NATIONAL QUALITY FORUM

+ + + + +

CANCER ENDORSEMENT MAINTENANCE
STEERING COMMITTEE

+ + + + +

TUESDAY
MARCH 13, 2012

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:30 a.m., Stephen Lutz, Chair, presiding.

PRESENT:

STEPHEN LUTZ, MD, Chair
JOSEPH ALVARNAS, MD, City of Hope
EDUARDO BRUERA, MD, FAAHPM, University of Texas, Anderson Cancer Center
ELAINE CHOTTINER, MD, University of Michigan Medical Center
HEIDI DONOVAN, PhD, RN, University of Pittsburgh School of Nursing
KAREN FIELDS, MD, Moffitt Cancer Center
JOHN GORE, MD, MS, University of Washington School of Medicine
ELIZABETH HAMMOND, MD, Intermountain Healthcare
BRYAN LOY, MD, MBA, Humana Inc.
JENNIFER MALIN, MD, PhD, WellPoint
LAWRENCE MARKS, MD, FASTRO, University of North Carolina School of Medicine
ROBERT MILLER, MD, FACP, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins
DAVID PFISTER, MD, Memorial Sloan-Kettering Cancer Center

ROCCO RICCIARDI, MD, MPH, Lahey Clinic Medical Center
PATRICK ROSS, MD, PhD, Ohio State University Comprehensive Cancer Center
NICOLE TAPAY, JD, National Coalition for Cancer Survivorship
WENDY TENZYK, Public Employees= Retirement Association of Colorado

MEASURE DEVELOPERS:

MICHAEL COHEN, MD, College of American Pathologists
KERI CHRISTENSEN, MS, American Medical Association
AMARIS CRAWFORD, American Medical Association
NADINE EADS, American Society of Radiation Oncology
JAMES HAYMAN, MD, American Society of Radiation Oncology
DIEDRA JOSEPH, MPH, American Medical Association
KRISTEN McNIFF, MPH, American Society of Clinical Oncology
MARJORIE RALLINS, DPM, American Medical Association
VADIE REESE, Society of Thoracic Surgeons (by teleconference)
FAY SHAMANSKI, PhD, College of American Pathologists
ALISON SHIPPY, MPH, American Academy of Dermatology
MOLLY SIEGEL, American Medical Association
SAMANTHA TIERNEY, MPH, American Medical Association
ANUSHREE VICHARE, American Society of Radiation Oncology
EMILY VOLK, MD, College of American Pathologists
EMILY WILSON, American Society of Radiation Oncology
CAMERON WRIGHT, MD, Society of Thoracic Surgeons (by teleconference)

NQF STAFF:

HELEN BURSTIN, MD, MPH, Senior Vice President, Performance Measures
HEIDI BOSSLEY, MSN, MBA, Vice President, Performance Measures
ANN HAMMERSMITH, JD, General Counsel
EUGENE CUNNINGHAM
ANGELA J. FRANKLIN, JD
ADEELA KHAN
LINDSEY TIGHE, MS

ALSO PRESENT:

KENNETH ADLER, MD, American Society of Hematology
DAWN ALAYON, National Committee for Quality Assurance
MAUREEN DAILEY, American Nurses Association
CHARLES HAMPSEY, Eisai Pharmaceuticals (by teleconference)
TOM MURRAY, American Society of Clinical Oncology
JONATHAN MYLES, MD, Cleveland Clinic (by teleconference)
ARTHUR SOBER, MD, Massachusetts General Hospital (by teleconference)

TABLE OF CONTENTS

TABLE OF CONTENTS

P-R-O-C-E-E-D-I-N-G-S

(8:33 a.m.)

MS. FRANKLIN: Hello, everyone. We're going to go ahead and get started. Welcome to the Cancer Endorsement Maintenance in-person meeting, Steering Committee meeting. And we will start this morning by introducing our Chair, Dr. Stephen Lutz. Thank you.

And we'll go ahead and get started with introductions.

CHAIR LUTZ: Hi, I'm Steve Lutz. And by the way, if they haven't pointed it out to you or you haven't seen it, there is a little speak button there for when you want to speak.

So this is my second time for an NQF meeting. I actually did the palliative care meeting in July so they asked me to be Chair. Please don't confuse that at all with any idea that I know what I'm doing. So I will ask the staff to step in when I am either inaccurate or leading us in the wrong

direction.

I am a practicing radiation oncologist in Findlay, Ohio. I also hold Board certification in hospice and palliative medicine and am happy to be here.

MS. FRANKLIN: Okay. And Ann, you can go ahead with the introductions.

MS. HAMMERSMITH: Good morning, everyone. I'm Ann Hammersmith. I'm NQF's General Counsel. We are going to start this morning by combining introductions with disclosures of interest. If you recall probably several months ago you received a form for us that was a disclosure of interest form where we asked you specific questions and we asked you to disclose anything you thought might be relevant to your service on this committee.

We went through those forms carefully but in the spirit of transparency and openness, we do ask the members who have been seated on the committee to go through an

oral disclosure of your initial meeting. It is not necessary for you to recount your entire CV. I'm sure you are all extremely capable. That's why you are on the Committee. But we don't want you to go through a laundry list of all of your publications, etcetera.

What we do want you to do is identify yourself. Tell us who you are with and then to disclose anything that you think is relevant to your service on this Committee. Just because you disclose something doesn't mean that you have a conflict. It is purely the spirit of disclosure.

We are particularly interested in any relevant consulting that you have done that might be connected with what is before the Committee. We are also interested in any grants or funding that you have gotten for work that might be relevant to what is before the Committee.

Two things that I would like to remind Committee members about. The first is

that you serve as an individual on this Committee. Sometimes we have members who say I'm Jane Doe and I am her representing the American Association of fill-in-the-blank. You sit as an individual. You are on the Committee because you are an expert. We are interested in what you think as an individual. You do not represent the interests of your employer, nor do you represent the interests of anyone who might have nominated you to serve on the Committee.

The last thing I want to remind you of is that someone can have a conflict, a real or apparent conflict and it is not financial. Often I hear people say I have no financial conflict of interest. Financials alone don't tell the whole tale. It is possible for someone to have been very involved in something that was entirely volunteer but it still could be a conflict of interest due to your involvement, even though no money changed hands.

So with that, let's go around the table so that you can introduce yourselves and do your disclosures. We will start with the Chair, Dr. Lutz.

CHAIR LUTZ: So I already introduced myself. I think the only disclosure that I would like to and would need to make is one of the measures, 1822 is a measure that I did not take part in creating but it is based upon a guidelines product that I did. So when we get to that part I will remind you of that and hopefully there won't be any concerns. But that is the only thing I can think of that might be perceived as being a conflict.

MEMBER CHOTTINER: I'm Elaine Chottiner. I'm a hematologist at the University of Michigan. I am on the Committee on Practice of the American Society of Hematology. And I do need to disclose that my previous practice was one of the two that was audited by the AMA for the validity

reliability studies that we are going to discuss today on the hema measures.

MS. BOSSLEY: Heidi Bossley, NQF staff.

MEMBER TENZYK: Hi, I'm Wendy Tenzyk and I'm Director of Insurance for Colorado's Public Employees Retirement Association. And I am here, I think, because I operate a large health plan for retirees. And I believe that the only possible disclosure is that we do have a contract with US Oncology for care management of our folks that are diagnosed and being treated by their physicians.

MEMBER GORE: I'm John Gore. I'm a urologist at University of Washington in Seattle. I have a disclosure. I am an expert panelist for BlueCross BlueShield of America for their blue distinction centers on complex and rare cancers.

MEMBER ALVARNAS: I'm Joe Alvarnas. I'm the Director of Medical Quality

at the City of Hope. I am also a member of the Department of hematology and hematopoietic stem cell transplantation.

With respect to disclosures, I am the Co-chair for the Practice Guidelines of Acute Lymphoblastic Leukemia for the National Comprehensive Cancer Networks.

I have grant funding through the Clinical Trials Network of the National Cancer Institute.

MEMBER FIELDS: I'm Karen Fields. I'm the Medical Director for Strategic Alliances at Moffitt Cancer Center in Tampa, Florida. And I am also a medical oncologist.

The only conflict is that I have served in the past as a member of a subcommittee for NCCN and then I also have worked as a consultant in the past for them working on clinical trials and development of a clinical trials network.

MEMBER LOY: Good morning. I'm Bryan Loy. I work at Humana. And my

disclosures include I am a member of the College of American Pathologists, I am a pathologist by training, and also a member of the American Society of Clinical Pathology and the United States and Canadian Academy of Pathology. And I serve as a committee member, a volunteer Steering Committee Member for the American Cancer Society CEOs and Companies Against Cancer. I've done that for two years.

I serve as a volunteer external counsel for Genentech's Oncology Institute. This will be my second year in that. And I have been appointed, a three-year appointment to the National Business Group on Health for the Cancer Advisory Committee. This is year one of three.

And then I have also served on the Molecular Diagnostics Workgroup for the NCCN for one year.

MEMBER HAMMOND: I'm Elizabeth Hammond. I'm an immunopathologist working at Intermountain Healthcare in Salt Lake City in

the University of Utah School of Medicine.

My conflicts are that I am currently a sitting governor of the College of American Pathologists. I am also a guideline co-chair for a combined activity between the College of American Pathologists and the American Society of Clinical Oncology for Breast Predictive Factors, none of which are on this panel's deliberations today.

I am also a previous tissue bank director and head of pathology for the Radiation Therapy Oncology Group, so I have familiarity with some of the marker studies that we are considering about prostate cancer.

MEMBER RICCIARDI: Good morning. My name is Rocco Ricciardi. I am a colon rectal surgeon at Lahey Clinic in Burlington, Mass. I can't think of any disclosures. Thank you.

MEMBER PFISTER: My name is David Pfister. I'm a medical oncologist at Memorial Sloan Kettering in New York. I'm Chief of the

Head and Neck Oncology Service there and I chair our Measures Committee there.

In terms of potential conflicts, I'm on the Board of the NCCN and I also chair one of the Guideline Panels that has to do with head and neck cancer.

I have also been involved in the ASCO guidelines process, perhaps the most relevant in lung cancer. I'm on the Data and Safety Monitoring Committee I think are relevant to deliberations here. I also do pharmaceutically funded research but again focused on head and neck cancer and thyroid cancer.

MEMBER BRUERA: My name is Eduardo Bruera. I work at the MD Anderson Cancer Center and my area of interest is supportive and palliative care. And my disclosures are that in the past I participated in one of the NCCN panels. I hold R01 funding in three different grants from NIH but none of them are directly related to the results of these

surveys. And I think there is nothing else I need to disclose.

MEMBER MARKS: I'm Larry Marks. I'm a radiation oncologist from UNC Chapel Hill. I'm interested in radiation-induced normal tissue and also lean in healthcare.

I serve on several ASTRO committees, American Society of Therapeutic Radiology Oncology. I serve on several committees there related to general practice guidelines and safety with grant support from the NIH and CDC. We have grant support also from Elekta and hopefully, and from Stevens Medical. And I serve on an advisory Board for Elekta as well.

MEMBER DONOVAN: Hi, my name is Heidi Donovan and from the University of Pittsburgh School of Nursing. I'm affiliated with the Oncology Nursing Society and American Nurse Association.

I have NIH funding from the National Institute of Nursing Research in

Symptom Management but not directly related to any of the questionnaires. And I serve on several committees at the Gynecologic Oncology Group and I have sat on a recent working group with NCI on identifying core symptoms and quality of life domains to be used in clinical trials.

MEMBER MILLER: Hi, good morning. I'm Bob Miller. I'm a medical oncologist at Johns Hopkins specializing in breast cancer. I'm also Chief Medical Information Officer at the Kimmel Cancer Center at Hopkins and I volunteer on several ASCO committees, none related to guideline development or anything relevant here. And I don't believe I have any other relevant disclosures.

MEMBER TAPAY: Good morning. I'm Nicole Tapay with the National Coalition on Cancer Survivorship. I serve as their Senior Director of Policy, heading up their policy in government affairs. I'm an attorney by training.

I volunteer on the Ovarian Cancer National Alliance Public Policy Committee. And NCCS also does receive and have partnerships with pharmaceutical companies and some payers but there is a clear firewall in terms of the positions we take. So I actually don't think that is a conflict. Thank you.

MEMBER MALIN: I'm Jennifer Malin. I'm the Medical Director for Oncology for WellPoint. I volunteer on several ASCO committees that have to do with quality assessment and quality improvement. And in the past, I've had a number of research grants related to measure development and quality assessment but none that I'm leading at this point.

MS. KHAN: Adeela Khan, NQF staff.

MS. TIGHE: Lindsey Tighe, NQF staff.

MS. FRANKLIN: And Angela Franklin, NQF staff.

MS. HAMMERSMITH: Okay. Anyone on

the phone? Any committee member on the phone? Dr. Naierman?

All right. Thank you for making those disclosures. Do you have any questions of me or is there anything that you want to discuss with each other based on the disclosures this morning?

Okay, thank you. Have a good meeting.

MS. FRANKLIN: Thank you, Ann. So with that, I think we will move on to a quick project overview. And just for everyone's information, we have phased this project and we are looking at 27 measures for review in this Phase I and we are going to be reviewing them over this two-day period and they address hematology, melanoma, prostate, lung, oncology, and palliative care.

We have moved to Phase II 21 measures and they will address breast and colorectal cancer.

With that, I will hand it over to

Heidi Bossley, our Vice President.

MS. BOSSLEY: Okay, so I am going to walk through the evaluation criteria again. I know all of you have been through the orientation and then on the workgroups but we wanted to spend some time again today just as you get into this to remind you of what this is.

I also wanted to note that there were a couple of people who were named to the committee that are not active at this point because we did identify some conflicts. One of them, Dr. Stephen Edge will actually be coming back for Phase II and he asked to tell everybody he can't wait to be with you during Phase II.

All right, so I am going to walk through this. Feel free to stop me if you have any questions. Again, most of this should be familiar. We also have a Quick Guide that we have developed. All of you should have copies. If you didn't, we will

get one for you.

But again, we are trying to make sure that you have everything in front of you because a lot of this has a logic in how you assess the individual criteria and is rather specific as we go through each one. So, if you want that -- and we are constantly working on it. So if there is any thoughts on how to make it better, please let us know.

So as you all know, there are four major criteria and the hierarchy and the rationale I am just going to walk through quickly. We are going to go through it more in depth in just a few minutes.

But importance to measure and report is first. That is a must pass. So if we get through a measure and we find it doesn't pass importance, you actually will stop discussion on it at that point and you won't move forward and evaluate the measure on the rest of the criteria.

If it does pass importance, you

will then move on to scientific acceptability, which is dealing with the reliability and the validity of the specifications. That again is a must pass. And there is a logic that we have in the Quick Guide and I will walk through. If it doesn't pass, you stop again. Okay?

Then usability and feasibility are our last two. Usability is currently being updated and actually approved by the Board but will not be implemented until the end of the year. So we are using the current one that looks at has this measure, if it up for maintenance, been in use, what uses has it been -- what programs has it been a part of or whatever. And if they have any information on how useful the providers and it is also the consumers are finding you will look for that as well.

And then feasibility is the measure really should cause as little burden as possible. So collected through your daily

care, etcetera.

Then we will move into talking about harmonization and best in class. I actually don't think you have competing measures but we will go through it. And then if you have related measures, you want to make sure that they are harmonized. But again, I am going to walk through all of that in just a minute.

So you do have in this project new measures and measures that are undergoing maintenance. So they have been endorsed before. You actually have a few flavors of all this.

So you have brand new measures that you will come in and you will look at and assess against the criteria. That should be pretty straightforward. The maintenance measures, you have some that actually were time-limited, which meant they hadn't been tested when they were first submitted but they were in use in a federal program so we allowed

them to come in. They have now come back for maintenance with the testing information. So the testing information is actually new. It has never been reviewed by another committee. But you will assess those the same as any other measure.

And then you do have a time-limited measure where they are coming in it is for use in a federal program now. It has been developed and specified but it has not yet been tested. And for that one, we will talk through exactly what you will do. You actually won't rate scientific acceptability the same way because you don't have reliability and validity information. So it will be more a subset of that. But we will walk you through that when we get there. That's tomorrow.

So for endorsed measures again, these measures should have been out in use, we would hope. Most of them are up for a three-year review. So we would look to see that

they have provided data on the implementation of the measure. Typically we see that in 1b under the opportunity for improvement.

You also should look at for the maintenance measures, if there is any that have a potential for reserve status and I am going through what that is exactly.

Most often when they are maintenance measures, we look to see if their reliability and validity testing has been expanded for quite a few of these measures the first time you are seeing the testing. So that doesn't quite apply here.

Usability, again, is it in actual use or are there plans in the timeline provided for it. And then feasibility, again, have they identified any concerns or issues with implementation or unintended consequences?

So it is a little different looking at a maintenance measure.

So again, we have a generic rating

scale that you will use across each of these, which is the high, moderate, low, and insufficient. High means that there is high confidence or certainty that the criterion is met; moderate being that there is a moderate confidence; and then low, obviously low certainty. And then insufficient I am actually going to walk through a little bit more in a minute.

As Drs. Bruera and Lutz remember from palliative, they were the first group to use the new updated evidence and testing. I think we have ironed out some of the kinks.

So for distinguishing between low ratings and insufficient, low ratings generally mean that they did provide the evidence and they did answer the question but the criterion is not met. So for example, you could get a low rating on say the opportunity for improvement is there but it is actually very low. Where it may vary is the quantity and quality of the evidence. It depends on

when you look at the three together, the quantity, quality, and consistency. In that instance, you may find that a low rating is a little harder to do.

Insufficient evidence means that the evidence does exist and was presented but it is not adequate for a definitive answer. So they answered it but it is not enough for you to be able to make a conclusion on that or the submission is incomplete or deficient. So they just didn't provide enough for you to be able to make an evaluation on that sub-criteria or criteria.

So you will use probably both today and again in the exception of the quantity and the quality, that will depend but as we go through that, we may ask you to distinguish why you rated something low and insufficient, just so that we can capture that in the rationale for the report that goes out for comment but it is kind of the guidance that we give and it should be in your Quick

Guide.

