases, and if I understand these numbers correctly, there was in fact a high rate of distant metastases in that population. So that does suggest that in a scar, you might think of the diagnostic issue of not wanting to determine what the scar is, but identifying the population at high risk for distant metastases. And from that point of view, that group may be, it's not really an interesting question, whether the scar represents tumor or not.

.00284

It's what you do. And the Schiffer study, which unfortunately is the only study they had that looked at that issue, if that's representative of it, that's a very high rate of distant metastases, so that would be a reason to do the test.

DR. SOX: Okay. Just to finish off the discussion of the use of PET in colorectal cancer where there's a potential for resectable metastases, are there high quality studies? I think we agree that at least there were adequate studies.

Does PET accurately identify patients who have additional mets not detected by CT? I think the answer is pretty clearly yes.

Is there evidence that the improved accuracy of the test will lead to better health outcomes? I don't know; I sense that the group's feeling is that the number of additional patients identified with metastases is good enough so that this would in fact alter management decisions.

Anybody want to take issue with that? Mike?

DR. MAVES: I don't want to take issue, but I think actually you get information well beyond just distant metastatic disease. I mean, even though this is a functional test that shows you the function of those tissues, it's certainly going to be able to

.00285

delineate where that tumor is at, and you may gain additional information even on the local resectability. As I thought about this, it is not so much an either or proposition, biopsy or PET, I think

there is actually -- and I don't know if this is a problem, but there's information to be gained from both, they are complementary, and may well be able to help you in many instances of localizing where that lesion is, particularly not so much with colorectal but in head and neck, the location, the accessibility, inaccessibility, are all local questions that you get information from PET that may be just as helpful as evidence of distant metastatic disease, or even regional metastatic disease.
DR. SOX: Is this an issue that's been studies systematically?
DR. MAVES: This is clinical empiricism here, but also I think if you look, there were some materials in our handout, not on head and neck, but I think showed some areas where they had gone to and looked at that.
DR. FEIGAL: And it wasn't just surgery, it was also helpful in treatment planning for radiation therapy. But it does hit at the issue of local.

.00286

DR. SOX: Okay. So, any more discussion of these two issues? Is there anybody from the audience who would like to make a comment before we move to some sort of formulation of a consensus? Yes, Dr. Valk?
DR. VALK: Just one thing. In the present coverage policy for Medicare and colorectal cancer, there is a remarkable anomaly, and that is that PET is approved for imaging for recurrent colorectal cancer provided the patient has an increased CA level. If the CA level is not elevated, regardless of whether you can feel a pelvic tumor or whether you go to biopsy which shows a lesion, or whether the CT shows lesion, if the CA is not up then Medicare doesn't cover the PET scan. That I think you would agree, is a remarkable anomaly.
DR. SOX: Go ahead.
DR. LIEBERMAN: I'm Dr. Lieberman and I'm a surgical oncologist at Sammons Cancer Center, at Baylor Hospital in Dallas. It has been an extremely valuable complement to my practice and where you're

going I think is also very strong. The one comment though that I would make is that most cancer centers work in a very multidisciplinary way. We meet, we discuss each one of these problems before we order a
.00287
PET scan. We're lucky we have excellent equipment and excellent physicians who interpret the PET scan, and that's just incorporated into the patient management, so it's a continuity of care. So as a surgeon, we get to a point where we have patients sent to us with a liver metastasis, or we have a scar after colorectal surgery, similar to a scar in a patient where there is cancer and you're worried about local recurrence and there's no sign of symptoms (speaker was inaudible) or a questionable CAT scan or a single liver metastasis, or a mass in the rectal perineum with a normal CEA who has had liver cancer.
So in a multidisciplinary way there is a decision point, an inflection point that occurs, and the PET scan is a value added, there's no question. Surgeons over the country, the letters that you got, the clinicians, the PET scan is proven because of its biologic testing to be value added, not to replace. It does replace CAT scan at a certain time, but where it is used in a clinical setting is to help the surgeon and the oncologist and the radiotherapist recommend to the patient what they should do with this scar, whether or not it's PET scan positive or not. The patient is incapable of making this complex
.00288
decision, but the multidisciplinary care of the cancer surgeons and oncologists can.
I think it's going to boil down to the fact that we haven't been allowed to study all these questions that you have, and you're going to have to trust the medical profession with patients in a multidisciplinary setting, and as I understand, all the PET scanning centers are data collectors. We assume we are analyzing our cases, but we know that the data collection is being done, and it's going through medicine.
We've seen things like gastric freezing.

It was done for a couple of years and then everybody realized it's not any good, but we're going to have to get this testing done on a broad basis in order to find out. I think that this biologic testing of tumors, I don't know of an oncologist, surgeon or medical radiation oncologist who doesn't feel it's a very big advance in the care of patients. Thank you.
DR. SOX: Thank you. Further comments before we try to formulate a consensus?
DR. HOVERMAN: Hi, Russ Hoverman with Texas Oncology. Just two comments on --
DR. SOX: Excuse me, sir. Could you restate your name and your affiliation for the
.00289
reporter?
DR. HOVERMAN: Sure. Russel Hoverman. I'm with Texas Oncology, a physician group in Texas with 200 oncologists, and I have no reimbursement relationships to PET scans.
Two points. One is, there was a study done a number of years ago that looked at what people would do given knowledge about treatment, and it had to do with high dose chemotherapy with breast cancer. A little less than 10 percent of the women would choose high dose chemotherapy if it gave them one month of life. Almost an equal percentage would not even have taken hormonal therapy if it gave them a year of life. So there is a whole spectrum of decision making that is related to the amount of information a patient is given.
The second is in regards to your algorithm about evaluating residual masses or scars with PET scans. One thing not considered is that it may change your whole algorithm. In other words, if you now have a positive PET scan in the face of a rising CEA and you have it on the CT scan, you may not do the biopsy at all, maybe then at that point you should be planning surgery and eliminate one whole step in your therapeutic algorithm. That was not
.00290
considered and may well be cost saving.
DR. SOX: Thank you Dr. Hoverman. Yes, please.

MS. ADAMS: Just a follow-up on my earlier question about the dedicated versus the modified systems. Is the diagnostic performance sufficiently similar so that you can apply or generalize the data presented here to these institutions that use the other systems?

DR. SOX: No responses to your question.

DR. GAMBHIR: The answer is not yet, no. We don't have a subset meta-analysis where you can say here is the accuracy for these dedicated. And the problem is the nondedicated themselves are evolving, there's not one nondedicated system that you can point to; that's actually itself several categories, so we can't easily give you an answer.

DR. PAPATHEOFANIS: Can I address that too, Hal?

DR. SOX: Okay. We really need to move toward this consensus process because people are going to start leaving.

DR. PAPATHEOFANIS: Yeah. Going back to the VA's lung cancer trial, one of the arms of that trial is looking at just those types of cancer, so I

.00291

agree with what Sam just said, we don't know yet.

DR. BALK: Just a question that was raised earlier, the VA group said that the study from Australia showed no -- it was concordant with their results, which appear to be rather discordant with the consensus of the panel so far, as I read it. I was wondering, I was in Australia as a visiting professor a few months ago, and the Australians were moving toward a final opinion. And my sense was since our fellows went back to Australia and now they're putting PET centers in other cities such as Perth and around Australia, that they in fact thought there was some favorable aspects to PET, and I wondered if maybe that should just be mentioned before the committee fully decides.

DR. VALK: In September, in Australia, where I have a particular personal interest, the government approved seven PET centers, and that's for a country with a population of 18 million, so I don't think that indicates lack of support.