So importance to measure and report. Again, this is a must pass, so they must meet all three sub-criteria. The first one is high impact. So it is a national goal or priority. The data on the numbers of persons affected is high resource or perhaps it is small numbers but the impact within that population is significant. So again this is one that there is a lot of ways to interpret high impact.

Performance gap opportunity for improvement is looking that they have demonstrated that there is considerable variation or there is overall less than optimum performance. You would also like to see data on disparities, if at all possible. And when you get to the reserve status, which I will walk through again, that is probably one area you will spend more time, if it is provided because we are finding often they may have overall high performance but when you

look at the disparities data, you actually do see some variation. And for that reason you might say it is not potential for reserve status.

And then evidence is the quantity, quality, and consistency and I am going to walk through that a little bit more.

Okay, so again the gap information variability and performance, overall poor performance. You will look for disparities. Look at the distribution of the performance scores that are provided. The number in the represented -- I can't even talk today. We are going to skip that word. Again, if they measured this in a small population and the performance is high, that actually may balance. You don't know how the rest of the population or the clinicians or hospitals are doing on that measure. You may want to wait, you know, balance that in your decisions here. Again, looking at disparities and looking at the size of the population at risk versus the

data that they provided.

Reserve status, okay. So I have been mentioning this a little bit. So this is what we are talking about. We are finding with some measures that are under maintenance that the measures meet all the criteria with the exception of the opportunity for improvement. So they have actually demonstrated overall that they are doing a good job but we are finding some committees say if we take that measure away, if we remove endorsement, we don't know the implications of that if everyone kind of stops using it.

So we have created what we call reserved status. And again we will look and see if you have any measures that you want to consider for this. But we would have you walk through all of the criteria and make sure that that measure meets everything. You, in this instance, would vote importance down and then come back and determine whether or not you want to consider it for reserve status because

it met everything with the exception of 1b, opportunity for improvement.

This isn't something we want to use all the time. There are some measures that there may be high performance. But when you look at overall it may be scientifically acceptable but the usability or the feasibility, the efforts of collecting that data, say may not be worthwhile, keeping it as a reserve measure and that is perfectly appropriate if that is what you determine. But again, this is available to you if you decide that you would like to use this today.

So the criteria again, evidence for the measure focus, you want to see strong direct evidence and a link, if it is a process measure, a link to the desired outcome and be as proximal to the outcome as possible.

We actually have some outcome measures that are reserved status as well. So don't limit yourself to just process or structure measures. And then you want

reliability and validity to be high ratings. And we will walk through what that is exactly. And then you want to look at overall how useful the measure is. Is it in use? All of the other pieces. So again you want to find this measure to rank pretty high in everything else, except for the opportunity for improvement.

So looking at the evidence piece, which is 1c, the last part of the importance criterion. So each of you were asked to, if you were in a workgroup, to evaluate this and I think everyone else was as well, if possible, rate the measure based on the evidence submitted. So sometimes we have had this and it has happened in the past where you know of additional evidence and we will discuss that. But for the point of what you are doing today, rate it with that and then we will have a discussion if you know there are additional evidence that should be looked at. We can always ask the developer to go back

and add that in and then you can re-rate the measure based on that. Does that make sense to everyone? Okay.

So the evidence rating scale again is quantity, quality, and consistency. This is for process measures. For outcome measures, you don't need to -- developers are not required to provide the body of evidence. They just need to demonstrate a rationale for why the outcome is important. They may provide the body of evidence and that is perfectly fine and you can rate that if you would like but it is not required for outcome measures.

So the quantity again high is five or more studies and this is in articles or papers for this actual study. Moderate is two to four; low is one; and then insufficient is either there is no evidence or it is only selected studies from a larger body of evidence.

Quality is looking at the certain

year confidence of the estimates of the benefits and harms across all of the studies provided. Okay? So it is high looking at there is randomized controlled trials, direct evidence for the specific measure focus. Moderate is there may not be RCTs or there may be but again, it may be a smaller set of information or it is, again, there are some confounding factors to it.

Low is again I'm not going to go through all of it but RCTs or non-RCTs, again, it doesn't have the precision that you would want to base a measure on perhaps but it is still okay but it is lower. And then insufficient is either again, there is no evidence provided based on what they are actually trying to measure and then also potentially it is also just select number of studies. They didn't include everything.

Consistency is looking at stability. So all of the evidence that they provide is showing the same meaningful

benefits or harms to the patient. Again, I'm not going to go through it but it is high, moderate, and low, or insufficient. So again, they didn't provide enough information for you to assess it.

So we have, and this is in your Quick Guide, there is a decision logic. So if the quantity, quality, and consistency of the evidence are moderate or high, and that passes, it is an automatic pretty much moves right on. Yes, it passed 1c.

If the quantity is low but the quality and the consistency are moderate or high, then you would say yes. If you believe that additional research would not change the conclusion. If you think it might, then you would say no. And then if the quantity is moderate or high but the quality is low, consistency is moderate or high, that might be another one where you determine yes, it passes 1c.

And then again if it is overall

low in consistency and then it ranges between low to high for quantity and quality, that would be no. Anything that is insufficient would not meet 1c.

So there are exceptions. This isn't something that we see used often but it can be used. And this was outlined in our testing task force. And I think I have another -- Let me skip forward, yes.

So expert opinion -- let me go back for a second. Sorry. So the exception is the empirical body of evidence for health outcome because this does vary either by outcome or other types of measures. So for outcomes, they need to provide a rationale to support the relationship of the outcome to at least one structure process intervention or service. So they need to demonstrate how that outcome would impact that.

I think you have one outcome measure before you may have a couple others. But this would be where you would look at

that.

For other types of measures, if there is no empirical evidence, expert opinion must be systematically assessed with agreement that the benefits of the patients greatly outweigh the potential harms. So that is what you want to see there. If you do see that when you rate the quantity, quality, consistency lower, if you want to take a look at whether there should be an exception to the evidence, you want to make sure that there is indeed some systematic assessment of what is provided.

Okay. So based on the evidence task force which met roughly a year and a half ago, they really took a deep dive and that is part of what you see with the new evidence criterion in front of you. They felt strongly that expert opinion is not empirical evidence and should only be considered in exceptional circumstances. So the conditions would be if there is no evidence available so it does not

exist rather than not submitted. So it is just one of those areas would be difficult to have evidence in.

I think a good example is often people say we can't do a randomized trial where we would separate out patients but not assess or treat in some way and that would be perhaps an instance where you would look for this.

Again, expert opinion should be systematically assessed and you need to have a strong rationale for why that specific structure or process should be the focus of the performance measurement. And again, that is where the closer that process is to the outcome, the better. So that would be part of your thinking as well, I would think.

Some additional considerations for exceptions. The impact in the opportunity for improvement must be met. So again, the measure still needs to pass the importance criterion in every other way. There needs to

be the strong rationale that I talked about with a link to the desired outcome. You want to look for the proximity. You want the measure that are closer to the outcome, rather than further away.

If there is a measure that is more proximal in process or an intermediate outcome that is before you, you may not want to put forward a measure that is further away from the outcome. Does that make sense?

And then it is important and this is something that our Consensus Standards Approval Committee often looks at and provides guidance on. It is important to distinguish between something that is important to do in clinical practice and things that are important to be putting resources toward for a national performance measure. So again, you want to balance that when you take a look at these measures.

So, any questions on importance?

Okay, scientific acceptability.

So this one is looking at again a must pass reliability and validity. Within reliability, that is precise specification and also that they have demonstrated either at the data element or the measure score level reliability of that measure.

Validity is looking at whether the specifications are consistent with the evidence. Again, validity testing can be at the data element or the score level. There should be a justification for exclusion and also they should show how those exclusions relate to the evidence.

If it is risk-adjusted, we will walk you through. You should take a look at that. And then identification of differences in performance for new measures that may not be something they can yet provide. But for maintenance measures, they should be able to begin to tell you how they think those measures perform and distinguish.

And then also if the measure is

specified for multiple data sources, have they demonstrated that you can compare across those data sources. That is often very hard for developers to do, especially the first time and then the second time maintenance is sometimes it is a bit challenging because it is a bit of work.

So, for reliability and validity you are going to rate these together in some ways. You are going to look at it for a high rating. For reliability it needs to be precisely specified and also the reliability data needs to be provided both at the data element and the measure score level. I'm not sure that they you have any measures here today that meet that but again that would be how we would ask you to rate it. If you see that, it would be the same thing for validity. So specifications are consistent with the evidence.

The empirical evidence of validity is provided for both the element and the

score. And the rest of validity were empirically assessed and addressed. This is a very high bar for most developers to meet but over time we would like to see them reach this bar. I'm not sure that you will see any today but that is no surprise, given the amount of work it takes.

So moderate is again precise specifications. That does not change. And for the reliability, you are looking for either the data element at the data element level or at the measure score level. It would be the same thing for validity.

For low, it may be that the specifications are not clear, so they don't perhaps have a definition in there to explain exactly what they are looking for or the coding may not be accurate are two examples that we have seen in the past. And they actually may have demonstrated that the measure is unreliable.

For validity again the

specifications are not consistent with the evidence, or the validity has actually shown it is invalid, or the threats have been assessed and there is clearly concern with the results.

And then insufficient is you couldn't determine what they did do, perhaps the method of testing or the scope of it or, for some reason, or validity threats were not assessed.

So there again is a decision logic associated with this. Validity if it is high, reliability can be moderate, or high and can move forward. If reliability is low in any instance, you don't move forward. It doesn't pass scientific acceptability. If validity is moderate, again the same instance; moderate or high reliability will allow the measure to pass. And then anything that is low validity with any rating of reliability should not move forward as well.

Usability, okay, so this one again

is once you get past scientific acceptability and the measure is passed, you move on to usability. And this is where you will talk about are the results that are provided, assuming it is the measure under maintenance where they have been able to provide it, are they found to be useful for accountability or public reporting? Is it in use? Is there a rationale for the use in that program or for that particular use appropriate or credible?

And then if it is in use for improvement, have they been able to demonstrate that -- I'm sorry. I have completely blanked on this one.

So if it is in use for improvement and if not, what is the plan of progress? So again, you are looking to see if that measure, especially if it is a maintenance that is in use, have they been able to demonstrate improvement in some way.

Feasibility is looking at again can the generate be generated and used during

the care process. So can you collect it in daily care provided to a patient? Do they use electronic sources?

We are hoping to move to an environment where we use electronic health record data but we are not yet there. But if they do provide that, we will walk you through exactly what you should look at for their -- Is it claims versus abstracted? Just overall assess whether or not they have taken a look at the feasibility and looked at unintended consequences.

So comparison to related and competing measures. So if the measures meet the above criteria and there are endorsed or new measures that either look at the same measure focused or the same target population, so not both. So if it is, say, a patient is looking at patients who have the same diagnosis, you want to make sure that they have used the same coding perhaps or the same definitions or logic to get that same

population, that would be what we would consider related measures. And competing would be they actually measure the same thing; the same measure focus and same target population.

We will walk you through if you have these measures. I think you have more related measures. They come from most of them the same developer so I'm not sure that you will have issues or questions with harmonization but we will walk you through that when we get to that point. This is just a nice table that shows how we define related versus competing. It's probably a little easier than the slide.

And then we have a logic. Again, I'm not going to go through this because when we get to that point, we will walk you through it, most likely tomorrow. And that is for competing and more for harmonization.

Competing measures, we would ask you if you do determine that you have measures

that are competing, again, measure the same measure focus and the same target population, can you determine if one is perhaps better than the other? So meets the criteria more than the other measure. And this is something that is often very challenging for committees but we will walk you through this if you get to that point.

For the most part, impact opportunity and evidence we would assume would be the same, other than developers. One developer may have filled it out better than the other but it would come down to the reliability and validity. We find untested measures cannot be considered superior. They haven't yet demonstrated reliability and validity. And there is a preference for measures with the broadest application and those that addressed disparities in care. And you would look for a preference for -- You would most likely rank a measure higher if it is used for public reporting or in widest use,

as well as those that use electronic sources. So these are kind of guidelines of how we would walk you through.

I don't know that I am going to spend much time on competing measures because I really don't think you have any. I have been proven wrong before so we will see if I will be proved wrong.

Related measures, again, if there are some that you identify that measure the same thing, either the same population or the same measure focus but don't do it in the same manner, we would ask you to provide justification on why you think it is okay that there isn't harmonization across those measures.

Again, this is something that is very challenging for developers. We worked with them on this. If you do identify something, if it is something very reasonable, we may be able to ask them to do it during this process. Otherwise, we will have to

figure out we have often had committees say we would like to see it by the next time this goes through in the maintenance review. Again, I don't think you are going to have that here.

So I will stop there and see if there are any questions.

Helen, do you want to introduce yourself?

DR. BURSTIN: Hi, Helen Burstin. I'm the Senior Vice President for Performance Measures. Welcome.

Heidi just did this overview. If you have any questions, we will walk you through this. We also gave you a Quick Guide on your tables which tries to at least -- As you are walking through it we found it helpful just to have something to refer to, particularly since some of our criteria now have a decision tree. We thought it would be helpful for you to actually see the decision tree. So I hope that helps.

MS. TIGHE: And if anyone needs the Quick Guide, just put your hand up and I will grab a couple copies. Okay, thank you.

CHAIR LUTZ: And one other thing we need to do. I think Dr. Ross was able to join us. So, Dr. Ross, good morning. If you could help us by introducing yourself, and then you sort of missed our phase, we also mentioned if we had any potential perceived conflicts of interest just so everyone can think those over.

MEMBER ROSS: My name is Pat Ross. I'm sorry to be a few minutes late and miss the early part.

So I am Chief of Thoracic Surgery at James Cancer Hospital at Ohio State University and have a busy thoracic practice there. I do have two consulting relationships, one with Pinnacle Biologics and one with Intuitive Surgical, the robotics company.

CHAIR LUTZ: Great, thank you.

MS. FRANKLIN: So with that, we will go over -- we will start our review of the measures.

Our consideration of candidate measures starts with the melanoma and hematology measures. And if you refer to your agendas, the first measure we have up is Measure 0561, Melanoma Coordination of Care. And measure developer is AMA-PCPI. And Wendy Tenzyk, I believe, is the person that we had assigned as lead discussant. And she will just tee up the measure for us and then the full Committee will discuss.

Oh, sure. Do we have someone here from the AMA-PCPI who would like to provide some input about this measure before we get started? This is Measure 0561.

Is there someone on the line from AMA-PCPI representing -- Sorry. The Physicians Consortium for Performance Improvement.

MS. JOSEPH: Good morning. Thank

you for the opportunity to present this measure. My name is Diedra Joseph. I am in measure development at the AMA-PCPI and I have Alison Shippy here representing the American Academy of Dermatology and some of my colleagues will be joining us shortly.

So just to kind of introduce the measure, some of this that I say will apply to the other measures as well, so I will just give a brief background.

The American Academy of Dermatology, the AMA-PCPI, and the National Committee for Quality Assurance formed a melanoma workgroup in order to identify and define quality measures for managing and improving outcomes for melanoma patients. The three measures were approved by the PCPI membership in October of 2007 originally.

And the measure that will be reviewed right now, Measure 0561 is supported by a consistent statement that was published by the American College of Physicians, Society

of General Internal Medicine, Society of Hospital Medicine, American Geriatric Society, American College of Emergency Physicians, and the Society of Academic Emergency Medicine.

The measure encourages communication within one month of diagnosis to the physician providing continuing care to patients with the new occurrence of melanoma. And communications between physicians within a timely manner will lead to improved outcomes by closing the loop of continuous care, thereby reducing morbidity and mortality rates due to delays in treatment and/or follow-up care.

I would also just add that the measure has since been tested for reliability, validity, and feasibility. And the measure is also in use in this CMS PQRS system.

Again, thank you for the opportunity to present the measure and we welcome any questions you may have throughout the discussion.

MS. FRANKLIN: Okay, thank you. So with that, we will go ahead and turn to Wendy Tenzyk.

MEMBER TENZYK: Thank you. So first just to review the measure. And I appreciate the fact that I was given this I think the least technical of the measures. My interest was in care coordination and taking care of patients. So I think this measure was of interest to me. It is the percentage of patients with a new occurrence who have a treatment plan documented in their chart. And this to me relates so much to coordinating care and especially the idea, the new ideas that are being talked about, accountable care organizations, and care transitions where the primary care doctor is aware of what treatment the patient has been recommended for the patient.

And would the expectation be now that I talk through each of the criteria?

MS. FRANKLIN: Yes.

MEMBER TENZYK: Okay. So importance to measure and report, it does seem like this was -- within our group we rated this as medium in terms of the impact of it and the performance gap. It is the cases of melanoma are rising. There is a high mortality. And even though the measure has been in existence for a number of years and is being used, there were still 12 percent of charts that didn't have this documented. So we felt that it was, again, even though it was in use, and certainly significant that 88 percent of the charts did have the documentation, there still was opportunity for improvement with 12 percent lacking that.

So within our group as we discussed it, it did pass in terms of importance to measure and report. So then we moved on to the next phase, which was the acceptability of the measure properties.

And as we looked at those, we were divided there, as you could see from the

results of our workgroup. We did feel that it was easy to identify that the plan was in the record but also there was concern that the data reported didn't demonstrate that the measurement tools were reliable. So there was, our group was somewhat split on that in terms of the results.

We did feel that in terms of usability that it was -- we really had a range there. The measure has been used and it was reported to us in all of the documentation provided that it had been used and there were a number of results studied that was reported but yet there was also feeling that there was really a question as to whether the quality was being improved from the fact that the measure was in use and that it was unclear how the results were being used. So that one, as I said, we were split on that.

And then in terms of feasibility, it seemed to be a feasible measure. Again, reported that it was being in use and that all

of the data elements were in electronic health records but also some concern there as to whether they were really easily extracted and reported.

So as you could see from the results of our preliminary assessment, we were also split within our workgroup on whether the criteria were met.

So if you could give me some direction on what next.

MS. FRANKLIN: Sure. Are there other comments from the rest of the Committee? Feel free to comment, workgroup members in particular.

MEMBER FIELDS: I wanted to comment -- a couple. My main concern was I don't think that the literature actually supported that communicating with the primary care physician improved quality or outcomes for patients. Also, the measure is open-ended. It said it can be verbal communication. So we are scanning the charts to look for one

note that says I told the referring physician that the patient has a diagnosis of melanoma.

I was also concerned about the role of a primary care physician in actually treating and monitoring these patients over time. So I didn't know that -- I needed more evidence that this actually contributed to a quality outcome for these patients.

MS. FRANKLIN: Other comments?

MEMBER ROSS: Yes, I agree completely. It is not clear at all that this is -- There are so many measures that we have to consider and so many things we will be requiring people to do. This does not seem to have the high impact that we might look for from this committee.

MEMBER MARKS: Can I ask a question? Does the criteria say the level of detail of the plan? Can the plan be I'm going to talk the other physicians to figure out what the plan is? Does that qualify as a plan?

MEMBER TENZYK: I would say just from the description that we received as we reviewed it, it just references have a treatment plan; document it in the chart that was communicated to the physician. It's pretty open-ended.

MEMBER MARKS: Yes.

MEMBER GORE: I just thought that the evidence base they submit for demonstration of performance gap all relates to in-patient treatment. It basically is care transitions for patients who are hospitalized. It is not really relevant to melanoma patients. So it just seemed that the evidence presented for that wasn't really relevant to the clinical situation the measure corresponds to.

MEMBER FIELDS: I also wanted to add that I thought that it could potentially increase cost. You are inserting another caregiver in a patient with a group of provides that might not have comfort or

expertise evaluating those patients in the long-term. And if you look at the fact that they are already going to be referred to a primary dermatologist for their follow-up, it is adding another layer of care, not that we would want to minimize the fact that communication is important among all the healthcare providers. I just don't think that this contributes to a quality outcome.

MEMBER MILLER: So I'm not sure how to deal with the -- if we are looking at the quantity of studies, this applies to, in our call, I think this applied to virtually every one of our measures was some of the citations were, they would say well in the NCCN guidelines, there were 93 studies and they seemed to use that, the measure developer seemed to use that as the justification for quantity. Looking at this one, let's make sure I get my numbers right here, they talk about a total of 73 studies meeting inclusion criteria but the 73 studies are just about

communication in general, as Dr. Gore said, between hospital base and primary care physicians. I didn't read the 73 studies but there is nothing in the material given to us to show that this specifically applied to this measure with melanoma. So again, I guess let me just put out there early on in this discussion, I am struggling with all of these because every single one of the ones in our workgroup, if I remember correctly every single one of them, seemed to have the same deficiency that there were studies that were cited as the quantity of the evidence, as quantity and quality but they were very general. They weren't specific to the measure.

So like I say, I'm new at this. And so someone help me. What am I supposed to do with that? Because if I just go with what was presented, then none of these pass.

MS. FRANKLIN: All right. Is there anyone perhaps on the phone or -- Okay,

go ahead, Nicole.

MEMBER TAPAY: I actually am not addressing this specific deficiency but I do want to put out there that the Institute of Medicine for quite some time actually, I think since their study lost in transition on cancer survivors that has called for a treatment plan. It is something that we are certainly looking at case studies to try to support at the NCCS, together with our legislative effort.

But you know, I don't know again whether this is an argument pro or con this particular measure but I do want to put out there that there have been experts in the field that have looked at this and have whole heartedly endorsed this type of measurement as an improvement for the cancer survivor.

MEMBER GORE: I would just say in responding to that this seems to be a classic example of something that definitely represents good clinical care but may not

represent something that requires resources for measurement as a performance measure.

MEMBER PFISTER: Because it looks like it was a maintenance measure, was there interval data that would be helpful in terms of informing the discussion to address some of the points that have been brought up or no?

MS. FRANKLIN: That is something I will ask the measure developers to speak to.

MS. JOSEPH: So with respect to the questions that are being asked about data, unfortunately there aren't any published studies that we were able to identify specifically related to melanoma and referral or care coordination, which is why we chose to reference the guideline that focus on patients that were being transitioned from hospital care to ambulatory care. There just isn't a lot of data in this particular area. I was hoping that Dr. Sober would have been able to join us to speak more to that issue but --

DR. SOBER: Well I am here. This

is Dr. Sober. And just to correct a misperception, most of the melanoma care is actually either outpatient office or ambulatory operating room daycare surgery. There is very little melanoma care that is not advanced disease that takes place as an inpatient for any period of time.

So I think there is a potential gap in what happens to the patient in an outpatient office or in an ambulatory care setting and what the primary care doctor knows about. You either have to value a communication from back to the primary care doctors so they know what is going on with their patient or you don't.

MS. FRANKLIN: Jennifer?

MEMBER MALIN: I think as Dr. Gore said the issue of good communication among providers is clearly very important. But I think the challenge I have with this measure is to have an impact on outcomes you want, you know, secondary prevention of future melanoma,

which is a longitudinal process. It is not something communicating right now with this primary care provider over the next three months. And I think with the evidence that patients stay with the same primary care provider or the communication that happens now impacts their long-term prevention and recognition of how to take care of themselves to prevent future melanomas, you know, I would like to see data like that as evidence in support for this type of activity in measuring this activity is going to improve patient outcomes.

MS. FRANKLIN: All right, thank you.

MEMBER PFISTER: Wendy, as far as the physician providing continuing care, just so I have it clear, is it what, the primary care doctors envisioned, the dermatologist, or who is it? Did they specify more precisely?

MEMBER TENZYK: I guess I would say no, I don't see the specification there

more precisely. The measure just seemed basic in terms of a treatment plan being documented, not that the treatment was done or the results of it. And I would echo what Dr. Miller said, we didn't see, because this was a measure that has been in place for at least 2009 or it sounded perhaps like 2007, we didn't see results and that was one of the big gaps that we looked for.

MS. FRANKLIN: Does someone on the line want to speak to that? I thought I heard something. No?

Any other comments?

MS. JOSEPH: So we just wanted to, in response to that question, say that we would be willing to take your suggestion to maybe further define the treatment plan that is documented or make any additional edits to the measure. We would be willing to take that measure back to the workgroup for consideration.

MS. FRANKLIN: Okay, thanks. Go

ahead.

CHAIR LUTZ: I was just saying we didn't get far in-depth to understand if maybe I am asking to have something done that has already been done but usually for patients that have had melanoma, I inform the family doctor but I don't anticipate or necessarily think that they should be the person following that closely.

I usually say that unless someone does skin cancers all the time, I don't think that they should be counted as the person following. In fact, as a radiation oncologist, I never say oh I will follow you for your melanoma or any type of skin cancer. I mean, is it feasible to say that the care will involve someone who does skin cancer regularly? Because that is, in my mind, an important criteria. I mean, it is one thing to have a family doctor follow where their range of knowledge could vary greatly, versus someone who has done this all the time.

MEMBER MARKS: I think the implication is, I'm just reading ahead, it says that the primary care doctor is integrating all aspects of that patient's care. Not to say that they necessarily are providing the care for the melanoma but that the primary care doctor needs to be aware of what is going on for the melanoma because they are caring for their global patient. That is how I interpret it. I don't know if that is how it was intended. I guess the argument they are making is that melanoma and skin disease in general is just so common, that that is why this is special.

You can say that these are great goals. They should be approved for every cancer patient. But they are saying this is special because it is so common, so primary care doctors deal with this a lot apparently.

DR. SOBER: Yes, this is Arthur Sober again. That is the intent. The follow-up of these patients would be through either

dermatologists or medical oncologists with the primary care doctor kept in the loop.

MS. FRANKLIN: If you want to make another comment, could you please put your cards to the side and let me know like so. I thought I saw someone who wanted to talk. No?

DR. BURSTIN: If people feel like they have enough clarity about the evidence question, it sounded like there was still a little bit of confusion. Okay?

Certainly just a simple count of the RCTs is not necessary. I think that we specifically made a quality, quantity and consistency to have that breadth of what is the available evidence. But I think in this instance what is most important is that one of our criterion is also that particularly for process measures, that process measure should be fairly proximal to the outcome. So the process outcome link is especially important here and that is what we would want to see that in some ways the evidence provides for

us. So I think that was sort of an issue many of you were kind of talking about. But I just wanted to put that in more clear terms in terms of evidence.

MEMBER MARKS: A process question. Are we, the Committee, going to vote on each of these? Do we all take this an up or down, approve or disapprove or do we vote on each of these four criteria? We vote on each criteria. Okay. So we should go through this in order and say did we pass number one because as you said before, if we don't pass number one, we can stop, obviating numbers two and three.

DR. BURSTIN: Exactly and that is the plan.

MEMBER MARKS: Okay.

DR. BURSTIN: As soon as you are done with your discussion, you will move on to voting on importance. And in fact if measures don't pass importance now, we stop evaluating the measure.

MS. TIERNEY: Excuse me. Could I just ask something about one of the questions about the use of the measures? I know there was a question -- Sorry, I'm Sam Tierney with the AMA. I know there was a question about the use of the measure and it has been in the PQRS program since 2009 and I'm kind of wondering how data has maybe changed over time.

So currently the only information that has been publicly available about the PQRS program is the most current is from 2009. So we have included that information in the opportunity for improvement section, which you had discussed with the 12 percent gap currently. And unfortunately the information provided for the public just had mean performance rate. So it didn't have variability across providers. So that is the current and best information that is available to us from the PQRS program. We are in discussions with CMS to try to get more recent

data but at this point they haven't provided that to the public or us in general. So I just wanted to speak to that question because I know that had been raised by some of you.

MEMBER LOY: I thought I heard one of the Committee members, the workgroup members comment on the broad base of evidence that was submitted yet I didn't really hear a response in terms of if there were pieces or trials within that body of evidence that you would want to bring to this committee for us to better understand what evidence exists that would support the importance of this measure.

So there really is no published data specifically for melanoma in this particular area. We did find some older studies, one from 1988 and one from 2001 that were kind of looking at the delay in diagnosis or delay in treatment, based on the length of time that it took for the referral to kind of happen. But since the data was so old, we didn't submit that. We can add it, if you

think it would help or better support the measure, that there really is no published data for melanoma specifically.

MEMBER LOY: Is there a broader topic of making sure that the treatment plan thing documented, and I am assuming by one or more oncologist or hopefully in combination with all oncologists participating in the development of that plan, that that documented piece of evidence conveyed back to primary care physicians or other physicians involved with a patient's care results in some sort of improvement of quality.

MS. JOSEPH: Not that we have identified to date. We can conduct another search of the medical literature to try and identify some more information but I don't think we specifically were looking in terms of the treatment plan. We were looking more or like closing the loop for care coordination. So I am happy to do that if that would be helpful.

MEMBER LOY: Either or both. Either the coordination of care I think would be informative.

MS. FRANKLIN: Dr. Fields?

MEMBER FIELDS: I would say that one of the things that would be the most helpful is you refer interchangeably to who is the primary care provider versus the following physician. And I think if you go back and clarify that, I mean, I think the goal is that the patient has continuity of care. And I understand that there may be a role for a primary care provider but sometimes you talk about the continuity of care for the treating physician and then sometimes you talk about the primary care provider being in the loop. And I think there are two different issues. I think the broad topic of should we have more uniform strategies to communicate patients being discharged into a system so everybody is aware of their diagnosis is one topic that is probably not related to melanoma as much as to

general care.

And then the other topic is making sure that patients stay in a system and get adequate continuity of care for their high-risk disease because a melanoma patient becomes a high-risk patient the instant they have melanoma.

And I think that there is other guidelines or measures that we get to that talk about a patient staying in the system and better ways to keep a patient in the system. That recall one that we will talk about next is a much better measure of quality for the patients, rather than making sure that the primary care physician got a copy of a report.

So I just think there is not enough specificity and certainly there is no literature to support in melanoma that this makes a difference.

MS. JOSEPH: And I do think that the original workgroup discussed the language of the measure leaving it at the provider that

would be continuing the patient's care. They left it general because there was a discussion of whether or not it would be specifically the primary care physician or if it would be a medical oncologist or if it would be a dermatologist that would be following incidence. There was kind of a sense that it could go either way. I think that was why the measure was left broader with respect to who would be following the patient in the future.

MEMBER FIELDS: But to Steve's point, the person with the appropriate expertise needs to be following the patient in the system. And so leaving it open-ended like that isn't necessarily a quality measure. And I think the point is trying to get to a few important measures that measure quality and continuity of care for the patients. And my concern is being open-ended, I understand you don't know who is necessarily going to be following these patients but our goal would be that they get followed by the right level of

provider.

MEMBER TAPAY: One more -- This is partly a question for those with the expertise in melanoma. But I think as a general matter, as we look at quality of care for cancer patients generally, they are not necessarily either or because of PCP maybe in coordination with other specialists and there may be a lot of comorbidities involved. And so I would almost understand a little bit.

Also, if you look in rural areas with specialists not available, I mean, you have to really realize what might be available for particular patients in terms of who was going to be able to follow their care not just for the melanoma but more broadly. So is that, I mean, are they necessarily mutually exclusive?

MEMBER FIELDS: I'm not necessarily a melanoma expert but I can tell you as a medical oncologist, I would feel that a well-trained dermatologist which would have

been the person that was diagnosing and treating the patient in the first place would be the appropriate person to follow the patient, regardless of whether you are in a rural setting.

I would venture to say that primary care physicians aren't going to have comfort or training that is appropriate to follow the patients.

MEMBER TAPAY: I'm sorry. I didn't mean to imply that. I just am trying to figure out to the extent to which this is a measure that is actually going to be promoting broader coordination of care in an improved setting for melanoma patients whether tying in the PCP in that factor not necessarily as the following physician but as a general matter, someone who is following the patient would be useful to do. I'm not disputing your point.

DR. SOBER: This is Arthur Sober again. Just to frame what takes place for melanoma patients up here in Massachusetts,

you know, we are sitting in a tertiary care center, so most of these patients with melanoma are actually initially diagnosed by a dermatologist and then the patient is referred in to our center for further dermatologic and medical oncologic care. And then we will follow the patient here or we will follow them jointly with the dermatologist in the community or we will send them back to the dermatologist in the community for that dermatologic element of their care. But when we send a letter back to the dermatologist, we also send a copy of the letter back to the primary care doctor who may know little about what is going on, as many patients self-refer to the outside dermatologist without going to their primary care doctor first.

MS. FRANKLIN: Go ahead.

MEMBER MARKS: The statement is made in the paperwork here that the point is to let the primary care doctor know how often, for example, the primary care doctor needs to

make sure the patient goes back to see the dermatologist. So it is to inform the primary care doctor so they know what is required. Not that they provide that care, but they know how to coordinate that care relative to the other care the patient is needing.

MEMBER PFISTER: One comment and one question. The comment is that I think that the precision specification of the responsible physician I think is critical to when you are looking at a process-type measure, looking for how proximal that link will be to the outcome that you are trying to connect to.

So I would say that, and I think the comment about the primary care follow-up is certainly very important but that probably comes with a different proximal timeline in terms of the outcome in the quality applications, as opposed to let's say if you were looking at a specifically dermatologic continuity of care in which you are probably

looking at something that is much more proximal.

The question I have and this may just reflect my own view of the process, is how do we access that original submission? Because I am trying to get through the SharePoint, the share web but I am hitting a hard stop to sort of get to it. And it is kind of helpful to be able to see it.

MS. FRANKLIN: You need the actual measure specs. Is that what you are looking for?

MEMBER PFISTER: Yes.

MS. FRANKLIN: Okay, we have got them on the thumb drive. Okay, we'll pull it down for you.

DR. BURSTIN: While we're waiting we will put the specs on the screen so that you can see.

MS. FRANKLIN: So while we are doing that, are there any other comments about the measure as we are getting ready to put the

specs on the screen for everyone?

MEMBER MILLER: So I just want to go back to rehash what I asked 15 minutes ago. If I am just looking at the under 1c, under the quantity of the body of evidence, I think I heard the measure developer saying that there really isn't any literature that supports the specific, you know, what we are discussing at hand. That having this care coordination in place specifically for melanoma provides some outcome of interest.

So if I am trying to be precise on our grid here it seems, therefore, that the answer is zero, that it is low. I mean, I guess if I am understanding the quantity question, if the literature doesn't exist I think either we say it is insufficient or it is low. It can't be -- well it doesn't exist so we have substituted something else that in general care coordination is a good thing.

And again I am just going to keep saying this because I think this applies to so

many of the things that we have reviewed that yes there isn't a literature that explains this but there is literature that explains something, a general principle.

The question is how granular is the expectation when there is no literature specifically for the measure.

DR. BURSTIN: This is really a judgment call for the committee. So my guess is your assessment of this would be that the rating of that is going to be low.

We do have an exception that we can apply but it is truly intended to be an exception. It is not something we do all the time but really at times the evidence may just not be there. So on your little Quick Guide, you should have it, it specifically does say that there are potential exceptions to an empirical body of evidence when essentially there is no empirical evidence but expert opinion is systematically assessed, and this is important, with agreement that the benefits

to patients greatly outweigh patient harms. This comes up, for example, in some of these coordination areas somewhat, although there is a fair amount of evidence in the broad literature around care coordination. But again, it is intended to be an exception.

So I think the issue would be you would still vote on evidence as you see fit. If you choose to, we could then have you consider whether you want to apply the exception if you think this is important enough to do that.

But again, it is intended to be an exception not really part of the evidence criteria.

CHAIR LUTZ: And I think it points out we are sort of moving headlong toward the voting part. And one of the things I will say from having been through this process once is that it seems that groups streamline themselves so that the discussion becomes smaller and smaller and the voting becomes

more and more important.

You see anyone who has ever done this before is already holding on to their little voting thing because we are saying yes, let's vote. Let's vote. You are exactly right. Unless someone else has something else to say, it is almost time to just get to that voting and give the thumbs up or thumbs down and deal with the implications thereafter.

So is there any other question or clarification anyone needs before we move on to the voting part?

MEMBER PFISTER: Does the exception thing, does that come up as an option on the voting or does the voting get that explicit? Or is it basically just come up high, medium, low, insufficient?

DR. BURSTIN: You would need to vote it down first. And I believe Heidi we have now added a slide. Right?

MS. BOSSLEY: We do have a slide that you will move to, if you choose to. Yes.