DR. SOX: Thank you. I would like a motion from somebody about whether the evidence, and we will deal with the first one first, scar, induration of a scar, possible recurrence at the site of resection. Does this represent inconclusive
.00292
evidence, evidence that is suggestive, or evidence that's pretty conclusive and if so, what does that evidence suggest? So, could I have somebody who would try to frame a motion and we can discuss, and then see if we can get everybody to nod their head without taking an official vote. Leslie?
DR. FRANCIS: I just want to ask you a question before this. When you say suggestive, pretty conclusive, what the significance is of that, I guess if I thought it was pretty conclusive, I would say the Executive Committee could recommend that HCFA cover it. If it's suggestive, one conclusion might be that it should go to a panel. I just want to know what you think the import is of those, or if there isn't any, then we would just drop it right here and recommend that HCFA not cover.
DR. SOX: Sean wants advice not about what to do but how good the evidence is that doing PET scanning under the circumstances we just described, induration of the scar, alters health outcomes. He wants advice about how good that evidence is, and I don't know what he's going to do with that evidence, that's his problem. He might make a coverage decision. So, does anybody want to say something? And since we are in an informal mode, nonvoting
.00293
members have the privilege of the floor to make that proposal if they want it.
MS. RICHNER: My only concern is we were given this information relatively recently, and it was a plethora of information, and new things have come out today that I don't feel confident about, for instance the Australian information, et cetera, and then the discussion we just had about recurrence versus scar, et cetera. There have been a lot of variables and unknowns.
And certainly on face value I would say

that it looks very adequate, the sensitivity and specificity of the exam. But once again, I feel that in a sense, the radiological panel should be wholly considering all this new information that has come out today. And so in a sense, you know, we want to give them advice, but maybe it's not the right time.
DR. SOX: So you might make a proposal for us to talk about it, we might say the evidence is suggestive but not complete enough for us to make a strong conclusion about it, and then Sean might say, well, I guess we better get the panel to work on this issue. That might be one proposal. Anybody want to make that proposal just to kind of get us off the dot, or something like it?

.00294

DR. FRANCIS: I will make a motion that the evidence is pretty good.
DR. SOX: Pretty good or terrific? Okay. Manual?
DR. CERQUERIA: Despite all my criticisms of the process, the evidence certainly looks overwhelming. I thought the Blue Cross/Blue Shield data was very supportive and we haven't really heard anything negative, so if we are going to go forward with the process, we have heard nothing negative, and I would recommend that we approve it for the indication suggested, for colorectal cancer.
DR. SOX: Remember, we're talking about the questioned local recurrence issue. Let's not talk about anything else but that until we're done with that. So Frank, what's your thoughts?
DR. PAPATHEOFANIS: How about prefacing our comments by saying in view of the interim guidelines for assessing diagnostic tests, the evidence appears to support the use of PET according to the questions that we have before us in the setting of colorectal cancer.
DR. SOX: Are you referring to this or the local recurrence?
DR. PAPATHEOFANIS: The local recurrence,

.00295

because we're doing two things simultaneously. We're looking at PET and we're also looking at the

application of these interim recommendations for evaluating tests.

DR. SOX: Okay. Bob, your thoughts?

DR. MURRAY: I think the evidence is conclusive that PET has greater diagnostic accuracy, but I don't think we've seen any evidence that it has improved health outcome, because all we are working with is indirect evidence.

DR. SOX: I agree with you, but since we are probably not going to get direct evidence on health outcomes, we tried to make inferences and I would be interested in your thoughts about whether you think they are pretty convincing evidence, let's say as opposed to the second application, that doing PET under these circumstances would improve health outcomes.

DR. MURRAY: Based on the comments of oncologists and others with direct experience in treating patients who had a diagnosis of this type, the patients then opt for follow-up biopsy regardless of a very accurate test. If it isn't 100 percent accurate, what I hear is that patients will generally opt for the biopsy and in that case there is no

.00296

change in management.

DR. SOX: Is this helpful, this discussion?

DR. TUNIS: Yeah. In a way it's sounding like, and correct me if I'm wrong, but from this most recent comment, that in a sense you might separate the question about the quality of the evidence whether it's direct or indirect into quality of the evidence on test performance, and then the quality of the evidence regarding clinical utility. And I would define clinical utility as whether it changes management, and whether those changes in management might affect outcomes. But rather than separate it into three questions, maybe just the two are enough and you could sort of give a different score for your level of comfort with the conclusion that the test accuracy is well known, or sensitivity and specificity, versus your confidence that that information indicates that it would have clinical

utility.
DR. SOX: He's our customer, so I would like to suggest that each person formulate their own thinking about those two questions, is the evidence adequate to conclude that this test has the accuracy that it says it does, and is the evidence adequate to
.00297
conclude that using the test under these circumstances would improve health outcome. Why don't we just split it like that, and I'm just going to go right down the group to get your opinions about it, and let Sean integrate that information as best he can. Randel, you're first up, or Kathy, did you want to make a general comment?
DR. HELZSOUER: Well, I don't know. I think there might be some clarification given some of the discussion, because, on whether you still want to separate out the two, scar issue versus the metastatic.
DR. FEIGAL: Yeah. I mean, you sort of assumed that all oncologists thought the same, and I thought that you were sort of bringing up the issue as though the discussion hadn't take place. And I think we did bring up the issue that metastatic disease may very well change your management.
DR. HELZSOUER: Let me just say, I think the question now that I think was raised is, can the scars itself be taken as an isolated case, and I think we have reason to question that, given that there is only one study and we haven't had a chance to look at that, and that study suggested that if you have a scar, you're likely to have metastatic
.00298
disease, and that is a very critical issue when we're trying to interpret how that test should be used, because it has a dual purpose in that case.
DR. SOX: How good is the evidence that it's metastatic?
DR. HELZSOUER: We haven't looked at it specifically with that issue.
DR. VALK: There is definitely more than one study.
MS. RICHNER: But we don't have it.

DR. VALK: There's one study that talks about (inaudible) specifically in those terms, but there are three other studies where they don't talk about recurrence at the primary site, they simply refer to it as pelvic disease. Pelvic disease in nearly all cases is in fact recurrent to the rectal primary site or adjacent to the rectum and would be managed in exactly that way. And the prevalence of disease at a second site is somewhere in the vicinity of 20 to 30 percent.
DR. SOX: But we haven't had a chance to review that data, and the general comment that you're making is you just don't know enough to make a recommendation on the second one, and that would be for Sean's meld, so Randel, please.

.00299

MS. RICHNER: Well, I'm going to say once again that I will say yes to the information we received today, that based on we've heard today, the accuracy is good, and the clinical utility, clearly there will be some differences in medical management associated with this intervention.
But I also want to say for the record that I believe that the process should be that this should go back to the radiological panel for further discussion simply because of this fellow standing up just now saying there are more studies. Well, we don't have all the information we need, and I think it's very important that we send it back to the panel of experts.
DR. SOX: So knowledge about metastatic disease might tip your thinking about clinical utility?
MS. RICHNER: Right.
DR. SOX: Frank?
DR. PAPATHEOFANIS: Taking into account Dr. Gambhir's cataloging of the experience and looking at that, at those tables at their face value, I would say that my recommendation is that there is very strong evidence for the diagnostic accuracy aspect of the technology, and very strong evidence

.00300

for its inferred impact on net health outcomes, so

yes to both.
DR. SOX: Bob?
DR. MURRAY: I think that the evidence is conclusive that it is very accurate diagnostically. I come to a different conclusion on the impact of health outcomes. I don't think it's going to have a significant impact.
DR. SOX: Thank you. Joe.
DR. JOHNSON: On the evidence, strong, yes. On the clinical utility, strong, yes.
DR. SOX: Ron?
DR. DAVIS: Well, I agree with Randel that there would be benefit in having the panel look at this whole question in more detail, but to answer the questions, I think there is evidence that the test performs adequately, there is some evidence that there will be resultant changes in management decisions and because of that, suggestive indirect evidence that health outcomes might be affected.
DR. SOX: Manuel?
DR. CERQUERIA: Yes for diagnosis and even though the evidence is a little bit less solid for making changes in management, it is probable, so I will say yes on that as well.

.00301

DR. FRANCIS: Yes, I think it's very accurate for diagnostic purposes. As for the management, with respect to the scar, I want to know more, but I think it's pretty good on the question of whether you would be interested in the correlation to distant metastatic disease. I never heard anybody with respect to the question of metastatic disease talk about the predictive value of a positive test and whether there might be cases in which people would still want to go to surgery on the possibility that they want to try surgery against the possibility that it's a false positive. But, I still think the evidence there is pretty good too.
Since I'm going to have to leave in about five or ten minutes, I'd like to say that I think we should refer to the panel any generalizations about this to other oncological situations, both because I think there are going to be questions about the

accuracy in other oncological applications, but even more importantly because I think the clinical management questions are going to be different in different settings.

DR. TUNIS: Could I just ask, could you clarify on that point, whether you would come to the same conclusion about generalizing your conclusions

.00302

on this issue to other uses of PET for the same cancer but for different clinical questions? In other words, some of the other questions, whether they have to do with diagnosis or --

DR. FRANCIS: It seeps to me there is a fair amount of evidence about a number of applications for colorectal cancer, but what I want is the panel to look at at least several more, pancreatic cancer and head and neck cancer, breast cancer, something like that, to get a clearer sense of the accuracy issues and the clinical management issues.

DR. MAVES: I would say yes for the accuracy. I think the test is accurate and reproducibly so, we've seen good evidence for that. And I would probably say the same thing with regard to health outcomes, provided that I think the quality of health outcomes here may not be so much either or, it may not be test or surgery. As I said before, I think there is value and information to be gained in the two complementing one another. We heard a surgeon here recently discuss the multidisciplinary approach. So, the answer would be yes to the second part with the qualification that it's a broader kind of health outcome.

.00303

If I could comment a little bit, I would also say, I think that this line of logic in the use of the algorithm actually has proved beneficial, Hal, I think this has helped and could be applied to others. And I sort of noticed, I actually marked this page here, I think there's an algorithm, I guess it's in the materials here from the petitioners, a matrix of PET use, which looks like it's about a four-by-twenty matrix. I assume what we're talking

about here is filling in one square, colorectal occurrence, in that four-by-twenty matrix, some of which has been filled in by HCFA already for us. But my sense would be is that I think there is information in all of these submissions that have been given to us today, but the problem I think is it's a little bit like trying to drink from a fire hose. We need to sort of distill this. Perhaps that matrix is the way to look at it. I do think this algorithm helps us, and I would concur with the recommendations about sending this back to the panel with instructions to just do that.

DR. SOX: Kathy?

DR. HELZSOUER: I would find that there is evidence that it improves the accuracy above what's already existing in terms of sensitivity and

.00304

specificity, and I think it would have an impact on management. And I just think the clarification in terms of the test scar, I don't think I would separate that out as being so distinct from the second evaluation based on the discussion. And I agree with what's been said, that you can't make broad coverages, extrapolate from one type to another, both in terms of the accuracy and also the management of PET.

DR. SOX: At least not yet. There's only a couple of careful studies under our belt, but I recognize that. Linda?

DR. BERGTHOLD: I don't think I have too much to add to what everybody has said so far. I agree particularly with Leslie.

I would just like to add one point which is, as the consumer representative on this panel, which is a very odd role that we all play, what I have not heard today, any talk about, is involving the patient much more in the decision making process, the whole process of informed consent but beyond that, sort of collaborative decision making with the patient about what all these risks mean, and I don't know how that fits in, but it seems to me that it has been significantly absent.

.00305

DR. SOX: But not entirely absent.
MS. BERGTHOLD: Not entirely.
DR. SOX: We did talk about it in relation to how patients would view an 8 percent probability of recurrence despite a negative test.
MS. BERGTHOLD: Right. But that's always from the point of view in the audience here from the provider's point of view, and so we really haven't heard from any patients.
DR. FEIGAL: Yeah, I'd like to say yes to both. I think the evidence is very strong for accuracy and I think there is strong indirect evidence for patient management.
I would like to address your issue just briefly about the patient issues because I think they are critical, and I think they haven't been fully addressed, and I think that will come presumably from getting more input from patient perspectives, both on some of the guidelines that you're attempting to put out, because it is the patients that are going to have to put up with these tests, and I think we should listen to what some of their comments are about what's actually required, and give them some autonomy in the types of things that are done to them.

.00306

The other issue I would like to bring up is that of extrapolation. I think more and more, and PET is just one example, there is going to be -- NIH is sponsoring a lot of research in all kinds of innovative technologies, and this group is going to be faced with dealing with technologies that look at functional or biological processes and together, we're going to have to think about developing an algorithm for how to evaluate those that don't fit into our usual algorithm of evaluating anatomic or structural imaging. So I just want to put that out as, it may not be a perfect set of criteria that you set up initially, but presumably it's going to be different than the criteria you have here, because just as a practical matter, you're not going to be able to go disease by disease by stage by condition and go through all this, unless we want to wait

another 25 or 30 years to get the answers. So, those are my thoughts.
DR. SOX: John?
DR. FERGUSON: Yeah, I agree with the comments on the accuracy of PET. I also think that it's suggestive that PET may affect health outcomes in the case of liver and distant metastases, and even in the case of scar versus recurrence, because of the
.00307
question of metastases.
As far as the process goes, what we have done today I think given what we swallowed and digested in the course of about four or five days, I would like to make a strong recommendation that we the Executive Committee not always do primary evaluation on complex material.
DR. SOX: Hopefully, never.
DR. FERGUSON: Even though I understand what HCFA has to deal with, I'm sympathetic with that.
The other thing is, we are using guidelines that we saw a day or so ago, and I think it's a wonderful thing that we've actually done. I'm surprised that we're still standing, or sitting. But I think it's important that we, and I think you do too, Hal, that we try to have these guidelines fairly well digested before we start actually using them. It's sort of like putting a car together and seeing how it rides before we actually check it out.
DR. SOX: So, it sounds like everybody thinks that studies of accuracy are reasonable and that there is some difference about the impact on health outcomes. Most everybody said either yes or a qualified yes, that's what I was hearing. Several
.00308
people suggested that we don't have enough information to make a really strong decision based on the data that we have heard.
Sean suggested that while it's fresh in our minds, we might talk about this framework that we have used today. A number of you have commented on it and my read of what you said, this went pretty well. We need to refine it, but we've learned some

things and we are at least on the right track. And maybe we can cut to the chase in terms of finding a conclusion about it just by seeing if anybody really disagrees with that, the way I just characterized it.
MS. RICHNER: One of the issues I think that still doesn't come through with that is you're talking about improvement in health outcomes as compared with established tests and once again, I think it was discussed earlier about more of an equivalency type of issue. But, it seemed to work when you moved forward with the questions regardless if you said improved health outcomes at the beginning. It's how you quantify and define what an improvement is that I'm concerned about.
DR. SOX: That's the hard part.
MS. RICHNER: Yes. But clearly that comes out later on, but it's important that, you know, we

.00309

don't stop at a no after question one, and because of that undefined quantitative marginal benefit, whatever that is.
DR. SOX: Manual?
DR. CERQUERIA: Well, I would like to reiterate that there is some value to using this model, but I think we need to define a little bit better the criteria for the data that we're going to feed into it. Again, we had so much variation in the selection criteria for inclusion of studies in the various analyses that were performed, and you're going to get that, and you're going to have selection, and I think this committee could give some guidance on what we feel the studies would be appropriate to include, and would help.
I also still think that, you know, there are some things that clinical judgment goes into it, and it's hard to gather the data in a way that is going to be convincing, but yet clinical practice still finds it has merit, and somehow that needs to be incorporated into the process in some way.
DR. SOX: Thank you. Ellen?
DR. FEIGAL: Yeah. The other caveat I would like to add is, a lot of us are used to dealing with therapeutic interventions, in which there's a