CHAIR LUTZ: Okay, anything else before we move on to vote for I guess question one in terms of importance to measure and report for this?

Dr. Pfister, are you okay with where we are? Okay.

MS. TIGHE: Does everyone have a voting control? Okay.

MS. KHAN: Okay, everyone, we are going to vote on importance to measure and report and we are looking at impact first.

So looking at 1a impact, it addresses a specific national goal or priority or the data has demonstrated a high impact aspect of healthcare. So you would press one for high, two for moderate, three for low, and four for insufficient. And you can change your vote. Whatever number you press last, that is the vote that is captured. And there is a little clock that I will start and we should be all set to go. So you can go ahead and start.

MS. TIGHE: And actually the thing that tracks the votes is it is connected to Adeela's computer. So if you want to aim at her.

MS. KHAN: So we are missing one person. If you all just want to enter your vote in one more time.

So we have four for high, seven for moderate, three for low, and three for insufficient.

So we are going to go forward and look at the performance gap. Does the data demonstrate considerable variation or overall less than optimal performance across providers and/or populations groups? Again, it is the same rating scale. One for high, two for moderate, three for low, and four for insufficient evidence.

So we have one for high, ten for moderate, three for low, and three for insufficient.

And again, looking at all three

sub-criteria on high impact, performance gap, and evidence. Looking at evidence is it a health outcome with a rationale or the quantity, quality, and consistency of the body of evidence is moderate or high?

DR. BURSTIN: We're missing 1c. Sorry, that's not right.

MS. KHAN: Oh, there it is. It didn't show up. All right. Okay, you can go ahead and start voting.

Can we do it one more time? We're only at 12. We have to get to 17.

MEMBER MARKS: Can I ask a question? Is it the consistency of the body of the evidence in terms of that there is a problem or that measuring this would lead to a better outcome?

DR. BURSTIN: Evidence for the measure focus.

MEMBER MARKS: Okay.

DR. BURSTIN: So does the measure, as intended, have evidence?

MEMBER MARKS: Have evidence.

MEMBER PFISTER: The low is special circumstances is the exception thing you were referring to or no?

MS. BOSSLEY: No. So you should rate this based on what you have been provided. And then if it is low, then we can discuss whether or not you want to have the exception applied and then we will move you to that slide or insufficient. Yes, if it is insufficient, we can discuss that.

MS. KHAN: So we have one yes, four no, and ten insufficient.

MEMBER MARKS: So are you taking the average of our scores for this? Are you taking the average of our scores to go through flow sheet?

DR. BURSTIN: No because you actually have to pass all three to pass importance to measure and report. So the fact that you have rated that insufficient on that third sub-criteria means it doesn't move

further unless you guys want to choose to invoke that exception.

MEMBER TAPAY: Do we want to invoke the exception?

MS. BOSSLEY: I'm assuming, by silence and a few people looking, I'm assuming you want to at least discuss it and then I think it would be helpful again for you to provide some information as to why you think the exception should be perhaps voted on and then we would do a vote if that is the collective thinking of the group.

MEMBER LOY: The exception pertains to all of the criteria or just this last --

MS. BOSSLEY: Just the evidence.

MEMBER LOY: The evidence.

MS. BOSSLEY: Just the evidence.

DR. BURSTIN: Only the 1c, right.

MS. BOSSLEY: So remember again, you want to make sure that it still meets the impact and the opportunity for improvement.

And then the exception would be that evidence is not there and that is why you ranked it insufficient. But then we need to go back and discuss whether it has been systematically assessed and all of that to make sure that the exception would apply in this instance. Does that make sense?

MEMBER LOY: And perhaps this will get addressed here. I heard two issues though in the discussions and one was that the way the measure has been crafted feels like that there may be some controversy around whether or not the primary care physician should be involved or not. So I don't know in this discussion or this exception process that allows for a rephrasing of that and if not, fine. Thank you.

MEMBER DONOVAN: So I'm hearing that this is a unique patient population with unique care coordination issues that we might want to look at. And for me that presents a possibility for why we would create an

exception when there is not specific evidence to be brought to bear on this measure.

MEMBER MARKS: So can I ask do we believe that that is harder for primary care doctors to care for their patients who are getting treated for melanoma than it is for primary care doctors to care for their patients that are getting treated for breast cancer or colorectal cancer or anything else?

My instinct would be no. I mean, I think melanoma care is probably on average a little easier than the care is for breast cancer or colorectal cancer but I yield to others' opinions or interested to hear other people's opinions on that.

CHAIR LUTZ: I think that is one good point. I think another point someone had mentioned, I forget who, there is sort of a lack of data for some of the other measures and I don't anticipate that if we are going to say, if we were going to use 1c as a stopping point that we are going to have an exception

for multiple numbers of them. So it is hard because this is our first one discussed. But if you can remember and prioritize in your head that there is a measure you think was best of the ones that don't have much data or one or two that you would like to push for an exception. I mean, I don't anticipate we are going to say exception on the first one, exception on the second one, exception on the -- I mean, is there is five or six that have limited data, we might want to prioritize in our heads which ones we think boy that one still is really good even though that one doesn't have data.

MEMBER ALVARNAS: And I guess my concern with respect to making exception for this one is one of the criteria you had early on is that this process measure is proximal to some adverse outcome. And I guess if we had some data that demonstrated that sloppiness or dis-coordination of care led to some concrete adverse outcome that could be quantified at

some level, then I think it would mitigate the lack of data and other aspects of the measure but we seem to be lacking there.

To me, that would be the boot that would push me towards wanting to make an exception either globally or with respect to this measure in particular but I have yet to see those data, unless the measure sponsor can articulate that in some way.

MS. FRANKLIN: Dr. Malin?

MEMBER MALIN: This may also not exactly be what the discussion of unintended consequences is supposed to be about in this forum but one of the kinds of unintended consequences I see is that when we have measures that we put out there for public reporting that don't really directly drive quality improvement, I think then there becomes a complacency among measure developers to try to put forth better measures. And I think this measure has been out there for four years now and we haven't seen much evidence

generated to support that it is making a difference. And I think we may see this on some of the others. I would tend to favor that we encourage development of measures that really can be shown to improve patient outcomes or quality of care.

CHAIR LUTZ: So not to ask a procedural question about procedures but so then if we are discussing whether or not to consider this for an exception, is this something that gets voted on or just discussed or where do we go from there?

DR. BURSTIN: Oftentimes you get a sense of the group. If people want to do that, I haven't heard a groundswell. If you feel it would be useful to have a vote, that's fine.

CHAIR LUTZ: I'll say it. Does anyone want to sort of carry the water for this needs to be considered for an exception or should we move on to the next one?

We're moving on. All right. So

if I'm reading correctly, I think the next one is 0650, melanoma continuity of care -- recall system. Dr. Miller?

MEMBER MILLER: So we will try to do this in less than 47 minutes.

CHAIR LUTZ: Like I said, there is always some streamlining and toward the end you actually have a hard time trying to pay attention long enough to give it the attention it deserves.

DR. BURSTIN: Yes, and the first measure usually takes 90 minutes, so you guys are way ahead.

MEMBER MILLER: Well, I'm going to slow down then.

So briefly this is another melanoma measure. This is actually a structure measure. So this is 0650. This is a measure that looks at whether or not there is a recall system in place for patients with a prior diagnosis of melanoma I believe up to Stage III. There is a recall system in place

to get them back for their annual skin exam. And is there a process as part of that recall measure that if they miss their follow-up appointment, how are they tracked down and otherwise reappointed.

So this was described by the measure developers as a structure measure. And I personally agree with that because I think this is a measure that says is there a mechanism, rather than a process in place. I guess that is maybe just more informational. It doesn't change our voting.

So in terms of going through some of the different parts of this, our workgroup had general agreement that this was important to measure and report because of the prevalence of the diagnosis, the increasing incidence of melanoma and the opportunity for impacting the outcome of these patients by early diagnosis of a new primary melanoma which occurs in up to ten percent of patients. And that a recall system could alleviate

that.

Under the -- Let's see what is next here. I'm sorry. So the performance gap that was identified by PQRI/PQRS data suggested that there was still up to ten percent of circumstances where this was not occurring. Some members on our call felt this was almost a never event, where there should really be close to 100 percent. So even though that may seem modest, I think there was consensus and I would agree that that is still a goal that can be improved upon further.

And then moving on to evidence. This measure is best with the same issues as the previous measure, which is that the studies quoted for the measure do not specifically address the recall system. That under the quantity of studies in the body of evidence, most of these were the articles that supported the NCCN and AAD guidelines but I'm not aware that any of those studies specifically addressed a recall system. So we

can decide individually, I guess, whether that is important or not. The quality of evidence was generally rated moderate.

And I think in terms of the other criteria, the usability and feasibility criteria -- I'm sorry. Let me start with reliability. Our workgroup generally were thumbs up for all of those that the measure was felt to be understandable, acceptable probably for reporting and because it is a structure measure the feasibility is perhaps a little easier to measure.

MS. FRANKLIN: Great. Thank you, Dr. Miller.

At this time, I wanted to pause and see if the measure developer wanted to say something about number 0650. Any comments?

MS. JOSEPH: Yes. So Measure 0650 is supported by clinical practice guidelines published by the American Academy of Dermatology and also the National Comprehensive Cancer Network. The measure

focuses on entering melanoma patients into a recall system at least once within a one-year period and having the structure measure in place at least through the process of melanoma patients being screened and examined at least once a year. And having the examinations on an annual basis will improve outcomes as it will lead to early detection of any signs or symptoms of a relapse and/or systemic spread of melanoma, therefore, potentially reducing morbidity and mortality rates.

And just to quickly speak to the point about the evidence not being directly related to the measure, the AAD and NCC and guidelines do recommend annual screening. And so the recall system was the workgroup's way of trying capture or trying to ensure that that process did take place. Thank you.

MS. FRANKLIN: Thank you.

DR. SOBER: This is Arthur Sober. There is actually a second factor that is not commented on in the information that you have

that actually takes place when you do these annual visits on patients for melanoma follow-up. And that is that this is also a group that is high-risk for basal cell/squamous cell carcinoma and the precursors actinic keratoses. So in addition to finding additional melanomas earlier and potential recurrences earlier, there is a big yield in this group in detecting basal cell/squamous cell and the actinic switch may not affect mortality but certain affects morbidity and being able to treat these other types of skin cancers on an earlier basis.

Also annual recall is also supported by the Australian and New Zealand melanoma guidelines.

MS. FRANKLIN: Thank you. Anyone else on the line? Okay.

Any other comments from the workgroup members who have reviewed this measure?

Okay, the Committee as a whole,

comments on this particular measure?

Wow. Okay. Dr. Lutz?

CHAIR LUTZ: Am I to understand this lack of comments as in you would like to vote or lack of comments you just can't think of anything at this moment? I see people holding their voting buttons. Is that --

MS. FRANKLIN: Dr. Malin?

MEMBER DONOVAN: All right, I have one question. I mean, this seems to be a process measure that taps into an outcome that is pretty easy to measure, which is did people come back on an annual basis. And if that is a quality measure, then people are likely to implement a recall system if that worked.

MS. FRANKLIN: Other comments? Dr. Malin?

MEMBER MALIN: I just couldn't tell from the discussion. Was there evidence provided by the measure developers on the link between structure process and outcomes in this measure?

MS. FRANKLIN: Dr. Miller, did you want to speak to that?

MEMBER MILLER: I'm not hearing that there was.

MS. FRANKLIN: Okay.

MEMBER MILLER: I think I heard that what I guess I didn't glean from the evidence from the documents provided was that some of the references that were used to develop the guidelines specifically spoke to having a recall system in place. So I will take that as new information that is important but I'm not sure. Again, I think that is part of the structure and I don't think anything was said about outcome unless somebody else wants to chime in.

MS. FRANKLIN: Do we have anything from the developer?

MS. JOSEPH: The outcome that would be improved would be the lead to early detection of signs and symptoms of a relapse or the spread of melanoma.

MEMBER MILLER: And is there a study that shows that?

MS. JOSEPH: No, I don't think we have any specific related evidence for melanoma.

MS. FRANKLIN: Okay, Dr. Fields.

MEMBER FIELDS: So just to clarify. I actually like this measure. But I think that the data is that up to a third of the patients have recurrent melanoma. So I think that just the epidemiologic data suggests that there is a high risk for recurrence and I don't know that you would do a randomized trial or have any -- So I think just the body of the literature suggests that this is a high-risk group of patients. That was my interpretation.

MEMBER MARKS: The fact that there is a high risk of recurrence doesn't mean that following them forward necessarily is a positive thing for the patient.

MEMBER FIELDS: No but also the

literature is early stage melanomas have a greater than 90 percent survival compared to the late stage melanoma.

MEMBER MARKS: Yes, so finding a new one, I think that is a useful thing. But screening for recurrence of the prior one that is the point I meant.

MEMBER FIELDS: Absolutely.

MEMBER ROSS: I have a question.

MS. FRANKLIN: Yes, Dr. Ross.

MEMBER ROSS: I don't take care of melanoma patients, other than those that have mets to the lung and my question is, is the 12 month the right number? I mean if we are saying that it is important for them to come back, I don't understand. Because the 12 month says they are seen sometime with 12 months but the follow-up visit might be one month after their initial treatment, which does nothing to detect subsequent recurrence or it might be at 12 months and their recurrence is at six months. So I don't

understand how we can arbitrarily say that 12 months is the appropriate surveillance interval for this disease when it is so clearly wrong for so many other diseases.

MS. FRANKLIN: Dr. Miller?

MEMBER MILLER: Yes, I think that what the way it was constructed was just that that is just the measurement period for this measure. I mean I think somewhere in the original specifications there was a comment that there needs to be lifetime surveillance. But I think we are just measuring. We have to measure something and I guess they picked 12 months as a logical interval of time to say did it occur in this first year after diagnosis.

But I agree with what you are saying. I mean, what does that really tell you.

MEMBER MALIN: Sorry, I didn't understand that. So it is only limited to the first year following diagnosis? It is not a

longitudinal follow-up?

MEMBER MARKS: Excuse me. The denominator in here is any patient with a current or history of melanoma. So presuming a patient with melanoma sort of gets re-put into the system every year to make sure they have a yearly follow-up is how I read it. I don't know if that is how it was intended but that is how I read it.

MS. FRANKLIN: Does the developer have something to add there?

MS. JOSEPH: Actually, that was the intention. That is why the denominator does say current diagnosis of melanoma or history of. And the annual, the guidelines do speak to at least annual screening. So that is why we have screening at least once within each year.

MS. FRANKLIN: Dr. Malin.

MEMBER MALIN: So I just wanted to respond to Dr. Fields. So I mean, you know, for whatever reason I also like this measure

but I guess the question is it seems like if the developers don't present evidence that there is a link that it is hard to say that there is high or moderate evidence when it is not presented. It seems like it is more of an issue of maybe one that we might want to consider an exception for.

MEMBER GORE: I agree. I mean, I think at least I know we are not supposed to compare to other measures but compared with the one we just discussed, you can at least hypothesize the link between the measure and the outcome. You can infer it.

And so this seems like something where we are going to rate the evidence as low but an exception seems very reasonable. And I agree. I like the measure as well.

CHAIR LUTZ: Let's see, is there anything else? Any other discussion or comments before we get to voting?

MEMBER GORE: So just to clarify because I mean you brought up the issue of

expert opinion. Because many of these PCPI measures their validity evidence is a review of a panel that they asked do you think this is good and they all think it is good and that is the evidence. so we are not support to consider that good evidence. Just to clarify.

MS. BOSSLEY: You are talking about the face validity information that they provided?

MEMBER GORE: Yes.

MS. BOSSLEY: Yes.

MEMBER GORE: Which is also sort of importance testing.

MS. BOSSLEY: Yes. Helen, what do you think? We have never yet had a committee take face validity and infer it into the evidence. We do see it as slightly different.

MEMBER GORE: Okay.

MS. BOSSLEY: I can see what you are thinking but we haven't -- it has not been part of the criteria. Does that make sense?

DR. BURSTIN: The exception is

that expert opinion is systematically assessed with agreement. So it is not just a systematic assessment, I think is the question. So there was some reference to Australian guidelines, for other guidelines. That might be something you would look toward but it would be a systematic assessment.

DR. SOBER: Yes, this is Arthur Sober again. I just wanted to reiterate that being seen at least annually is the standard of care in the United States. So I think this measure looks to see that implementation of the standard of care is actually being addressed.

MS. TIERNEY: This is Sam Tierney. If I could just add one comment about the evidence.

So although as Diedra said, there is no evidence specific to patients with melanoma, there has been some literature conducted by the task force on community preventive services from the CDC that looked

at whether client reminders increased screening rates for breast cancer, cervical cancer, and colorectal cancer. So they did show that client reminders do lead to increased screening for those cancers. Now, obviously this is a different cancer but just to provide you with that additional background information about evidence for other cancers. I know that was the question.

MS. FRANKLIN: Dr. Pfister.

MEMBER PFISTER: No, that was just some clarification when we think about the strength of the evidence and it kind of alludes to the comment that was just made is that there have been a few different issues that came up. One had to do with how -- You know, the 12 months, does that make sense? You know, one had to do with do client reminders work.

And it is unclear to me in terms of when we are looking at the evidence for this particular measure what are we looking to

rate it for. Because it would seem to me it would vary based on what we are trying to focus on.

So I would say I guess I would argue that as a structure measure folks are getting the idea that it is going to cause a reminder that what was just said is highly relevant, although it wouldn't necessarily have come up in the discussion of an evidence-base for this melanoma specific measure per se. And it would also have little relevance to the relevance of 12 months as an interval.

And so and I think how we would rate the evidence, I would think, probably at least to the extent it seems to me that the litmus point is between that moderate or higher or less than moderate category where you might come in based on what you are focusing on that the evidence is supposed to apply to.

MS. FRANKLIN: Dr. Miller?

MEMBER MILLER: Well I would just

be cautious though with the idea of client reminders as always a good thing. I mean you could have a primary care practice that sends client reminders out to do annual PSA screening in 80 year olds and we now know, I think we know that that is probably not a great thing. So you know, I am not disagreeing with that. It is just I don't think you can use that to infer anything about this measure because I think there needs to be more specificity. So that is my original objection.

CHAIR LUTZ: Any other discussion before we move on to the vote? All right.

MS. KHAN: So again, we are voting on impact. You can vote one for high, two for moderate, three for low, and four for insufficient. And you can start voting now.

I think we are only one person short. So if you could just press it one more time.

So we have nine for high and eight

for moderate.

And voting on performance gap, the data demonstrated considerable variation, or overall less than optimal performance across providers and/or population groups. So you can start voting.

We have one more person. So we have four for high, 11 for moderate, one for low, and one for insufficient evidence.

And then we are rating evidence. So you can go ahead and start voting.

So we have seven yes, one now, and nine for insufficient evidence.

CHAIR LUTZ: It sounded like, if I understood the conversation, there are some folks who would like to have this considered for an exception, in the event that we are now I guess more insufficient than yeses. Is there anyone who wants to sort of encapsulate and give us that point so that we can work with it?

MEMBER MARKS: Sure. The experts

in this field suggest this should be a never event and it does make logical sense. And patients with one melanoma I think are at high-risk for other melanomas but they don't explicitly show that. So having regular follow-up by someone skilled and looking for melanoma it sounds like a very reasonable thing to do.

DR. SOBER: The data from Australia says if you follow 1,000 melanoma patients for ten years, you will get 61 new primaries. So it is about, in their data, six percent over ten years.

MEMBER FIELDS: Actually, I thought that the data was different when we interpreted this one compared to the previous one, which was the quantity was moderate to high because we know that the patients have recurrences and because of some of the data. The quality was lower but the consistency was moderate to high because we know that early diagnosis leads to improvement in outcomes.

So I checked the box yes because the potential benefits to the patients clearly outweigh the potential harms, if you look at how you could rate those different bodies of literature.

So I didn't know that -- I don't think it is the same thing as measuring whether or not PCPs get their reports qualities affected. I think there is more data. Whether or not quality is high, the quantity and the consistency is high.

MEMBER ALVARNAS: Well and I guess to think about it from a slightly different perspective, if we are thinking about which of these metrics deliver value to the patient, I think with respect to the former that was the first measure that we considered that was dubious. With this one, I think based upon the data including those sighted from the Australian experience, there is real value to be conveyed by doing this intervention, which has been recommended by other expert organizations. I think despite what we might

perceive as a lack of data I think there is still extraordinary value that could be conveyed to the patient by implementing those.

CHAIR LUTZ: Is there anybody that wants to argue against that before we find out how we are supposed to vote on that?

How do we vote on that?

DR. BURSTIN: I just think at this point it is your decision. Do you believe that there is sufficient benefit to patients that you would want to potentially invoke the exception? And just again, from the information we gave you, just to remind you, it must have met 1a and 1b, which it did, the first two sub-criteria. A strong rationale links to the desired outcomes and you have talked that through. And consider the proximity of the desired outcomes.

So distinguish important to do in clinical practice, versus importance for national health performance measures. That

will drive significant gains and quality and outcomes.

So I think at this point this is really intended for the committee to have an opportunity to say it didn't pass but for full transparency, so again you are very early in the consensus process at this point. You will have an opportunity for public comment. You will have an opportunity for others to weigh in as well so we try to have your deliberations be as transparent as possible so others can weigh in as well to see if they would have considered the same way.

CHAIR LUTZ: All right, so here is our vote. Is this an exception that you are good with?

MS. KHAN: So looking at the importance to measure and report a potential exception to empirical evidence 1c. Is there an exceptional and compelling reason that the measure should be considered further? So you are going to press one for yes and two for no.

We have 16 yeses and one no.

CHAIR LUTZ: All right, so then we move on to the rest of the voting, the other parts to the evaluation of this measure.

MS. KHAN: Looking at the scientific acceptability of the measure properties, 2a reliability, you are going to vote high, moderate, low, or insufficient and you can start now.

MS. BOSSLEY: Do you want to -- I think you should have a little conversation about the scientific acceptability first, perhaps.

Did you want to -- I think you were the lead on this one. Did you want to talk a little bit about this?

MEMBER MILLER: I didn't have anything else further to add. I said in my opening remarks that --

MS. BOSSLEY: Oh, okay.

MEMBER MILLER: -- for all what is

left, these three that are left, that there was general consensus in the workgroup call that we didn't have problems with these. The problems were the earlier things.

DR. BURSTIN: The second criterion is about testing. So again, it would be helpful since it is a measure for maintenance if you could also just reflect on the adequacy of the testing.

MS. FRANKLIN: Any comments on the testing from the group, workgroup or --

MEMBER MILLER: Well I guess I will say then just for completeness sake, that it should be easy to tell if this is in an electronic health record system, it should be easy to identify that this is built into an EHR. And if it is done on paper, that likewise it shouldn't be hard to extract those data.

MS. FRANKLIN: Other comments? Okay.

MS. KHAN: Okay, so again

scientific acceptability of the measure properties, looking at 2a reliability. And you can go ahead and start voting right now.

We are still missing one person. There we go. And we have seven high, nine moderate, and one insufficient.

And we are going --

MS. FRANKLIN: Our next discussion will be on --

MS. KHAN: We have one more vote.

MS. FRANKLIN: What's that?

MS. KHAN: We have one more vote on --

MS. FRANKLIN: Usability.

MS. KHAN: -- validity.

MS. FRANKLIN: Oh, I'm sorry.

CHAIR LUTZ: Right, so next we go to validity.

MS. KHAN: Scientific acceptability of the measure properties 2a validity. So you can start voting.

So we have four high, 12 moderate,

and one insufficient.

MS. FRANKLIN: So let's discuss from the workgroup or our discussion lead this reliability. I'm sorry. Usability. Any discussion about this particular piece of the measure before we go on?

CHAIR LUTZ: Actually, I guess I am going to request since we haven't gotten to vote on usability yet and since you said this was in flux, can you remind all of us once again how we should look at usability? Because I get confused about exactly what that means for these measures.

DR. BURSTIN: Sure. And in fact we just, the Board just approved an updated definition which we are not applying yet to usability because it is kind of confusing for folks to understand what it really means.

Essentially we are trying to get at is the measure useful. Will it provide useful information for accountability or quality improvement? And since it is a

maintenance measure, you would actually want to have information on actual use as part of this. So they provided information for you that was part of PQRS and beyond that I don't know other information.

CHAIR LUTZ: Any comments?

MEMBER MILLER: I think the usability issue as I understood it was basically can the end user -- is this something reasonable for public reporting? Can the end user make some sense of these data? It is not something so obscure or something that is so granular that it kind of loses its relevance.

And so I think the workgroup's feeling and my feeling is that this is, as the discussion was going, this was something pretty clear, easy to understand. As we said it seems to be a reasonable connection between an outcome and this structure measure. So I would speak to this meeting those criteria.

DR. BURSTIN: And just one

expansion. We do look at all accountability applications including public reporting but for example pay-for-performance, other uses of the measure be appropriate as well, as well as whether the measure is useful for quality improvement. It is supposed to be both an accountability and a QI.

CHAIR LUTZ: Okay, does anyone need clarification or have any other thoughts they want to share before we vote on usability? All right, let's vote.

MS. KHAN: Looking at usability, 3(a) meaningful, understandable, and useful for public reporting and accountability and 3(b) meaningful, understandable, and useable for quality improvement. So again high, moderate, low, or insufficient. And you can start voting. And we are missing one person.

So we have four for high, 12 for moderate, one for low. So we can move on to feasibility.

CHAIR LUTZ: So feasibility is

next. Can you help us again with just a quick thumbnail, since this is our first time voting on anything, just so we are caught up?

DR. BURSTIN: Sure. So is the information something that you could readily collect without a lot of burden, particularly the EHR action here is helpful.

MEMBER MILLER: So I probably misspoke when I spoke to this earlier. But basically yes, this is something that could be easily seen embedded in an electronic health record or collected on paper.

CHAIR LUTZ: Okay, and anyone need clarification or to comment before we get to the vote? On with the vote.

MEMBER LOY: I would just ask a quick question.

CHAIR LUTZ: Sorry.

MEMBER LOY: Did your committee look at the issue of if you didn't meet the measure in the data, did you have, is there any understanding to be gained from whether or

not it was the target date that was missed or was it a process that was missed to recall and follow-up? We don't get to learn anything from that from the process measure. Is that correct?

MEMBER MILLER: I don't think the workgroup really addressed that.

MEMBER LOY: Okay.

MS. FRANKLIN: Does the measure developer want to speak to that question?

Dr. Loy could you repeat that?

MEMBER LOY: I'll try to make it more succinct. The measure tells us whether or not if you met the measure, then you hit both aspects of the measure. If you didn't meet the measure, we don't know whether or not they didn't document at target date or whether they failed to have a process to follow-up on the patients who did not make an appointment within the specified time frame. But it just seems to me that if you are really looking to drive quality, you would want to know which

aspect of that that you missed.

MS. CHRISTENSEN: That's a great question. We have the data. We didn't analyze it that way but we could analyze it that way. But for the measure specification, it is not specified that way.

CHAIR LUTZ: Okay, anything else before we go on to the vote for feasibility? All right.

MS. KHAN: So voting on feasibility, we are looking at 4(a) the data generated during care, 4(b) electronic sources, 4(c) susceptibility to inaccuracies or unintended consequences are identified, and 4(d) data collection can be implemented. You can start your vote.

We are missing one person. So we have six for high and 11 for moderate.

CHAIR LUTZ: Okay, then I think we just go on to the final vote for overall suitability.

MS. KHAN: Right. So for overall

suitability for endorsement, does the measure meet NQF criteria for endorsement? So you vote one for yes and two for no. And you can start your vote.

So we have 15 yes and two for no. So the measure will pass.

CHAIR LUTZ: All right, the next one is 0562, overutilization of imaging studies in melanoma. And I believe Dr. Laver was the first --

MS. FRANKLIN: Actually, Dr. Miller is going to cover this one for us.

MEMBER MILLER: So let the record reflect I was deputized an hour ago to look at this one. I wasn't the primary reviewer, so bear with me. I was on the call.

So 0562 is another melanoma measure. This is a process measure that looks at the question of, the important clinical question of overuse of imaging studies. So the background is that there are many cancers where perhaps we physicians, we are in love

with our tests and we like to do a lot of diagnostic tests and there is very little evidence that the pre-test probability of finding metastasis, for example, is high enough to justify the expense of the radiation exposure and the use of resources for the test.

So in this measure, since there is a suggestion that patients with lower risk melanoma, Stage 0 through IIC who do not otherwise have signs or symptoms suggesting a systemic spread, that these patients generally would not benefit from diagnostic imaging studies. This is a negative so we are looking that no diagnostic imaging studies were performed inappropriately for these patients.

And I guess we will get to this in the discussion but the denominator to this is all patients with a current diagnosis melanoma Stage 0 through IIC or a history of melanoma of any stage. But the important exclusion is that patients have some comorbid condition or

other medical reason why they need said diagnostic imaging studies. So this is the exclusion.

And that I think was probably the most problematic issue in our phone call was it is very hard to specify these exclusions and I will leave it at that and let the discussion start. But that is my introduction.

CHAIR LUTZ: That is a good introduction and good enough that any more where we have to decide at the last minute who is going to do it, you inherit all of them. We appreciate that.

Anything from the developer?

MS. JOSEPH: Thank you. Measure 0562 is also supported by clinical practice guidelines those that have been published by the American Academy of Dermatology and the National Comprehensive Cancer Network. The measure focuses on the process of identifying signs and/or symptoms prior to ordering

imaging for a melanoma patient. The measure aims to include outcomes, including reduction of radiation exposure and also focuses on cost reduction. Thank you.

DR. SOBER: This is Arthur Sober. I just wanted to add that if you do these studies a false positive rate is about 15 percent. And so that usually leads to either additional testing or repeat testing, which is associated with additional costs, patient anxiety and, in the case of biopsies, especially invasive ones increased potentially morbidity associated with it.

The true positive rate of finding cancers when you do these kinds of screenings is actually less than five percent. So it is a tremendous ratio of false positives to true positives here.

MS. FRANKLIN: Thank you. Dr. Malin?

MEMBER MALIN: I just have some clarifications in terms of how NQF views

overuse measures like this. I mean, is it -- do these measures need to show that they improve quality of care in patient outcomes but have to be linked to quality? Is reducing inefficient resource use a sufficient bar? What is the sort of overall view?

DR. BURSTIN: Yes, it is a great question. So measures that assess inappropriate use are considered an element of quality, essentially.

MEMBER MALIN: Okay.

DR. BURSTIN: So I don't know that there is yet another bar to reach. They are brought on the issues of appropriateness. So appropriateness brings in more than just utilization because it says based on evidence this is not needed. So that is where the quality piece is already kind of built in to an overuse measure, as opposed to just looking at the rate of utilization of a test without that built-in appropriateness.

MS. FRANKLIN: Thank you. Other

questions from the workgroup who discussed the measure? From the Steering Committee as a whole?

MEMBER PFISTER: You know, most measures tend to be underused measures. So on its face, an overuse measure has a certain kind of conceptual feel.

But the one thing that was alluded to earlier that in terms of the exclusion issue here and to what extent this was discussed or data provided that if you, it is amazing how, if you are based on what was written on the rec for the reason to obtain a study, if you have a history of cancer, in the handoffs between the ordering and also getting it done, how that ends up being history of melanoma and you know, car accident, cough, whatever that might have led to the reason to order it often kind of starts to, you get some extinction en route to actually what is written and to what extent you end up in this sort of gray zone when you go to quantitate

this that a lot of stuff gets counted as being done as sort of a pseudo for cancer reasons when it was sort of just suboptimal ordering process reasons.

MS. FRANKLIN: Dr. Malin?

MEMBER MALIN: Could the measure developer or someone in the workgroup clarify how signs and symptoms are captured in the denominator?

MS. FRANKLIN: Does the developer have a response?

MS. JOSEPH: Yes, we actually have -- I don't see it in the form that is posted online but we actually have definitions of signs and symptoms.

MEMBER MALIN: I guess this is -- the denominator is specified using CPT codes. So are the exclusions only symptoms captured by CPT codes or is there some other way that symptoms are captured to exclude people from the denominator?

MS. FRANKLIN: Go ahead.

MS. JOSEPH: So the way the denominator is captured, it would be the signs and symptoms would be captured with a CPT code.

MEMBER MALIN: Just so I am clear, if a physician -- Let's say a patient had a cough but the physician didn't code the CPT code for cough during that encounter, that patient would be included in the denominator.

MEMBER LOY: Just for clarity, we are talking about ICD-9 coding, are we -- aren't we? Not CPT coding.

MS. TIERNEY: The denominator is identified through a combination of codes. So it is an ICD-9 code for history of melanoma or -- I'm sorry I don't have it right in front of me, but history of melanoma or a current diagnosis. The staging criteria, obviously, are not part of ICD-9. So we have developed CPT-II codes to identify that for administrative claims reporting. And so the numerator is reported by a CPT-II code and

then the denominator exception would also be reported by a CPT-II code.

I will say for electronic health record reporting we get more into the granularity of the exception examples. And so we have like hard-coded those using the available terminologies of SNOMED and other coding to be able to capture those from an electronic system.

So hopefully that answers your question.

CHAIR LUTZ: I have, I think, a separate question. In terms of overuse phenomenon, I don't take care of any melanoma patients but I take care of other categories of patients where I would see no reason for someone to order a study. Anyone that that person sees as a physician will oftentimes order a study. So maybe the family doctor says oh my God you have got a melanoma. We are going to get a PET scan, we are going to get a CT scan, we are going to do an MRI of

the brain.

So when we say this is a measure, who does that attach to? In other words, there are going to be physicians who see this patient and it will be listed as your patient got these studies for this stage of melanoma, when it may be someone who would say I would never do that. And I don't know if that applies or if that is an issue but it is just one of the things that comes up in our clinic a lot. There is lots of patient who say you ordered what? It is not our overutilization. It is someone else's.

MEMBER LOY: I direct this to the workgroup as well as the measure developer. But I am hearing that there may be a number of exceptions. And I think Dr. Miller alluded to there may be some other reasons that we might want to order a CT scan. But I am also hearing that there possibly is something that might have been missed in the diagnostic workup initially that someone may have gone

back and required an advanced imaging study like a CT scan.

So is there anything in literature that would say there is an acceptable target of exclusions?

MS. JOSEPH: I'm not aware of any data in the literature that talks about acceptable exclusions. I don't know if Dr. Sober would have anything to add to that.

DR. SOBER: Yes, the exclusions would be things like patients enrolling in clinical trials. You could imagine that there would be adjuvant therapy trials for IIC. So someone enrolling in clinical trials is probably going to have the scans done for staging purposes.

If patients are symptomatic, then by all means the true positives then zip up northward. I think the other indications was that if somebody was ordering a CT scan for some other clearly defined indication that had nothing to do with the melanoma. But part of

this measure is to try to promote the fact that doing these scans is not beneficial to melanoma patient care from Stage 0 through IIC where there is an absence of symptoms or signs.

MEMBER PFISTER: You know, I think if the -- This surveillance question is certainly not unique to melanoma. There is a breast literature, there is a colorectal literature showing that. And so I don't think that is so much that there is unnecessary testing done but I still come back to sort of like the robustness of the measure to inform what we are trying to measure.

And so I think what Dr. Loy might have been alluding to but I will give an example, is that you do your original staging study for a patient with melanoma and I wish we all had negative CAT scans but probably a lot of people in this room have what we call incidentalomas. So you kind of look at it. You kind of make that judgment that it is

probably M0. You treat them as there is a problem. But that is kind of tucked in the back of your head that sort of well I probably want to keep an eye on that even though technically they are staged early stage melanoma. And that will probably prompt another CAT scan, another CAT scan. I would think that this comes up with the thoracic surgery all the time.

And so I am just struggling with how that is going to be captured and coded in a way that you are going to end up really getting at what you are really trying to get at, which is this asymptomatic person, pristine scans, and then you are just kind of, you are just doing gratuitous things that makes me feel better, you know, which I think we would all say. But I am not sure how this is specified when relying on electronic measures to sort this out that you are going to be able to get a handle on, I think scenarios for the all commissions in this room

you see all the time.