.00310
large industry, there's biotech, there's pharmaceutical industries, NIH has clinical trial networks to support it. Diagnostics are a very different kettle of fish, and here you see it, you don't have one industry supporting PET or supporting the tracers that get utilized in the device, so you don't have that kind of control and coordination of the type of studies that can be done, and I think you can see that how that leads to fits and starts in the types of studies that get done. Also, at NCI, we only within the past two years started a clinical trials network in biomedical imaging.
So we're, you know, starting to think about changing the culture of how these studies get done, but it's a very different type of investigator who don't have control of their patient. These are patients who get sent to them for a test. They are not the primary physicians that see the patients. So there's a lot of complicated issues that go into trying to get these studies done that I think this group needs to take into account as you set up your framework for what you would like to see. It's not just extrapolating from a therapeutic setting and trying to plunk it into the diagnostic setting, because you're dealing with a very different
.00311
environment.
DR. SOX: Well, Sean, my thought is that we ought to talk about the issue of generalization. We've heard a few comments about that and maybe you've heard enough to formulate your own opinion about that.
DR. TUNIS: I mean, it's such a crucial issue, I wonder if I could invite Sam, if you wouldn't mind coming back up, and sort of representing your notion which as I understand it, you said we need to get out of the mind set of condition by condition, that's an anatomical mind set, this is a functional mind set, and I believe that, you know, the framework that we're sort of coming to some consensus about really drives us in an empirical condition by condition approach, and so I'm

not sure that's going to be -- I feel like maybe there can be some constructive engagement between you and the panel as far as coming to some better at least understanding of that difference of view.

DR. GAMBHIR: Yeah. I'd like to reiterate that, you know, as I stated, given enough time, the best way to approach this would be that we take each disease entity, each category, look at the accuracies, criticizing the literature very

.00312

cohesively, doing a meta-analysis, and then moving to a decision on it. That's the ideal world of how you would do this. But practically, as was just stated as well by Ellen, the amount of time that would take is enormous, and it does an injustice to the patients that are in clinical trials now and to the patients that are outside of clinical trials, and also doesn't really do justice to the fundamental biology of what we're discussing.

We don't want a CT scan to prove the difference in its ability to diagnose a broken bone in my left pinky versus my right pinky. Yet in my view of thinking, cancer as we look back, you will look back at molecular mechanisms, and you've got to get away from categorizing them based on the organs in which they originally emanate from. That's not what's underlying molecular biology of cancer. That's a classic way of thinking, it has some bearing because of our false positive notions in the background signal we get. But beyond that, it's not the right mode of thinking here. The mode of thinking is to go away from these kind of categories and to go to a category looking at a molecular abnormality, and FDG is looking at a molecular abnormality in cancer cells, regardless of where the

.00313

cancer cell originated. And if we don't do that, basically we'll never get through -- we'll get through two or three or four indications, but we will never get to the whole battery of other cancers where those same cancer patients don't have the voices to be heard because the incidence of those cancers is low, we can't recruit enough patients to get those

kinds of studies done. So I think we have to back away from this type of approach to a more unified approach looking at the molecular biology.

DR. SOX: Thank you. If I could just respond and then perhaps to start the discussion. For now, my personal reaction is, the devil is in the details. The devil's in the details of test performance, specific location, specific form of cancer, and the devil is in the details about the management options and about the effect of those management options on health outcomes. So maybe with those two points of view, we can start a discussion. Randel?

MS. RICHNER: Well, taking a step back then, clearly how we defined the technology assessment to begin with then was probably inappropriate, based on what you're suggesting, that we need to step back and look at the body as a whole,

.00314

and looking at PET in cancer, rather than each diagnosis. So it's really, you know, we're sort of down this collateral path then, if that's not the path that you think is appropriate.

DR. GAMBHIR: No, that's right.

MS. RICHNER: So how do we do that?

DR. GAMBHIR: Well, I think part of it is what was alluded to by one of the panel members. If you go back to the grid, the grid concept that was actually the HCFA related concept, saying well look, we can't fill in every portion of the grid, there's not enough data. So instead, you look at the pattern of the entire matrix, and you say how many holes are there and are they sufficiently low enough, are enough Xs filled out to make sense to go for broad --

MS. RICHNER: But the fundamental research question is wrong then. From what we have here in front of us, we have no choice but to look at individual indications. So I think what we need to do is step back and say, how would we frame a question that would meet your research needs and give us the answer to one.

DR. GAMBHIR: I think, to just quickly respond to that, I think the questions being asked

are well intentioned. The idea was could you apply a set of rules for a given disease entity, sort of break it down, but then the next part of the question is well, okay, but how do you generalize the answer you get to this across a multitude of diseases? And all I'm saying is, it's okay to go through this process, but now as we try to go through and generalize to all the different disease states in, let's say cancer for starters, we can't use this process of each individual piece, not unless we're willing to wait 30 years.

DR. SOX: An alternative would be to fill in some of the big holes, see what direction it's going, and then apply that same reasoning to the less common cancers that are going to be very difficult to study. I guess I would argue that this MCAC panel has only filled in a couple of holes so far, and actually we haven't talked about the second application. We have to keep remembering, we just finished talking about one. So, more responses? Ellen, and we'll go this way.

DR. FEIGAL: Yeah. I'll try to keep it brief. My only comment is, maybe a hybrid approach is to look at where there is the most complete information in some particular diseases, which I think was the attempt at this meeting. And then to sort of look at it, not a tree approach, but a forest approach, with the results across the broad spectrum of cancer, and see whether or not is there consistency of results. Look at the trend, rather than look at the precise estimates of the magnitude of the difference, look to see if there's a consistency. And there may be instances in which there is insufficient data but that doesn't mean it's going in the other direction, it just means there is not a lot of data. So I think you may be able to take a hybrid approach with looking at some common diseases or some good diseases in which there's a lot of data and other diseases in which there is a smattering, and just try to look at the consistency of results across a variety of investigators and a

variety of different conditions.
DR. GAMBHIR: No, I think I would agree with that.
DR. SOX: Linda, and then Mike.
MS. BERGTHOLD: Well, this probably complicates things, but it seems to me there are also some other issues that have to do with the treatability and the aggressiveness of some kinds of cancers. And for example, I thought that some of the data about I think it was pancreatic cancer was very
.00317
interesting but -- or was it ovarian -- one of the two of them is very difficult to detect in early stages. So do you put a lot of resources into your sort of diagnostic phase or do you say we don't really know enough, we can't catch it early enough, so we have to, so maybe it's not worth putting the resources in. But these are policy questions, and really important ones for HCFA, to put some kind of framework of policy and priority onto this. If we can't do every disease, technically, can we make some priorities about the, you know, the importance, the treatability, the whatever, other priorities.
DR. SOX: Mike?
DR. MAVES: I brought up the matrix and I think it's actually a good way to look at that, and I agree with what you're saying. I think that at a certain point you may well be able to make some broad categorizations across disease entities. But, having said that, as you know, cancer has different anatomy, different histology, different responses to treatment, and as we saw even in this one example, some very different implications, or thoughts at least, about what that means in terms of health outcome. You know, is it absolutely going to mean that you're not going to do the biopsy? Well, that
.00318
doesn't appear to be the case. It's a relative thing and in fact, it may be more complementary. So, I think you've got it.
I will also tell you that Alan before he left, sort of gave me his proxy to say he didn't think we could lump the indications together. If Bob

was here, Bob would probably say something like, there's plenty of patients, you've got the most notable cancer centers in the world with these machines, it would certainly seem that there is the opportunity to collect this type of information and to bring it forward, and I think to fill out the forest a little bit, so we have a little better assurety about making those kind of decisions.

DR. GAMBHIR: Yeah. I think the only thing I would add to that is to fill in this forest, if you back up five years to the lung cancer data, if I were to show you stuff we presented five years ago, you would look at that and say yeah, it looks like you're in the right direction but you need more analysis, more data, keep doing what you're doing. So the vicious cycle here, though, is to keep doing what we're doing because as was pointed out, this is not big dollars by drug companies pushing these clinical trials. We need the reimbursement because

.00319

in fact what's driven up all the lung, the colorectal, and the data that we've shown in the more established cases is the fact that reimbursement has allowed those studies to get done. So what I'm arguing is, leave to it to clinical judgment perhaps in these less incident cancers, let those be gauged at the clinical level, don't dictate that you've got to study each one in this way. We've got the bulk of proof in the cancers that are more prevalent, and I think reasonably good proof.

DR. MAVES: Hal, if I could just -- I agree with you and in fact, the situation, the conundrum you find yourself in is not dissimilar than we have had in this room before with other panels discussing other types of technology. And in fact one of the things we have done is turn to Sean and turn to HCFA and say wait a minute, these folks do have a problem, we understand how things get funded and things get going forward, and I don't think that's an illogical conclusion for the panel to recommend to HCFA that perhaps there be some sort of investigative role for the Agency to play in helping to fund these thing to help get the answers so we can

fill in the forest better.
DR. SOX: Manuel?

.00320

DR. CERQUERIA: I'm in favor of broad indication approval. I guess from the perspective of the payor, the only question I would ask is when you're doing it at academic centers, you have some control, but when you're reimbursing, what happens when the floodgates open, I mean people start doing it indiscriminately. What steps does the PET community recommend to sort of drink responsibly as it were, to avoid inappropriate utilization, and what steps have the professional societies taken to that end?
DR. GAMBHIR: I think those are very important questions. In part they have been addressed by the exact same way in which the current reimbursement mechanisms have been worked out. That is, what's in place is to tightly monitor the current utilization of the reimbursed techniques, to rereview it in a limited period of time yet again. That is the only way to answer this is to in fact look at the usage patterns, look at this abnormal or normal kinds of usage and try to correct them by revisiting it down the road. On the other hand, if you don't open the floodgate as you will initially, there's no way to assess it. That's been the limiting problem. There's been no way to go forward with the data

.00321

because there has been no way to get these studies done.
MS. RICHNER: To get back to what we were supposed to do in terms of the guidelines for diagnostics and testing this out today, I think what this is sort of saying to us is that we need to figure out exactly what we want as the key markers for approving, essentially, a diagnostic intervention, and it's accuracy and it's some sort of clinical utility, and that seems to work with what we've done today. Now the problem is once again with PET. We have this broad indication and opening the floodgates for use, well, isn't that all about medical management, et cetera? I mean, I think it

should be approved for use and the physicians should be able to have their, use their best judgment on how it's used, and through natural use it will eventually select the path it should take.

DR. JOHNSON: I also support the broad application use with it, and I think that looking at cancer from the molecular basis as opposed to the old model of organ basis is -- it requires not only as we've got on part of our number two, (1) breakthrough technology, in some instances it takes a breakthrough mind set, and a rebooting of the computer, and I

.00322

think that should be applauded as visionary. I think as you brought up, some of the aspects of the devil being in the detail, some of the floodgate issues might be worked out by HCFA and that aspect, but one of the questions that the Executive Committee were to look at is can we make that leap with what's been presented, looking at it on the molecular basis to say is there adequate evidence that we can make that jump and to recommend broad coverage, and I would support that.

DR. SOX: Anybody else want to weigh in on this issue of matrix versus generalizations based on molecular mechanisms? Ron?

DR. DAVIS: Well, my comment is twofold. First of all, I don't feel like I know enough about the biology of cancer to intelligently answer that question. But if we do allow generalization, and we cover PET for example, for many other cancers, it gets me into a policy area that I want to just throw out there, and obviously we're not going to talk about it, but I just want to get it out on the table, and some of us raised this before this meeting. And that is that one of the most common causes of all these cancers is cigarette smoking and if we are to start paying for a lot of PET for many

.00323

many cancers and yet Medicare does not pay to help people quit smoking, there is just a looming irony in policy making there that I think HCFA needs to address, especially when we have an evidence based clinical guideline from Department of Health and

Human Services that says that the treatment of tobacco use and dependence as updated in June of this year is very efficacious and very cost effective.

DR. SOX: Dr. Phelps.

DR. PHELPS: Yeah. I'd like to make a comment back to Mike's statement and also Manuel's. You know, where there are a lot of different teachers, Mike, of the biology of cancer cells by the origin or the organ system that do differentiate out the targets therapeutically. In terms of glucose metabolism, that's not the case. So this particular assay is ubiquitous, although there are other features which are specific to the organ system. So that's one of the reasons we moved broadly, because we take the fact that cancer biology has proven it to be a broad feature of neoplastic generation. Manuel's question, we formed an institute for clinical PET where we train over 700 physicians every year. We went into the American Board of Nuclear Medicine, 35 percent of the questions are on

.00324

PET now. We trained 100 people and we'll train 200 people, physicians, next year at UCLA alone. So, we've reached out to the community, and I don't mean to just focus on UCLA because other universities do it, and when they come to UCLA, they get from three to six months of clinical training and when they leave we overread to them, so we help them to help them learn to be able to do that, to read the scans the right way, and to progress over time as their skill increases. And we do special cases for them over time, and we do that internationally. So we're trying to be responsible in the use.

DR. CERQUERIA: No, I think that's good but realistically though, once you approve reimbursement, anybody who's out there who's board certified in nuclear medicine or radiology will be able to basically bill for this test, so we can't guarantee that they will have the training.

DR. PHELPS: You know, we can't control all the world. We can educate, we can intervene, we can criticize, but there is going to be misuse of PET. I don't know of anything in medicine that's not

abused.
MS. RICHNER: Controlling access through reimbursement seems a little naive in a sense
.00325
because, you know -- it's not naive, that's the world that we live in, but there are other ways to control utilization of technology other than through coverage.
DR. HELZLSOUER: May I make a comment? While the process may be ubiquitous, we've heard from you all that the false positive and false negatives will vary by site, so there is a signal of another issue to deal with. And that, and let's get back to the patient. False positives and false negatives can have a devastating impact when we are talking about how it is utilized. So I am concerned about broad coverage and to make that based on one review of colorectal cancer that we've heard today. But I think it's not a matter of controlling how it's used by reimbursement, it's a matter of doing what's right for the person.
And as we also heard, it's not just Medicare that's looking at this, because not everything is covered by the other carriers who have looked at this issue. So I think while you want to be all encompassing, you also have to be protective in the sense that you don't want something -- and some site may have a lot of false positives that makes it unuseful and also damaging.
.00326
DR. GRIFFITH: May I respond to that?
DR. SOX: Please. If you haven't spoken before, would you identify yourself?
DR. GRIFFITH: My name's Landis Griffith. I'm the director of nuclear medicine at Baylor University Medical Center and the medical director of the North Texas PET Institute.
Before I respond, I would like to appeal to the chair that now that we've finally gotten to what we thought we were discussing today, which is broad coverage, I and several of the other of the scheduled public speakers came specifically to talk about broad coverage and extrapolation into the

community, which I think are several key issues, and I would like to appeal that we can be heard.
Now, the answer to the question regarding false positives is that this panel has appeared to take the approach that we are comparing PET to some perfect ideal currently existing practice, and that is far from the case. We've seen the numbers on CT; yet, clinical decisions are made on CTs and MRs every day. Clinical decisions are made on needle biopsies every day, and they are not that good. PET is at least as good.
In terms of false positives, yes, there