CHAIR LUTZ: Dr. Alvarnas, do you have --

MEMBER ALVARNAS: Sure. I guess kind of distilling what they have said is that if we are looking to add value to the patient in this sense, keeping them from having to live out the negative consequences of a false positive test and also avoiding radiation exposure, you know, I think we all see that as the intrinsic value of the measure. The problem is the way the measure is constructed or at least the data I had mentioned are being captured.

You can't discern between that, somebody ordering a test that is in appropriate or somebody whose documentation is just poor. And at the end of the day if you are looking in terms of implementing a discreet quality process improvement distinguishing between someone who has poor documentation and somebody who is ordering

tests that are capricious or ill-advised, I'm not sure that the measure will allow you to discern between those two things unless we have a better way of capturing that perspectively.

I don't know that that undercuts the value of the measure but that would be my concern and the ability to take the data and apply them towards particular process improvements.

MEMBER MILLER: So I was going to address my remarks mostly under the reliability section but since we are talking about it, I will just say it now. Which is that I feel very strongly that the reliability of this measure is very suspect for the reasons everyone is saying. I mean, if you think about this, if you were trying to publish a paper on this and you -- and this goes against all the principles of intent to treat analysis. I mean, if you said I am going to exclude these people because I

decided that this really isn't part of the melanoma, this is a secondary thing as I think it was Dr. Pfister was saying. You know, as a clinician you see someone, okay, they had a Stage I melanoma. You are a primary care doctor. They come in with a cough. You know, you are going to approach that patient differently. You might order CT scan because you know they have that history and that may not be part of the initial staging. I just don't know if you do this of ICD-9 codes how you are ever going to pull that out.

And I just think if this going to be held up as a quality measure that this is going to have some meaning, I just thinks this fails at every level. And I'm speaking to someone as a clinician. I see all the time the pain of ordering scans that lead to pain for myself and for my patients because you are always chasing and have these false positives. I just don't know how you get around the denominator exclusion issue. So I'm having a

lot of trouble with that and I think that falls under reliability, as opposed to the -- number one. But I will speak to that now.

DR. BURSTIN: I do think it might be helpful to have PCPI explain the exceptions are not ICD-9 code based. Please, Sam, because I think there is a little confusion in the room about the way the exceptions are coded.

DR. RALLINS: This is Marjorie Rallins, AMA and I worked with the specifications team. I think the goal for these specifications ultimately is to capture the exceptions in the clinical vocabularies, such as SNOMED which captures things like signs and symptoms. However, if you are using another data source, then you would have to use a combination of codes, I-9 for certain disorders but also a CPT-II code to capture the fact that there is documentation that a symptom is present in the record.

MEMBER MILLER: And if the

beneficial outcome is forcing or encouraging the clinician to think twice about ordering a test and documenting that symptom because Lord knows the radiologists struggle with the patient comes down, the requisition just says cancer but maybe they do have a pain. So the radiologist has a hard time doing their job unless the medical record clearly documents the reason for the symptom. So I think the idea that the documentation would have to be better is not a bad thing. That is a good thing, I think.

MEMBER GORE: So just to question the steward and to clarify. Because we in urology we have to report an overuse measure. And so is this something that you report by explicitly denoting a CPT code?

So like for example the question about systems concerns where someone orders it and you now are penalized, for urology we can denote that as a system based on CPT reporting. Because that I think would obviate

some of the reliability concerns that people have. So is that accurate for this measure as well?

MS. TIERNEY: Yes, that is accurate. So a physician who ordered the imaging for another reason like another comorbidity or because the patient had signs or symptoms, they would be reporting a CPT-II code. And then there would be an expectation that that information is substantiated in the medical record somewhere.

So for a claims system reporting like PQRS, they would just be reporting a CPT-II code that corresponds to the clinical action or whatever is based on the measure specifications.

CHAIR LUTZ: Can I ask a cynical question? Say there is a clinician who owns their own CT scanner, whatever incentive they have to order more scans or even a few of being sued if they miss something, if they just document well and every single patient

that comes in they say have you coughed at any point in the past six months? Yes, I did once about three months ago, this patient absolutely has cough and a diagnosis of melanoma. So if they do that, we smile but I actually know physicians that do this type of thing. So it sounds cynical but it is maybe not.

Do we then never capture the physician who simply just codes everything as an exception? You don't get a chance to go back and say gee, 85 percent of their patients have an exception. That is not captured here, is it?

MS. TIERNEY: So the intent with the exceptions is that they would be reported out separately. So a physician would get a report back related to the performance and then how many exceptions they had. So an usually high exception rate hopefully would trigger maybe some potential concerns, possible gaming.

I will say we have done some research and there has also been some research done in the UK on exception reporting. And generally, the research shows that exceptions occurred generally infrequently and they are usually valid when they are put to clinical judgment as to whether or not those exceptions were appropriate in that circumstance.

CHAIR LUTZ: Karen?

MEMBER FIELDS: I wanted to ask, I guess, the developer and possibly all of us for our interpretation of this measure. Were we -- I guess my interpretation was we are trying to get to not aggressively initial staging of patients with early stage melanoma, which is different than following patients over time. And I think we are sort of blending both of those issues. Because you know, overuse and gaming like you talked about in the follow-up is different than how do you initially stage a patient.

So what was the actual aim of this

measure? I interpreted it as not staging an early stage melanoma with anything but physical exam and pathology.

DR. SOBER: This is Arthur Sober. Your understanding is correct. This was meant not to be using these tests in the initial staging of an asymptomatic melanoma patient but the staging would be clinical in pathologic.

MEMBER FIELDS: Just because I think then two different issues -- We are discussing two different issues, which is then how do we follow the patients is a different topic than how do we diagnose them initially.

And I guess one of the questions is then it says current or ever diagnosis of melanoma. I don't think ever diagnosis. I would think that when that patient is sitting in front of you and they had melanoma three years ago and you are going to get follow-up tests, that is a different medical decision-making process. So I assume we are trying to

get to not staging these patients aggressively initially.

So I think there is literature to support that initial not staging or at least there is inference in that very few of those patients have evidence of metastatic disease at the initial time of their diagnosis, which is different than following the patients in the system.

MS. FRANKLIN: Dr. Miller?

MEMBER MILLER: Well, I think you nailed it because I think that is the sum of my objection is the denominator says the history. It is not just a current diagnosis or initial diagnosis. It says a history of melanoma. And I think that is where the entirety of my objection about how are you really going to separate out the "appropriate studies" versus "inappropriate studies" if it is really any melanoma patient at any point and whether it is a new diagnosis or not. And I just think, in my opinion the measure fails

in that measure.

MEMBER MARKS: Can you clarify that for me? So are saying evaluating it it says at the time of diagnosis for early stage or at the time of follow-up for any stage. Right? And you are saying that we should be doing routine staging?

MEMBER MILLER: No, the opposite. I am saying if you include any melanoma patient in that, then any scan that is ordered for any valid clinical reason is going be counted as a denominator could theoretically be a denominator exception. How do you, for the reasons we were talking about earlier, whether it is coding correctly or gaming the system, how are you going to separate out those, as opposed to -- and I guess I would like to see it -- Let me say it the other way.

I would like to see a measure that says that there are no denominator exceptions, figuring that the denominator exceptions will spread across the entire population. So if

you are looking for quality measure, the doctors that have fewer scans ordered, bottom line are probably the ones that are doing it right. They are the ones that are not ordering inappropriate scans. When you start allowing exceptions, how do you justify the exceptions? It makes it worse by saying it is any history of melanoma, as opposed to just the initial diagnosis.

MEMBER FIELDS: Right. So if the goal was we don't over-stage people initially, then we wrote the measure incorrectly and we should change the measure.

MEMBER MARKS: Well, it's both. It is do we stage, do we over-stage in the diagnosis and do we do too much surveillance in follow-up? They are both combined, is how I read it. That's okay.

MEMBER FIELDS: All right but I would think -- Well are they two different measures then? Because we should stage them appropriately and then we should follow them

up appropriately.

MEMBER MARKS: But they are saying for the early Stage 0 to IIC, there is no reason to stage them beyond clinical exam at the outset. And then for all patients, all asymptomatic patients, there is no reason to scan in the follow-up.

They are both valid, I think.

MEMBER FIELDS: Yes, but I guess my point is because we are having the muddy waters of appropriate diagnostic imaging, the first question is different than the follow-up question because the first question is how do we stage an early stage cancer and what are the appropriate exams?

The second question is following them with surveillance scans is not appropriate either. But you are blending too many variables in the decision-making to really get a measure that is helpful and concise and can improve quality is my point.

I mean I understand they are both

-- I mean because when I am hearing the questions about are we every patient that comes back are we doing to say they have a cough and therefore we are going to get a CAT scan is a separate topic from when a patient walks through your door with a Stage I, early stage melanoma how much work-up should you have?

I am just listening to the discussion and trying to understand what the goal of the measure was supposed to be.

MEMBER MARKS: I guess it is both. They are both overuse concerns and they have lumped them together. And they are cared for by the same group of doctors, presumably, so it sort of makes some sense that they are addressing these patients. I think they make sense to put together. It doesn't bother me so much that they list them both.

MS. FRANKLIN: Could we hear again from the developer on that point on the intent of the measure?

MS. JOSEPH: So as Dr. Sober did state earlier, the intent of the measure is to capture any imaging that is not being appropriate based on the patient being asymptomatic.

So for patients with a new diagnosis of melanoma, if they are at Stage 0 through IIC, then they would only use physical findings or pathological diagnosis versus using imaging to stage those patients, based on the guideline recommendations. And then for a patient that is being followed with a history of melanoma, there is no evidence that suggests that imaging is necessary. So they would be followed by the annual exam or the annual visit to the doctor if they don't have any signs or symptoms and there is no reason, there is nothing justifying imaging in that set of patients. And so instead of having -- so the intent was to capture both of those different populations in the one measure. Initially the measure has been recently

updated. Initially the measure was Stage 0 to IA, but since the evidence changed, the measure had to be updated in order to be consistent with the evidence.

MEMBER PFISTER: I mean I think Dr. Fields presented a key point here because I think a lot of the measurement issues I think are unsolvable for surveillance electronically. As opposed to I think if you are doing like I think it would potentially be great value to efficient staging. And also the efficiency of staging to leverage that behavior initially also avoids the incidentaloma problem on the CAT scan you didn't get initially and appropriately.

You know, when I read the measure, clearly everything is lumped together but I think I would be more sympathetic to this measure if it did just limit it to initial diagnosis because I think that would be a more measurable event. I think lumping it together I think that reliability will take a huge hit

and I just don't see that as fixable.

MEMBER LOY: I supposed. This is Bryan Loy. I suppose that this speaks to the reliability issue.

How did the workgroup deal with the issue of seen for an office visit during the one-year measurement period? What if you didn't show back up or you showed up back in a year and a day? Do you get excluded from the measure? How did you all deal with that aspect of the measure?

MS. TIERNEY: So what you are speaking to kind of relates to identifying patients for the denominator. So in a program like the PQRS program for the claims reporting, they would look for a CPT E&M code that indicated the patient had the visit sometime within the reporting period.

So if a patient didn't have a visit within that year, then they wouldn't be part of the denominator population of the measure.

MEMBER LOY: Okay, so you are neither given credit for or --

MS. TIERNEY: Right.

MEMBER LOY: -- discredited for not showing back -- being lost to follow-up?

MS. TIERNEY: Yes.

CHAIR LUTZ: Can I ask a quick question just in general? I don't of many overutilization measures in any branch of medicine. I don't know if through NQF or through anyone else's experience --

DR. BURSTIN: I think there are dozens.

CHAIR LUTZ: Dozens? Okay. And in those dozens, is there a common discussion about whether there is one time frame either at diagnosis or follow-up or is it more common to have either? I mean, I am a little lost.

DR. BURSTIN: I think it is not so much an issue of how you frame it. It is really just the evidence and I guess that is the question that was raised earlier. If the

evidence is identical that you wouldn't screen for either period of time then lumping seems reasonable. But again, it is really up to you. That is why you are assembled as the clinical experts here. If the evidence is the same, then it is not clear why you would not lump those two together, since you do have the ability to have the clinician provide the exception.

MEMBER GORE: We do have an overuse measure in urology and it is limited to the diagnosis. And so I think at least our experience with doing it clinically it is very reliable. It is very usable. And I think if that were extended to the surveillance period, it would be much more complicated from a usability and reliability standpoint.

CHAIR LUTZ: All right, are we getting anywhere near the ability to start voting on this or do we --

MEMBER MARKS: Are we allowed to vote on it with a friendly amendment to take

out the follow-up patients or it is an all or nothing?

CHAIR LUTZ: My understanding is it is all or nothing.

DR. BURSTIN: I believe it is. I think the question is, I would also like to hear from PCPI if there were any differences in the reliability. They did test this measure. Can you tell us if there were in fact differences in the reliability of the measure when you looked at both patients at initial diagnosis as well as follow-up. I mean, this should be an empirical question rather than just a --

MS. JOSEPH: So, to that point I don't believe -- Okay I don't believe the initial version of the measure included the history -- the patients with a history of melanoma. So that part of the measure has not yet been tested.

DR. BURSTIN: So which measure did you test? The one that has history in it or

the one without? Because technically you should be presenting the measure you have actually tested.

MS. JOSEPH: I'm not sure.

MS. SHIPPY: So, hi. Alison Shippy. I'm from the American Academy of Dermatology. So this is included in the PQRS program and it has for a couple, for 2011 and it is again for this year. So when we gathered the data to send to PCPI to run the testing, we asked -- so essentially the AAD runs a registry for PQRS reporting. So we took the information that practices had entered into the registry system. We then recruited additional practices to give us copies of each one and we sent them to a medical chart abstractor who then kind of duplicated the entry. And then I know that the testing group sent through their analytics.

But we captured both history of and new melanoma patients. So it was tested

for both.

DR. BURSTIN: And was there any difference in the reliability between those two cohorts?

MS. SHIPPY: Right. So I don't think that there was.

MS. CHRISTENSEN: So to clarify in what I was shaking my head on, we did not do the analysis at that level. We could do the analysis at that level. We have the data. We just did not do the analysis that level.

CHAIR LUTZ: Are we now getting closer to voting? I would just like to measure how many people are holding the black things and staring intently. You kind of get an idea of where we are.

MS. KHAN: So importance to measure and report, 1a on impact. You can start voting.

So we have seven high, six moderate, three low, and one insufficient.

So looking at 1b performance gap,

you can start voting.

So we have one high, 12 moderate, two low, and two insufficient evidence.

And looking at 1c for evidence. So we are missing one person. If someone could just click it one more time.

So we have seven yes, three no, and seven insufficient evidence.

CHAIR LUTZ: So do you stop because more than half are either no or insufficient?

DR. BURSTIN: This is where it might be helpful to get a sense of the group. And the developers are certainly welcome to come back and provide additional information. For example on this question that you guys raised about initial presentation versus history of. So it might be helpful just to have the group have a discussion of those who thought it was no or insufficient. Is there anything the developer might be able to come back to in terms of additional information on.

Essentially were people voting on that issue of history versus initial presentation?

So I think for now you are just probably, what do you think, Heidi, stop?

MS. BOSSLEY: I think stop and we will huddle with PCPI and see if there is anything that they may be able to do that we could then bring back to you, unless you all say no. But I think it may be worthwhile seeing if we can pull something together for you and have them respond and then we will bring it back.

CHAIR LUTZ: All right, can I take the most important vote so far this morning? How about a 15-minute break?

(Whereupon, the foregoing matter went off the record at 11:14 a.m. and went back on the record at 11:38 a.m.)

CHAIR LUTZ: All right. Shall we work our way back in? I think 0377 is next. And the question had come up in terms of order since we were a tiny bit late what our plans

are, I think we were hoping to get through the next four, the hematologic ones before lunch. So that it the carrot all the way out at the end that we are chasing. If we can get through these four hematologics, then we are allowed to eat.

And the staff has suggested if we could for 0377, we start with AMA giving us sort of the presentation and then we will look to Dr. Alvarnas after that to comment.

DR. ADLER: My name is Ken Adler. I'm a hematologist in Morristown, New Jersey. I have been a member of ASH for 25 years and I was on the original working group in conjunction with the AMA and other members of our ASH Committee on Practice to try to develop measures back in 2006-2007 that would improve patient outcomes and improve patient care. So I will present the four measures that we have developed from ASH and what has been in practice the past several years.

The first measure is Measure 0377

and this is the use of baseline cytogenetic testing in patients who are newly diagnosed with myelodysplastic syndrome and with acute myelogenous leukemia. The numerator is all patients who have baseline testing done and the denominator is everybody diagnosed with a diagnosis of AML or myelodysplasia.

And we feel that this is important in terms of improving patient outcomes, that it helps stratify patients with myelodysplasia, that it shows what the risks are and the prognosis of patients with myelodysplasia.

And I will open up for discussion. Any comments or questions?

CHAIR LUTZ: I think Dr. Alvarnas was our primary reviewer, if you want to go through your thoughts.

MEMBER ALVARNAS: Great. I appreciate the opportunity to speak towards this measure. It is one of the things that when we sat down in our group over the phone

we viewed this measure as being of importance with some caveats as towards the specificity with which it was articulated.

As Dr. Adler mentioned, the numerator here would be those individuals who are evaluated with baseline cytogenetics testing, the denominator being those with a diagnosis of myelodysplastic syndrome or acute leukemia.

One of the things that had come across in our initial review of this was much of the data and much of the focus appeared to be on myelodysplasia, whereas we viewed a focus on acute leukemia as being of at least equivalent in performance and also to make sure that that referred to both acute myelogenous leukemia and acute lymphoblastic leukemia for which we believed that karyotypic data might provide important stratification means to decide, to make major therapeutic decisions with respect to the patient population.

In terms of the construct that we have here for reviewing this, we felt that this was important to measure and report as a group because it played such an important role in the evaluation and management of these patients.