.00327

are false positives. Are there false positives on CT, more than there are on PET. So to say that we shouldn't do it because there is the possibility of false positives ignores the limitations of the techniques that are out there. It also ignores the clinical judgment of the surgical oncologist, the medical oncologist and the radiation oncologist to take the data and do something constructive with it. Many times we find lesions that are outside of the field of initial concern but are much more accessible.
If a surgeon can get to a supraclavicular lymph node and prove that there's metastatic disease, if it's necessary to rule out a false positive, that's a heck of a lot easier than trying to open the patient up and get to a presacral scar after the patient has been treated for rectal carcinoma, because the CT can't tell you where the active tumor is. So, a wealth of information on PET used out in the clinical setting, and the false positives are a problem, no doubt, no test is perfect but as we've seen, PET is considerably better than what we're using.
DR. TUNIS: Can I just direct a specific question here? And by the way, we do plan to

.00328

actually come back and offer anyone who signed up to speak a chance to do their testimony, so you see, we still have 40 minutes.
DR. GRIFFITH: Sometime before the flights

leave.
DR. TUNIS: The question for you is, and I'm really querying to understand, so --
DR. SOX: But we are going to quit at 5:30.
DR. TUNIS: So say for, again, I don't know the literature on use of PET in prostate very well, but let's say, you know, no good empirical studies have been done to characterize the false positive or false negative rate for patients with prostate cancer or suspected, you know, spread of prostate cancer, et cetera. How is a, in the absence of that data, how is a clinician supposed to intelligently use a test like that, in the absence of that information?
DR. GRIFFITH: Well, you know what he does, he does what he's supposed to do for every other nuclear medicine test and that is, he or she calls and consults. The panel may not be aware that part of the regulations for all nuclear medicine tests is that every test that's done has to be
.00329
preapproved by the nuclear medicine consultant.
DR. TUNIS: Just to interrupt you, I said nobody knows, not even the nuclear medicine person knows the sensitivity or specificity because no one's done the study, I'm asking in that circumstance, how does anybody use it clinically. That's the question.
DR. GRIFFITH: You have to address that based on extrapolation of the known data. Is the clinical question one of bone metastases from prostate carcinoma? Then my answer, and I think any responsible physician's answer would be, probably a bone CT is more accurate, given the limitations of FDG right now. Is the question, there's a solitary two centimeter retroperitoneal node that you think may be prostate carcinoma but is not, then I think we do have some data in regards to soft tissue in prostate carcinoma that would say that it would be a valid valid test to do in that case.
DR. SOX: Just so everybody understands where we're going, we are going to quit at 5:30. I hope everybody will stay if they can until then. We

need to go back and make sure, and see if we have a consensus on the second application of PET scanning for colorectal cancer, and then we will spend the rest of the time hearing from people who would like
.00330
to speak, giving first preference to those who had signed up, and we will simply divide the time between those who signed up.
So, we are currently operating without a quorum so we are not going to take a vote on the issue that we've just been discussing, but Sean has been listening closing and ultimately it will be up to him and Jeff to decide which way to go. So, if I may --
SPEAKER: Is a quorum six? You have six.
MS. CONRAD: A quorum is seven.
SPEAKER: Seven out of ten?
DR. SOX: The question is, does PET scanning provide useful information about the extent of additional metastatic disease in patients in whom another imaging test shows a resectable metastasis. My take is that on both the issue of the test accuracy and complementarity to other tests, as well as impact on outcome, my take is that the evidence is quite good on both of these. And if anybody wants to register nonagreement with that attempt to characterize what I think I was hearing, speak up. Otherwise, we will take that as an expression of consensus. Everybody agrees? John.
DR. FERGUSON: I thought we already more
.00331
or less did agree on it, but maybe -- I certainly did.
DR. SOX: I think we only did the one for the indurated scar. Okay. So it sounds like you've got your answer on that.
(The chairman and executive secretary conferred off the record.)
DR. SOX: Why don't you name the people who are signed up and have them raise their hand to acknowledge that they still want to present, and then we will divide the time up.
MS. CONRAD: Okay. Norman la France, do

you still wish to --
DR. LAFRANCE: What I have to say will take about 3 seconds. Should I do it now?
MS. CONRAD: Yeah.
DR. LAFRANCE: My name is Norman la France from Brockwood Diagnostics, Princeton, New Jersey. Thank you for the panel's opportunity to present. Given the time and the types of discussions that we have had, you have a hard copy of my presentation, and in all due consideration for the lateness of time, one of the connections I wanted to make around FDG, in fact to complement Dr. Love's presentation around the FDG review was in one of her last slides
.00332
around the FDG cardiac indications was the requirement around FDG viability requiring a perfusion study for optimal evaluation. And the rubidium generator, which is a HCFA approved PET perfusion agent, provides the unique opportunity in a single setting to have FDG be utilized in that situation. If anybody would like a copy of the Power Point presentation that you have a hard copy of, please let me know, or Miss Conrad can certainly let me know, and I can send that to you. Thank you very much.
MS. CONRAD: Sue Halliday.
DR. SOX: Well, I suggest you identify everybody who wants to speak and --
MS. CONRAD: Oh, okay. You want to speak?
(Inaudible response.)
Okay. How about Dr. Hoverman, do you still have something to say?
DR. HOVERMAN: Sure.
MS. CONRAD: Dr. Griffith?
(Inaudible response.)
Dr. Lieberman.
DR. LIEBERMAN: A brief comment.
MS. CONRAD: Dr. Maddahi.
(Inaudible response.)
.00333
Dr. Merhige.
DR. MERHIGE: I will defer my time to Dr. Maddahi, thank you.

DR. TUNIS: How about, can people live with five minutes each? Okay.
(Inaudible response from audience.)
DR. TUNIS: Okay. So there's four. So five to seven minutes each.
MS. CONRAD: This is J. Russel Hoverman.
DR. HOVERMAN: Let me reintroduce myself. I am Russ Hoverman, vice president for managed care for Texas Oncology, which is a 200-physician all oncologist, radiation oncologist, gynecological oncologist, medical oncologist, in Texas. I have responsibility for managing 500,000 lives with various insurers in the Dallas Fort Worth area and in Austin. And we have had access to the North Texas Clinical PET Center since November of 1998. I'll show you some slides about how many studies we've done; we've done over 2,000 by this time. About 36 percent are lung cancer, 17 percent lymphoma, about 5 percent breast, 11 percent are miscellaneous, about 6 percent are referred directly from oncologists. I'm actually going to skip some of the things that I had available.

.00334

Of interest in the information that I have printed out for you, there are recent very good studies, within the last two to three months, regarding the use of PET scan and other diseases. A summary of the PET scan in lung cancer is in your packet. This is the distribution of the referrers -- this is now a community based PET scanner. Two-thirds are from oncologists, pulmonologists, surgeons, internal medicine and other, and this is our distribution of the diseases we see, a little over a third lung, 20 percent lymphoma, melanoma, brain, breast and colon, with small amounts of head and neck, and others.
I want to skip through these and I will get to the take-home message. This was a recent study; this is a summary of the study in the New England Journal at the end of July. PET changed clinical staging in 62 of 102 patients. This is used as the gold standard, everybody received thoracotomy after both CT scan and PET scanning. The way that

the disease was changed, the disease treatment was changed, had to do with mediastinoscopy. This is the way we viewed the treatment pattern for a diagnosis of PET scan and -- diagnosis of lung cancer -- and this is what's happened since we've added PET scan.

.00335

We have found that an additional 10 percent have metastatic disease and are taken out of the thoracotomy surgical cure phase.
The amount of -- the number of patients who needed invasive staging, i.e., mediastinoscopy, has dropped by 50 percent, and go directly to thoracotomy, so that's where 60 out of a hundred patients get changed.
Dr. Coleman earlier referred to monitoring lung cancer. This summarizes that study, or is very similar, in that if you have PET downstaging after chemotherapy, your prognosis is much better. The issue for us is how do we use that. This is a group of -- these are our lower cost physicians, these are our higher cost physicians, looking at 800 cases within Texas Oncology and dividing up into one standard deviation who are lower cost and higher cost. This is the average cycles of chemotherapy in the lower cost and higher cost groups, and the commercial population, which is here. If we look at the number of patients who got secondary and tertiary regimens, secondary is here and the lower cost up here, nearly 50 percent, and the upper cost, 30 percent. Again, this is in Medicare age, and then the red is tertiary, so that we see in our higher

.00336

cost physicians, more secondary and tertiary chemotherapy. This is a survival curves between the two, blue is lower cost, red is higher cost, and there is no difference statistically.
So how do we plan to use prognostic information based on PET scan? This is our algorithm for metastatic lung cancer which 90 to 95 percent of our patients with lung cancer will flow through. We decide on further chemotherapy based on progression of disease or deterioration of performance status.
If we had a better stopping rule, we may be able to

have more patients get better and earlier supportive care.
The same is true if you have chemotherapy and invariably, you will progress. Again, performance status is the key. If we had a better stopping rule here, we may be able to avoid second line chemotherapy. We don't know how much this is, how much this will involve, but I think this is a promising use of PET scan.
Let me just review. This is a patient with breast cancer, with a -- she actually presented for PET scan because of a rising marker and she had a positive scan. The take home message is 29 percent of folks on breast cancer in this study and recent
.00337
studies confirm or support that, will have a change in their management based on the PET scan.
DR. SOX: Could you wrap up in the next 30 seconds or so please?
DR. HOVERMAN: I can. Let me just go to -- this just summarizes the M.D. Anderson experience without PET scan, in which 23 of 35 patients will have metastatic disease demonstrated, even though they have had laparotomy within a year. And I think, again, there's a recent study within the last three weeks in the Journal of Clinical Oncology, again using the gold standard of thoracotomy and esophagectomy, that shows that PET staging was changed in 22 percent, 10 of the 11 patients in whom distant disease was found had T-3 N-1 disease.
And again, just looking at where, when we talk about broad coverage, we look at similarities of uses of PET scan, so that you avoid surgery by better staging in lung, colon, melanoma, possibly lymphoma, very good data now in esophageal, and I think speculative but by inference, high possibility that it's going to be effective in pancreatic cancer.
And if you look at unnecessary therapy --
DR. SOX: I really do ask you to wrap up; that was more than 30 seconds. Other people are
.00338
waiting.
DR. HOVERMAN: I'm sorry, last slide.

Carcinoma of unknown primary, less radiation treatment, less chemo for lung cancer, and possibly less chemo or additional surgery for melanoma and brain. Thank you.
DR. SOX: Thank you.
MS. CONRAD: Dr. Griffith. Dr. Lieberman, next.
DR. GRIFFITH: Landis Griffith again. I have a few comments. First -- well, I'm known to be a straightforward person and that reputation I'm sure will be intact by the end of the day. The first thing I want to do is reinforce the rebuttal of Dr. Valk and Dr. Gambhir regarding the VA and the Tufts reports, and their critique of the methodology. I feel like I should go back and tell my chairman that we should close down the entire radiology department because as a reviewer and an associate editor for several major medical journals, I can tell you that we don't have a single imaging modality whose literature could withstand that type of scrutiny. PET is at least as good in terms of the information that we got.
Now, the second thing I really wanted to
.00339
say is this critique about the disjointed nature of the studies, well PET grew up in a time when imaging money, money for imaging research, was hard to come by, and so most of these studies were funded by intramural research funds, or from money granted off of the clinical departments, like Dr. Coleman's or like ours at Washington University when I was there. And so that very much limits the size of the studies and the multicentrality of the studies that you can do.
Now I want to talk today mostly about the extrapolation of PET in the community setting. A critical question that has been brought up today for any new technology, whether it's surgery, chemotherapy, or an imaging modality is will it work out in the community, not just in the academic center but out in the community. And in our hands and I think everybody else's hands, Dr. Valk's hands and at multiple other sites, the answer is an emphatic yes.

We have been open at our particular site -- I have been involved in PET for 14 years; we've been involved in this community PET center for two years. Dr. Hoverman has given some of our results. Just -- we entered every patient into a clinical database for follow-up analysis. After two
.00340
years, you can well imagine we're only beginning to scratch the surface of that mountain of data. So far, one of the studies we most recently completed was to look at the first 284 studies we did in patients with colorectal cancer. Of those, 139 were done without a rising CEA level. They were done for other indications and because Medicare won't pay, they were either paid for by private insurers or paid for by themselves or for certain indications we did them as freebies. The PET imaging relative to CT, MR and clinical diagnosis, the PET imaging upstaged 47 percent of those patients and downstaged 25 percent. Now that doesn't necessarily mean that there was a huge change in patient management of those 70 patients, 70 percent of the patients, but it does mean that according to our data at least 45 to 55 percent of those patients had a substantive change in their management.
Now, our oncologists have adapted very readily. I'm an imaging physician and apparently during this arduous evaluation process, HCFA and other entities have been skeptical about potential bias on the part of imaging physicians, and to a certain degree that's understandable, but only to a certain degree. I bristle at the implication that our
.00341
motives are somehow less honorable than those of physicians in other specialties trying to advance their own fields. Yes, we're imagers, but first we are physicians. I don't know how many of you ended up doing exactly what you thought you would be doing in your career but I certainly didn't start medical school thinking I was going to be spending my career in a darkened room reading images off a monitor and teaching nuclear physics to residents and fellows. This is the way we take care of patients.

We entered medical school thinking we were going to take care of patients; we take care of patients. We may not do it with a scalpel or a stethoscope, we do it with the tools of our specialty and I cannot, I just cannot overemphasize the frustration that we feel at being forced to deliver suboptimal medical care, at being forced to delay patient diagnosis, being forced to waste health care dollars by performing repetitive CTs, MRIs, CEA scans, oncocyte scans, all those other sorts of things, all the while knowing that in a large number of those patients, PET scans will answer the question more accurately and quicker, earlier in the disease process so that better decisions and more cost effective decisions can be made.

.00342

The situation with PET is similar to what it was with CT, it was initially approved for a few indications and then took several years before people realized the broad applicability. Metabolic imaging with PET has at least as much validity, as we've heard today, for broad application in cancer imaging particularly and in a host of other disease processes as morphologic imaging has.

Now, we have heard that there were 2 million patients that have been studied so far. Radiologists and nuclear medicine doctors did not order those studies. Those studies were ordered by physicians taking care of patients who had no financial stake in ordering the studies. You've seen the PET studies. You know that they didn't order those PET studies instead of a CT or an MR just to look at pretty pictures, because the pictures aren't that pretty. They ordered those studies because PET makes a substantive difference in the way that they manage patients.

So, I really can't overemphasize that I believe HCFA must allow all the physicians who care for these patients from all specialties access to this technique. Broad approval is the way to do it frankly because as we've heard today, the piecemeal

.00343

approval of these applications with bureaucratic

hassles and guideline by guideline is going to take years. Every day there are hundreds or thousands of patients in this country that are going to be denied access, and not just senior citizens. The decision by HCFA is monitored obviously by other payers and they follow those guidelines in a lot of circumstances and so this, the decision that is made by HCFA extrapolates to the general population. Thank you.

DR. SOX: Thank you.

MS. CONRAD: Dr. Lieberman, followed by Dr. Maddahi.

DR. LIEBERMAN: I will just make a few comments. I think the one that I'm going to take home is the tremendous value that HCFA and the Executive Committee's evaluation and how this conference has moved in a direction that I think is beneficial to patients, I think that you have talked about patient advocates and as a surgeon, that's all we are. And we use PET scan and this whole concept of biologic testing to avoid excessive surgery. I think the oncology community is different than most people understand. It is an integrated group of diagnosticians, of imagers,

.00344

radiologists, radiotherapists, medical oncologists, and patients, and you can't be an oncologist without being patient oriented all the time because we never stop in the treatment plan. Whatever we do, we follow the patient we hope for life, or help through death. So oncology is a different field, and I don't think any of us would be surgeons, I'm sorry there aren't other surgeons here, none of us would be surgeons if we had to make every decision by ourselves.