In terms of the performance gap, this was something that was relatively striking in the 2008 data that were originally cited as part of the impetus for this measure. Nearly 50 percent of patients did not actually have baseline cytogenetic data which we believed would compromise their potential outcomes. There was a partial data point from 2009 where approximately 90 percent of patients may have had that but that assessment was based upon an incomplete dataset. So we have concerns about the reliability of that data point to make major decisions regarding this particular proposed metric. And towards that end, we still look towards the 2008 data as being the most robust dataset upon which to

evaluate this. And we would, in terms of the performance gap, view this as a significant performance gap with a high potential to affect patient outcomes.

In terms of the evidence, broadly the citations referred to the National Comprehensive Cancer Network's practice guidelines for MDS and to AML, which cites an extensive number of papers using the NCCN parlance category IIA data accepted by the committees. Based upon that, these aren't based upon prospective randomized trials for the most part, but there is still a robust dataset.

Again, in terms of definition, there is still validation data excluded for the diagnosis of acute lymphoblastic leukemia and that was one of the areas that we thought might need to be addressed further by the measure's sponsor.

In terms of the additional issues of scientific acceptability of the measurement

processes, these are things that we thought could be measured both reliably and with great specificity.

In terms of the latter characteristics of usability and feasibility, the only concern that was raised in this regard is that the data related to MDS were largely abstracted out from outpatient records, which made them more amenable to the sort of assessment methods that were being utilized. Because it was viewed that the majority of individuals with acute leukemias were diagnosed on inpatient basis, some of the Committee members raised questions about the capacity to access those data in reliable fashion so as to provide a fully robust assessment performance under this metric but we felt that it was important to attempt to do so.

CHAIR LUTZ: Is there anyone else from the working group that would like to add insight? Karen.

MEMBER FIELDS: I did the bisphosphonate one and some of the MDS comments made their way into the bisphosphonate one, including the one important one which is they thought that perhaps, at the bottom of 8 of 14, one of the reviewers thought that perhaps this won't be the gold standard for treating and diagnosing and triaging patients in the future. That is only --

And I don't think all the reviewers caught that one

CHAIR LUTZ: Dr. Chottiner?

MEMBER CHOTTINER: In answer to that, I think FISH is probably becoming as important but I think that is under the heading of cytogenetics.

The concern I had was with the acute leukemia population because these patients are captured in the office. It requires an office visit. And my practice was one of the practices that was audited. And

when they asked me to come up with 20 acute leukemia patients seen in the office, Medicare patient, we had three. The problem being that they present in the hospital. They decline treatment. They die in the hospital and it is rare for them ever to end up in the office.

So I just think that in terms of the feasibility of collecting that patient population in the outpatient setting is low. There aren't going to be big numbers.

CHAIR LUTZ: Does the developer have anything to say to help us with that?

DR. ADLER: We had talked about the problem of collecting data on AML patients. ALL actually did not come up in our discussion. And MDS is almost universally an outpatient diagnosis and the marrow is done as an outpatient.

And I would tend to agree with Elaine that there is that difficulty in collecting data on AML patients that are almost universally inpatient diagnoses.

CHAIR LUTZ: Is it a difficulty this is going to make it hard to get any data or do you think that it would introduce a bias? In other words, is there going to be a difference between someone that has data collected in the clinic versus the hospital or is it not going to matter?

DR. ADLER: Yes, I'm not sure.

MEMBER CHOTTINER: I don't think there is going to be any bias. I just think they are going to be very small numbers. It would be nice if there were a better way to get at that patient population but I don't think that is going to be done with this measure.

DR. ADLER: I guess again the issue comes up and I would like Elaine's opinion on it, is that outside academic settings, is it assumed that all patients in the community setting are having cytogenetics done on their AML diagnosis? That is what most of the measure is trying to look at.

You know, there is always those concerns that I often worry what goes on in the field and I know you see patients from parts of the Mid-West where the level of sophistication is not as great. And to have a measure in place that tries to ensure that patients are getting proper baseline testing with AML is appropriate. And that is why we developed this measure to begin with.

MEMBER FIELDS: I guess I didn't understand the distinction of you wouldn't -- It would be harder to be inpatient versus outpatient because I thought when they described the measure it is linked to the presence of an initial bone marrow biopsy and then whether or not cytogenetics were done. So I don't know how that would make a difference where the bone marrow biopsy was performed.

MEMBER CHOTTINER: And you can correct me if I am wrong but I think that the CPT codes that are collected are all office

codes. Is that correct?

MEMBER FIELDS: There is also a list of ICD-9 codes that they were --

MEMBER CHOTTINER: Right but these patients are often, they are rarely seen in the office. So if they are never linked to an office visit, then they don't get pulled in. So it is not an issue of whether it is the right thing to do, an important thing to do, whether it is done. It is an issue of this population not getting captured in the office setting. So can you can you correct me about that? That is how you pick it up is from the office codes.

MS. TIERNEY: Yes, that's right. So that is how we would identify patients for the denominator. As I mentioned with that other measure, with CPT E&M codes and they are all outpatient codes.

I will just add to the discussion, I don't know if Dr. Adler or Dr. Chottiner could add more but when we spoke to some of

the other hematologists that were on the workgroup, they did indicate that they are increasingly in some parts of the country doing initial induction therapy on an outpatient basis, especially in patients of Medicare age. So it might depend on the practice, in terms of how many patients would actually be seen on an outpatient basis with AML.

MEMBER HAMMOND: I have a question about the use exclusively of the cytogenetics and whether or not in view of the changing practice of the diagnosis here, that other molecular FISH tests might not be included, specifically including the CPT codes for those other diagnostic modalities. It would seem like if one used the pathology codes or the presence of bone marrow biopsies rather than the E&M codes for these diagnoses, you would get around the problem of outpatient versus inpatient because bone marrow biopsies have specific codes, SNOMED and STS codes.

I would like the developer to answer.

DR. RALLINS: Speaking to that, again for the electronic data source, you can capture those additional types of tests that are currently not captured in CPT. And we have clinical vocabulary standards that are able to capture clinical data more so than administrative data for these types of things.

Does that help?

MEMBER HAMMOND: But you are not capturing the information from the pathology?

DR. RALLINS: What I am saying is we have the capability to specify that information.

MEMBER HAMMOND: Right but you have not looked at that yet or you have the data or you don't?

MS. TIERNEY: So for claims reporting for this measure to identify patients for the denominator, the measure really focuses on the outpatient management of

patients with MDS and acute leukemia. We are identifying patients based on the ICD-9 codes and the CPT service codes.

It sounds like what you are referring to is the actual CPT codes that indicate that the test was done.

MEMBER HAMMOND: The procedure code. Yes.

MS. TIERNEY: And that really more speaks to the numerator of the measure. And the numerator of the measure in the PQRS program is done through a report of the CPT-II code but not necessarily through analyzing whether any of those CPT codes for the pathology testing were actually recorded.

MEMBER HAMMOND: Or the procedure code. There is a procedure code for bone marrow biopsy which would be done on virtually all of these patients and would help you diagnose. If that was added into the measure, you would be more likely to capture the information whether they are inpatient or out.

MS. TIERNEY: So I do think that we could probably specify and I know for one of the other measures we have specified the numerator to be reported either with the CPT-II code or with the CPT procedure code that relates to the actual performance of the testing. So we could probably look into specifying that as an option for reporting on the measure.

MEMBER HAMMOND: And what about the addition of these other types of tests rather than just cytogenetics?

DR. ADLER: I think it is a really good question about looking at these molecular panels but I think the, and I will defer to other physicians here, but the utilization of molecular panels tends to be still variable I think around the country. And it is not universally being done looking at all the new ways of characterizing MDS and AML. And I think it would be hard to get that to take place at this point. I think it is early in

the use of molecular panels.

That is my opinion. I will defer to Elaine who is in more of an academic setting.

MEMBER CHOTTINER: No, I agree. There are parts of the state where it is very difficult to get the FISH studies. We are trying to educate but we can't get them.

But I would say that I don't think that this issue with acute leukemia invalidates in any way the measure. I think it is just an issue and when you come down to the validity reliability studies, the percentage of acute leuks are always going to be much smaller on this but it doesn't invalidate the importance of the measure.

CHAIR LUTZ: Okay, any other questions or thoughts we should get into?

Does that mean we are good to start voting? Let's do that.

MS. KHAN: So looking at importance to measure and report, you can go

ahead and start voting.

So we have nine high, eight moderate.

And moving on to the performance gap. So I think we are still waiting on one more person.

So we have 11 high, six moderate.

And going to 1c on evidence. Thirteen yes, one now, three insufficient.

CHAIR LUTZ: Okay, then if we move on to question two, is there any discussion anyone needs to have before we get to number two? I guess we already did all of our discussing and did it well?

MEMBER LOY: I just wanted to ask a question in terms of proximity. I'm looking at the description and it says we either got it at baseline or prior to therapy. So I'm just wondering, you know, ordering the test is different from getting the result and I'm wondering at what point do you say you got the cytogenetics but it was three weeks out or

three months out? Do we say that is appropriate or not appropriate?

Was there any consideration given to the time factor as it relates to the initial diagnosis or the initiation of treatment?

DR. ADLER: I think the hope was that everything would happen at baseline at diagnosis, in terms of doing the testing.

MEMBER LOY: Well what does that mean? I mean, was there a time factor that you had?

DR. ADLER: There was no time factor put in except that presentation when the diagnosis was being established to do the cytogenetic testing at that time.

MEMBER LOY: Okay because clinically many things can happen, inadequate material, etcetera, and time delays and getting information back to the folks making treatment decisions.

CHAIR LUTZ: Dr. Marks?

MEMBER MARKS: Is it ever clinically or often clinically acceptable to initiate therapy and then modify therapy based on the cytogenetics when it is pending?

Do you need the cytogenetics to start therapy or can you start based on traditional -- I think you do. Right? Right.

So this business about you feel that we could initiate therapy and then send the cytogenetics -- but the way it is written is --

MEMBER ALVARNAS: Sure. I mean for most of these things, other than APL, acute promyelocytic leukemia where you want to have a good idea and have a very different intervention for most of these individual, you are going to use this for decision-making for either post-remission therapy or for stratification of intensification of therapies, including consideration of hematopoietic cell transplantation. So I think you get started with standard of care

outside of APL and then make your decisions based upon those data. So it is okay if they come a little bit later.

MEMBER MARKS: The way it is worded, it says you must have. The way it is worded it says that it doesn't have to be done before therapy is started.

MEMBER ALVARNAS: The test needs to be set up before therapy starts but you don't have to have the results back before therapy starts.

MEMBER MARKS: Well does the testing have to be set up?

MEMBER ALVARNAS: Because in theory if you induce a remission, you may not have that clonotypic abnormality in the future to analyze. So you may have lost your ability to adequately stratify the patient by having had a good response to therapy.

CHAIR LUTZ: Any other discussions before we move on to voting on reliability?

MS. KHAN: Okay, so voting on

reliability. You can start now.

We have seven high, nine moderate, and one low.

And moving on to validity. We have eight high and nine moderate.

So now we can move on to usability.

CHAIR LUTZ: So if it's all right with you guys, we will just go straight through. Then usability, feasibility, unless someone needs us to stop.

MS. KHAN: So voting on usability. We have ten high, six moderate, and one low.

And voting on feasibility. We have five high, 11 moderate, and one low.

And then overall suitability for endorsement, does the measure meet NQF criteria for endorsement?

We have 17 yeses.

CHAIR LUTZ: All right, so it looks like it was a good measure but maybe also the promise of lunch has really moved us

forward quickly.

The next one up is 0378, documentation of iron stores in patients receiving erythropoietin therapy.

DR. ADLER: So Measure 0378 I think it is fair to say that in 2006 and 2007 we felt that again myelodysplasia was becoming a more common entity. As the American population ages, we are just seeing many, many more cases of MDS. So this measure is the documentation of iron stores in patients receiving erythropoietin therapy to document iron stores prior to them starting erythropoietin therapy. And the numerator is by documenting iron stores either by a bone marrow examination or by a serum iron, iron-binding capacity or by a serum ferritin. And the denominator is all patients diagnosed over age 18 with a diagnosis of MDS.

And again, it is interesting over the last five years how controversies have evolved about the use of epo therapy in ESAs

but nevertheless there has been much less controversy over the use of ESAs and myelodysplasia. And for patients prior to starting this rather expensive form of therapy, it is important to know that they are replete with iron prior to starting therapy. And that is the purpose of this measure.

CHAIR LUTZ: I think Dr. Chottiner was the primary reviewer for this.

MEMBER CHOTTINER: So it is interesting that the FDA REMS program excluded myelodysplasia I think because they felt the potential benefits outweighed the risks. For patients who are receiving Procrit or Aranesp because of chemotherapy or chronic disease, those patients need to have verification of iron stores but myelodysplasia fell outside of that.

So in terms of randomized trials and evidence we don't have anything that falls into that category but there is a large body of support, evidence-based guidelines, the FDA

REMS program. So I don't have issues with that. I agree that it has very high impact. I don't really have any issues going forward. It is very easy to measure. We do it for all of our other patients receiving erythropoietin stimulating agents.

CHAIR LUTZ: Anyone else from the smaller workgroup that went over this? Dr. Fields.

MEMBER FIELDS: My only -- I agree with everything you said. My only concern was one of the measures of documentation of iron stores was looking at the iron bone marrow stores. And that is more of a subjective measure when obviously you could have a more quantitative measure if you measured it in the serum. So I just wanted to ask the authors why they included that when serum measures are cheap and easy and more quantitatable.

DR. ADLER: Yes, I think it is interesting if you have someone who looks like they have MDS but have absent iron stores,

then that is excellent proof of iron deficiency. You know, sometimes the iron stores may be more reliable in that in that setting than measuring the ferritin or serum iron. So I think using both measures of iron deficiency seems appropriate.

If you have somebody with sideroblastic anemia, then the irony is you will see increased iron in their bone marrow. And I think the point is to try to capture these patients prior to them starting therapy. Because there may be patients who are iron deficient who fail to respond to ESAs because they are iron deficient and this would help promote the proper use of iron therapy in patients who are iron deficient. So I think the marrow can complement that.

MEMBER FIELDS: Well I agree that can complement it. I was just wondering if it made it as a reliable of a measure when it just came down to the reliability. That was my only question. Otherwise, I think it is a

very good and appropriate measure.

DR. ADLER: Okay.

CHAIR LUTZ: Yes, sir?

MEMBER LOY: I may be misreading this but under the numerator statement I am seeing we have got measures of iron stores or serum iron in total iron-binding capacity. I am wondering, wouldn't you want both?

DR. ADLER: It would be nice probably to have both and probably both should be obtained. that is reasonable.

MEMBER PFISTER: Yes, my question also went along those same lines which is that the -- Actually I think I had somewhat a different spin on it than Dr. Fields did, which is that actually I was thinking that the bone marrow iron would be probably a little bit the gold standard of what you want and that the other test is complimentary to that is one thing. But let's say if again I am not a hematologist I don't want to sort of imply that but if my understanding is that ferritin

can be kind of an acute phase reactant and that you can have an overestimate of your sort of -- if you are otherwise sick that your ferritin may be factitiously up. How big of an issue that is.

Or that similarly that when you have iron in TIBC there are other things that can kind of make that like you know compared to let's say if you routinely did a bone marrow on everyone and you had that test and you did the predictive value and so forth, that you would end up with -- I have always been taught sort of like the marrow is kind of the gold standard. These other things have other reasons why you can kind of be off. And I guess if you are really doing for what sounds like a very good reason to know what the deal is because I am struck in my own experience or seen people prescribe when it was a lot easier to prescribe them, that they often didn't adhere to any of this.

You know, you probably want to do

it in the way that is going to optimally inform the decision. You know, you are doing it up-front and I guess that it would be helpful to have some reassurance that are we getting good enough if we were not to do a bone marrow?

MS. FRANKLIN: Dr. Alvarnas?

MEMBER ALVARNAS: I guess in that regard I am just going to mention briefly. You know, when you look at serum irons and ferritins, they are not as bad as you might think. For a single ferritin I mean the coefficient correlation is like 0.55 but if you do two, that number approaches over 0.8. So they are not bad, even though the bone marrow is still the gold standard.

I guess if you could do a bone marrow aspirate in biopsy without causing pain or discomfort but for some of these patients they may be down the road from their initial diagnostic study which we know from the last metric is essential. But it may be that we

just don't want to subject them to a procedure that can be quite uncomfortable to using the ESAs. And I think that the serum iron measurements or assessments, whichever constellation of them we use, represents a suitable alternative. Because I think one of the standards we talked about was that the benefit exceeds the danger caused to the patient. And I do worry that bone marrow biopsies are uncomfortable and patients really are reticent to undergo them. I hate to have that be the stopping point to compliance with this metric.

So that is why I do view the serum iron assessments done in an appropriate data-driven way, as being a suitable alternative to the bone marrow aspirate and biopsy.

MEMBER PFISTER: But just so I understand it, so you are saying on one value did you say that the correlation coefficient was 0.55?

MEMBER ALVARNAS: The serum

ferritin. It is like flipping a coin if you do just one of them, but if you do more than one, the coefficient correlation goes up significantly.

And you are right, the qualifications are that there is no hepatitis, that there is no concomitant inflammation, infection, or some other state that is going to drive up the serum ferritin, including autoimmune disorders.

But I think in the hands of a knowledgeable practitioner, those are things which would be appropriate.

MEMBER PFISTER: I mean, just again on its face I would think that it is hard for me to jump up and down about correlation of 0.5. I mean for something that is pretty high technology thing we are doing and we are doing intervention with it. Like the 0.8 sort of passes the Smith test a lot more to me but you are making this decision up-front. You know, the measurement is

clearly an up-front measurement that we are measuring that behavior and it seems to be sort of a low correlation to me.

CHAIR LUTZ: So would your concern be that it needs to be defined with greater specificity?

MEMBER PFISTER: Well I think that it is -- I am totally onboard that this is very important information to know before you start giving something that potentially has down sides and that is highly expensive because you want to make sure you are getting the bang for the buck that the indications there.

I guess it sounds like if you have it done by these less-invasive means it would be helpful for me to have more, like what percent of the time is done by the less invasive means in real life? Like all these people have a marrow. Don't people generally get the iron stores as part of the initial marrow?

I guess that would be helpful because I think the correlation coefficient of 0.55 seems to me that a lot of the time you said it was a coin toss, I mean 50 percent of the time it is almost like -- Then I would question well gee how does that inform your decision if it is a coin toss after you are doing the test.