So we rely on the medical oncologists, we rely on the CAT scan, and now we have a capability of biologic imaging to help us in this process. I think that it's a tremendous advantage. I think you can trust the multidisciplinary oncology community that's all over the United States in every hospital to evaluate this test and to use it appropriately. It won't probably be used for prostate cancer. It will

be used for esophageal cancer. I have had young patients who are sent to me to have their esophagus, to do an esophagal gastrectomy, to do a PET scan, and find disease in the supraclavicular node, a biologic test, we can't even feel, but seeing that type of biology occur as a surgeon is an impetus for us to continue.

.00345

Liver, patients who come with liver lesions, I won't do a liver resection or approach a liver case without PET scan potential, whether it's colorectal or what, because without it, we're subjecting a patient either not to the appropriate procedure or to an excessive procedure. So I think this is just a continuum of oncological development and I really applaud you.

DR. SOX: Thanks for the inspirational words as well as the kind words.

MS. CONRAD: Dr. Maddahi, followed by Dr. Merhige.

DR. MADDAHI: My name is Jamshid Maddahi from UCLA and I represent the American Heart Association, American College of Cardiology and the American Society of Nuclear Cardiology on the topic of PET FDG imaging to assess myocardial viability. The reason that this is included on the agenda is that the original submission did include an application request for approval for PET for myocardial viability that, I would like to first address the two criticisms of the Tufts group, and then also demonstrate some additional evidence for the clinical utility of these tests.

First, as to what societies I represent,

.00346

the American Heart Association is the world premier in the field of cardiovascular disease and has 31,000 members and 4.2 million volunteers. The next society is the American College of Cardiology. This college was founded in 1949, has 23,000 members and 38 chapters in 41 states; more than 90 percent of the practicing cardiologists belong to this society. And the American Society of Nuclear Cardiology has 4,430 members, and is dedicated to fostering optimal

delivery of nuclear cardiology services. Now the issue of myocardial viability is basically targeted at the very specific population of patients in the United States with congestive heart failure that are increasing in number year after year, with 4.6 million of these patients currently, 550,000 newly diagnosed cases each year, the five-year mortality of 50 percent, 250,000 deaths each year, and based on HCFA, they say in 1996, $3.6 billion was spent to Medicare beneficiaries for congestive heart failure, so this has a significant cost impact.

Looking at this data, it shows that the majority of patients suffering from congestive heart failure as shown in the last two blocks of the bar graph are patients over the age of 65, equally

.00347

distributed between men and women, and therefore, this issue is of particular importance to Medicare population.

In the original submission, the question of what is a definition of myocardial viability was raised and because of the brevity of the original submission, it was not clarified that the only issue that was addressed there was the issue of whether PET can identify which patient population would benefit after revascularization with respect to improvement of original left ventricular dysfunction. I will address this issue shortly but I would like to show you as much as time permits, some of the evidence that we have gone beyond that specific question. We do have data on other end points that are very clinically relevant.

With respect to improvement or original left ventricular dysfunction, the original submission showed a 90 percent sensitivity and 73 percent specificity in 11 references of 432 patients. The document was criticized for including one abstract and two references from prior to 1993. If you take those out, the numbers remain the same, 89 percent and 73 percent. In fact, if you go back and look at the older literature from 1986 to 1992, the data is

.00348

88 percent and 71 percent, not significantly different.
DR. SOX: Excuse me. I just want to remind you, you only have a couple more minutes, so --
DR. MADDAHI: Sure. With respect to global functional improvement, there is consistent data in the literature that in patients who do have evidence of viability by PET imaging, ejection fraction improves following revascularization, while without evidence of myocardial viability, ejection fraction doesn't change, and also the same is true with improvement of heart failure symptoms. In our own data, 73 percent of the patients with evidence of viable myocardia who were revascularized had evidence of improvement of heart failure symptoms. While they did not benefit from revascularization, if there was no PET evidence of myocardial viability.
The same data applies to an average of four data in the literature, 339 patients with respect to prediction of survival. And the benefit of PET imaging in selecting a subgroup of patients with heart failure who would benefit from revascularization, and here these two points show that in patients with viable myocardium by FDG,

.00349

medical treatment is associated with a very very high risk and that could significantly be reduced by revascularization. However, if there is no evidence of viability, whether the patient is revascularized or undergoes medical treatment, the results are identical and actually no better than the patient not having revascularization.
Let me skip from these slides and perhaps just show you one that shows the influence on patient management and the data published in 1997. It shows that 63 percent of patients who were initially decided to have, prior to PET scanning, to have transplantation, the decision changed to revascularization after the PET scan was obtained. In 44 percent of patients who were destined for medical treatment, revascularization was done after PET imaging, and in 42 percent of patients who were

destined for revascularization before PET imaging, the decision changed to medical treatment. And overall, 71 percent of patients did low ejection fraction, the decision was changed as the result of PET imaging.
I would like to summarize with a few summary slides at the end, that first of all, the two criticisms of Valk and associates regarding the
.00350
utility of FDG for assessing myocardial viability are addressed in my written document to the committee as well as this presentation, that the exclusion of the 3 of 11 references did not change the results that were originally submitted, and the results have been consistent from prior to 1993 and after 1993, and the definition of viable myocardium as a reference standard of functional improvement if further clarified.
It is important to recognize that the new data that I have submitted in my written document as well as this very brief presentation, that in patients with left ventricular dysfunction, PET FDG imaging predicts post-revascularization improvement in original dysfunction, improvement of ejection fraction, heart failure symptoms, and survival. These are the very relevant, clinically relevant end points for a cardiologist, and influence patient management and is cost effective. I didn't get a chance to show this data but it is given to you in your handouts.
PET imaging is widely accepted by the cardiology community as the gold standard for myocardial viability. Based on this and the 1995 radionuclide imaging guidelines of the American Heart
.00351
Association and American College of Cardiology, which was also approved by the American Society of Nuclear Cardiology, PET imaging with FDG was a Class I recommendation for the assessment of myocardial viability in patients with left ventricular dysfunction in planning revascularization. And currently, the dilemma that we have is third party payers other than Medicare approve the vast majority

of cardiac PET procedures; however, Medicare patients are always turned down and they have to either pay out of pocket or be denied the service, and at this point, Medicare patients overall are being denied of a service that other insurance recognizes to be valuable.

This is my last slide. The conclusion again representing the three societies, the American Heart Association, American College of Cardiology, and the American Society of Nuclear Cardiology, strongly urge the Medicare Coverage Advisory Committee to make a favorable recommendation in support of reimbursement for cardiac PET FDG imaging procedures. Thank you.

DR. SOX: Thank you very much.

MS. CONRAD: Dr. Merhige.

DR. MERHIGE: I've already given my time

.00352

to Dr. Maddahi.

MS. CONRAD: Oh, okay.

DR. SOX: Sean, is there anything else you want from us before we disperse?

DR. TUNIS: No. I want to thank everybody for their input, I want to thank the Executive Committee for their thoughtful discussions. I think it has been tremendously helpful, and that's just thanks to you all.

DR. SOX: I would like to thank everybody who worked hard to make good presentations today. I would like to thank the panel for giving up a lot of time and taking this assignment very seriously. I think it's through experiences like this that we grow together and become more effective, and at this point we're adjourned.

MS. CONRAD: Wait. Can I have a motion that this meeting be adjourned?

DR. MAVES: Motion to adjourn.

DR. HELZSOUER: Second.

MS. CONRAD: Thank you.

(The Executive Committee meeting adjourned at 5:25 p.m.)