MEMBER ALVARNAS: An isolated ferritin if you are also adding in measures like iron saturation, TIBC, you can make that a far more robust measure.

I guess my concern if we make the only acceptable measure bone marrow aspirate and biopsy that your compliance rates are going to be really, really low.

And I think for this particular disease, this may be one where somebody has a bone marrow aspirate and biopsy and years later when they become transfusion-dependent, then you are asking this question. So I think because this is a disease that can have a

certain amount of chronicity before the patient becomes either symptomatic or transfusion requiring, that is why I think the bone marrow isn't the only means by assessing the iron stores.

Dr. Adler could probably speak to that better than I could.

DR. ADLER: Yes, I just think if you have a zero ferritin level of low iron saturation, that will be enough documentation.

MEMBER ALVARNAS: Yes, I think if it is low you can be pretty sure. If it is high, then I think the reliability of the high number, particularly in light of those covariates that you talked about make that more suspect.

MEMBER CHOTTINER: The primary reason for testing it is to replete iron stores in somebody who is iron deficient before you start the ESAs. So we are really looking more for the low ferritins and worrying too much about why they are high.

MEMBER FIELDS: And I guess to -- I think you articulated it well. I think the main question is again we want it at the baseline prior to initiating therapy. But you are going to continue to want to measure iron stores along the way to make sure that the patients don't become iron deficient while you are treating them with ESAs.

So my question was why don't we use both or why would we use just the iron stores? Because otherwise, how would we continue to document whether or not the patient wasn't iron deficient.

So I'm just trying to make it so we have a delay until we get lunch.

(Laughter.)

MEMBER PFISTER: So I'm not quite that hungry.

So it sounds like you are not so much worried -- if it is low then sort of it is low. But so if it ends up, let's say it ends up being a false positive. Then so it

comes back and the ferritin is up. And you say oh, I feel good. So you wouldn't supplement them. You would just treat them. Or would you give them iron anyways?

MEMBER CHOTTINER: We treat them and follow.

MEMBER PFISTER: So just to be clear, is it you are going to give them iron anyway while they are getting the ESA? It is just that you would --

MEMBER CHOTTINER: No, I don't think we routinely do that. I mean a lot of patients with myelodysplasia will come in with high ferritins because of ineffective erythropoiesis or because they are iron overloaded. So we would not generally treat unless they were iron deficient or if it looked like they were become iron deficient over time.

MEMBER PFISTER: So when the value comes back either normal or high, then most of the time -- there would be a minority of

circumstances that that would be like a false positive in terms of what their store situation is.

MEMBER CHOTTINER: Probably. You know, the REMS program I think requires that the ferritin be kept over 100 for the duration of treatment.

CHAIR LUTZ: And if you were giving an ESA and you didn't get the response you wanted in the endpoint, you would potentially repeat the ferritin anyhow.

MEMBER LOY: If you get a normal -- Let's try to get to this question when we were talking. When you get to a normal ferritin, don't you still need a total iron-binding capacity?

MEMBER CHOTTINER: Ordinarily, we would get all three. I don't think the ferritin is included in the measure but it is something we usually follow.

MEMBER LOY: When I was reading this, it sounded to me like you could get

either or.

MEMBER CHOTTINER: Usually we would get all three.

MEMBER LOY: And I'm thinking I'm hearing agreement that the measure could reflect that. I heard the word reasonable but I don't know if that is an option for us here today at this table.

MS. FRANKLIN: Could the developer -- I know some things were added to this measure, this particular measure I believe in 1c.16, would that answer or just the issues around ferritin?

I believe it is the guideline for ASH.

MS. TIERNEY: So I think the measure as I think Dr. Loy, hopefully I got that right, was saying, does account for either/or. I can see actually that the statement after the or is a little confusing because it is a serum iron measurement by ferritin or serum iron in TIBC. But here

where the or and the and comes in, so we could probably confer with our workgroup and try to clarify whether TIBC is always required with serum iron measurement. It is a little unclear from this and the documentation where the or comes in with that.

CHAIR LUTZ: Are we good enough to vote on that? Do we have any --

MS. BOSSLEY: I guess the one question I have is when it was tested and you abstracted, how did they pull it? Because that may help clarify it and I don't know if you have the answer to that.

Okay, because that is what your testing results are based on. And it would be useful for the Committee to see that part and then I am assuming you would need to know that before you move on. But I don't want to put words in your mouth.

MEMBER CHOTTINER: They pulled all the iron studies, they pulled the ferritin, iron, iron-binding capacity.

MS. BOSSLEY: So if you looked at how it was written, Gene can you scroll back up?

So here it says documentation includes either bone marrow examination including iron stain or and I am assuming it is iron stain then it would be bone marrow examination including serum iron measurement by ferritin or serum iron and TIBC. How did --

MEMBER CHOTTINER: I know they went through and looked for all the iron studies because that was very painstaking.

MS. BOSSLEY: Okay. So Sam, I think we need to clarify that it looks like. I'm not sure even by reading it. By the large or, I assume it is iron stain or anything else that is listed, I would assume.

CHAIR LUTZ: So where does that leave us right at this minute?

MS. BOSSLEY: So I guess the question is how quickly can you get that

information? Oh, okay.

So would you like to defer discussion on this and move to the next measure and see if they can answer it?

CHAIR LUTZ: I think we should.

MS. BOSSLEY: Okay, let's do that.

CHAIR LUTZ: Next is 0379. We will let our developer discuss 0379.

DR. ADLER: 0379 is the use of baseline flow cytometry at the time one is making a diagnosis of chronic lymphocytic leukemia. And the numerator is to include all patients who had baseline flow cytometry at the time the diagnosis is being attempted to be made and the denominator is all patients with a diagnosis of CLL. And again the impact here is the importance of differentiating CLL from other types of lymphocytosis and the use of the flow cytometry will confirm the diagnosis by demonstrating a monoclonal cell line of lymphocytes and would help differentiate the diagnosis from other

conditions such a mantle cell lymphoma, non-Hodgkin's lymphomas, hairy cell leukemia, infections and a newer entity called monoclonal B cell lymphocytosis.

So CLL is felt to be really confirmed by the diagnosis of baseline flow cytometry initially at the time of diagnosis because many patients will not require treatment for an indefinite period of time. At least at the time of diagnosis, this should be performed.

CHAIR LUTZ: I think Dr. Chottiner gets to speak again.

MEMBER CHOTTINER: Thank you. So chronic lymphocytic leukemia high impact. It is a common disease, especially in the elderly. The problem is that the lymphocytosis is very difficult to look at morphologically and flow cytometric analysis is very important, particularly if you are going to initiate treatment because obviously treatment is very different for some of the

other low grade lymphoproliferative disorders.

I was surprised at the performance gap that was identified in the 2008 PQRS but it is high. It is very easy to extract the information from the chart. And I didn't have any issues with this other than trying to correct the timing of the numerator and the denominator because what is difficult is if you make a diagnosis of chronic lymphocytic leukemia in 2002 and you initiate treatment, which we sometimes do in 2012, then it is a little bit difficult to reconcile the numerator and the denominator.

CHAIR LUTZ: Does anyone else from that small workgroup have any comments to add? Anybody from the larger group?

MEMBER HAMMOND: I'm wondering because of what you said, does that mean that would that cause the performance gap if you didn't adequately find the flow on the patient, then it would explain a performance gap and it might just be an artifact of

measurement.

MEMBER CHOTTINER: I'm wondering that too because I watched the auditors go through, they had to go way before our electronic medical record back into the paper charts down in storage. And I think that the intent of the measure needs to be clarified so that we are simply saying that for newly diagnosed chronic lymphocytic leukemia, a flow cytometric analysis is required if you are going to treat.

MEMBER LOY: I like this measure. I am a little disturbed by the gap that you have identified. It is surprising to me.

But having said that, I have also heard in the presentation by the measure developer that a lot of this was geared towards trying to get an accurate diagnoses. And having flow performed, in my view, doesn't necessarily guarantee that the diagnostic workup was appropriated and addressed the differential diagnosis. It doesn't feel like

it really drives towards that quality outcome.

MEMBER CHOTTINER: If there is a pathologist here you can correct me but I think the flow cytometry profile for CLL is pretty well established.

MEMBER LOY: Agreed. It is established but I think that there is, in the differential, trying to make sure that you have ruled out a mantle cell is problematic for many folks who are doing it. If it is not adequately performed, then --

MEMBER HAMMOND: Yes, if the flow is not adequately performed, absolutely. But I think the standard -- I don't think we have any data here to help us understand what the typical way in which flow is done in that situation. It should answer the question if it is done. But I think there should be some change in the way the measure is written so that the flow has to be done in some period coincident with the treatment instead of basing it on a flow that might have been done

a long time before that because that would be, I think, a more valid measure.

Would the developers please comment about that?

DR. ADLER: I think the initial goal is to really do the flow cytometry to confirm that the disease indeed was CLL and not another entity. And you know, the flow may or may not be repeated at the time of treatment initiation but the intent was to confirm the diagnosis of CLL at diagnosis with flow. That was the initial intent in 2006 and 2007.

CHAIR LUTZ: Dr. Marks?

MEMBER MARKS: I have two questions. In a patient being treated, this should be a never event? You should always get this 100 percent? Okay.

And then are there a bunch of patients that don't get treated? So the way it is worded it is at diagnosis. So if you have another new person you are not going to

treat, you shouldn't have to get this. So the measure should be reworded to be those who are initiating therapy.

And then another comment would be in the patient who is sort of older who isn't going to tolerate aggressive therapy, you might treat him with prednisone anyway or you might treat him with something in any case? I'm not a hematologist. I'm just asking.

Okay, so should it be reworded to be patients receiving therapy with, I don't know if you call it a class of drugs or a curative therapy or aggressive therapy or some --

MEMBER CHOTTINER: I think that is going to turn it into a very complex measure.

MS. FRANKLIN: Dr. Malin, you had your hand up. And then Dr. Hammond.

MEMBER MALIN: I have a couple of questions. First is I am trying to understand the numerator and the denominator. But it looks like the denominator is people with a

diagnosis of CLL who are starting treatment or who just have a diagnosis?

DR. ADLER: Just a diagnosis.

MEMBER MALIN: Just a diagnosis.

DR. ADLER: Not starting treatment.

MEMBER MALIN: So regardless of starting treatment, you have a diagnosis of CLL. And then the numerator is a flow cytometry in the prior 12 months. Right? No?

DR. ADLER: It doesn't say 12 months.

MEMBER CHOTTINER: No, it doesn't. It is ever.

MEMBER MALIN: It says consecutive -- oh. So it's an ever. So how far back, realistically do you look?

DR. ADLER: Well the point was to have it done at some point in time to confirm the diagnosis.

MEMBER MALIN: I mean, it says at least once during the measurement period. So

what is the measurement period for the indicator?

MEMBER CHOTTINER: That is a year. That is 12 months. That's where the confusion is.

MEMBER MALIN: So if it is once during the measurement period, then it is an annual flow cytometry the way the measure is specified.

MS. FRANKLIN: Dr. Hammond?

MEMBER HAMMOND: I have a question, too. When we come up with all these suggestions about things that might be done to improve the measure or ways in which it is confusing, then we vote. What happens about measures that could be improved? Is there any requirement by the developers to do the things that we are suggesting or to consider doing them? Or how do we know that something changes?

DR. BURSTIN: That's a great question. So that is why would ask you to

vote on the measure as it is before you. The measure developer then has an opportunity as they go back and reassess. They could always try to bring it back forward to you with making those changes.

We just try to keep it pretty clean. This is what you have got. This is why the developers are always invited. They are here listening to the suggestions, if they choose to want to try to come back with some minor tweaks.

And again, keep in mind we can't -- they can't completely rewrite a measure. It has got to stick pretty close to the measure as submitted because it has been tested in that way.

MEMBER HAMMOND: So if we want the measure to be changed, how should we vote to make sure that that happens? Not vote for approval, is that what we would do?

DR. BURSTIN: Yes, I mean some of it really comes down to the kind of changes

you are talking about. A very slight tweak or a clarification which the developers can usually just agree to is fine. If you are really suggesting a substantive change to the measure, then probably what makes sense is to actually vote it down as is and allow the developers to bring back a revised measure.

MEMBER MALIN: The second question I had, which is isn't -- I mean I just wonder this measure almost seems tautological to me. I mean, I just can't imagine the definition essentially of the diagnosis of CLL to go from lymphocytosis to CLL involves doing flow cytometry. And if you are using the ICD-9 code to identify the measurement population, you are excluding all the people who basically didn't have the test done to make the diagnosis.

MS. FRANKLIN: Is there any comment from the developer on the specifications there?

DR. ADLER: The implication also

was that the dong the flow cytometry does have treatment implications and that there may be information coming forth from flow cytometry such as finding someone has CD38-positive CLL and has a higher risk for rapid progression of disease and a poor prognosis.

Now whether we would want to take this back at your suggestion to say that flow cytometry should be done prior to treatment rather than at the time of diagnosis, that is something that we could certainly entertain as an approach.

MEMBER FIELDS: In the numerator statement it says refer to testing -- baseline flow cytometry studies refer to testing that is at the time of diagnosis or prior to initiating treatment for that diagnosis.

So I assume that the group was already looking at both of those scenarios.

MEMBER MALIN: I think the difficulty we would need to know how it has been operationalized is then below that it

says the measurement window is just the prior 12 months. So those two statements are in conflict, basically.

MEMBER CHOTTINER: Actually, looking at that the denominator is still all patients aged 18 years and older with a diagnosis of CLL. So it is not really an issue again of the measure itself. It just makes it somewhat difficult when you get down to feasibility to extract data that may go back several years prior to the denominator time window.

So we are still looking at patients with a diagnosis of CLL who have either been diagnosed or are going to get treatment. That is the numerator. The denominator is still all patients with CLL. it is just that to get at the data piece that you want, which is the flow cytometry, you might need to go back prior to the denominator time window.

MEMBER MALIN: Correct. But the

material that was submitted states that the numerator time window, that is the time period in which flow cytometry would be looked at is the measurement period, which is one year.

MEMBER CHOTTINER: That is not how it is being used, I don't think.

MS. TIERNEY: This is Samantha Tierney. I think that that is correct. I'm not sure if that is an error. I think unless -- I think the expectation is that you would report on this measure in a year but you are reporting that the baseline cytogenetic testing was done. So maybe that is why that language is worded like that. We can confer with our specifications colleagues who completed this. But it should be similar to the other baseline measure because they are kind of mirrors of each other.

CHAIR LUTZ: So is this another measure we have to wait on more information?

MEMBER DONOVAN: Can I get a question just on the process here?

So I'm not a clinical expert in this area. So I sort of fall back on looking at what has been presented for the measure and what other experts in the room are talking about. And it seems to me in both of these cases, we are going through another exercise in sort of face validity of the measure, which I think this dangerous. Over half of the review panel in each of these past two measures have said that there is insufficient evidence for this measure. And all of these discussions around here, extensive discussions suggest that there is questions about the validity of the measure. So I sort of wonder about going through a second exercise of consensus or expert review on something that that has been done and presented already.

DR. BURSTIN: I see. The workgroup level is intended to give us sort of a preliminary sense of what people need to focus on but we really do rely on this group, which is multi-stakeholder, lots of different

voices at the table.

You certainly are, it sounds like giving a fair amount of deference to our hematologists at the table on evidence, which is fine.

MEMBER DONOVAN: So I guess it would be helpful for me if we could hear sort of the evidence piece. Because I think when I look at the summary from this working group preliminary review, half of the people on the Committee felt like there wasn't adequate evidence for this measure. And since that is a stopping point at some level, I wonder whether that working group could address that issue.

MS. FRANKLIN: Dr. Alvarnas?

MEMBER ALVARNAS: Yes, I guess being a hematologist on the working group, I think our issue wasn't so much when we discussed this whether or not there was appropriate evidence for doing this. I think unequivocally flow cytometry is the

appropriate thing. I think where we ran into some issues of discussion was the timing. Because you are right, if the question is ever then you may have someone who as the doctor here mentioned has an interval of ten years between their initial diagnosis and when they might require treatment. But the relevance of flow cytometry in that instance may be the decision not to commence therapy. So that you have distinguished that patient from somebody with mantle cell lymphoma in a leukemic phase.

So the value is what you don't do based upon that information, namely starting appropriate chemotherapy.

So I think at least from a working group perspective, I think one of the things if we could clean up the definition of the numerator and denominator with respect to some of the timing issues, it would obviate a lot of the concern, a lot of the angst that we have raised in the last 20 minutes.

MEMBER PFISTER: The one thing

that was a little confusing to me when I was looking at originally what was submitted by the proposer is that they say that the NCCN data was category 2a and then say it was in non-uniform consensus, which I think is sort of automatically sort of makes it a 2b. So one of them, I think, is not -- is inaccurate.

I mean, I just brought up the NCCN guidelines. It looks like it is a 2A but again, I do not --

MEMBER CHOTTINER: It's a 2A.

MEMBER PFISTER: Okay.

MEMBER CHOTTINER: They put the wrong qualifier in there.

MEMBER PFISTER: Okay.

MEMBER CHOTTINER: So I don't have any quarrels with the evidence either. I think the problem with both of these, myelodysplasia and chronic lymphocytic leukemia is that they are chronic diseases. You can follow patients for a very long time without treating. So some pieces of the

diagnostic evaluation are just going to be remote from the year that we are looking at. But I don't think that that makes the importance of measuring them or the validity any less.

MEMBER PFISTER: Like what percent of like if you take like a hundred people who come in, they kind of look and smell like CLL and then you do the flow, what percent of those end up being something bad that you intervene on?

MEMBER CHOTTINER: Five to ten percent end up being marginal zones that are less ominous or mantle cell that is more ominous.

MEMBER HAMMOND: I think though that the way we are talking really compromises the validity of the measure. Because if you don't know when the flow was done, you may under report flow cytometry and therefore invalidate the measure.

What we are trying to do is get an

accurate measure of how often flow is done in people that might have CLL. And the way that the measure is constructed, I think it really fails on two. The measure properties do not allow us to measure this important outcome.

CHAIR LUTZ: Well I guess at some point we have to decide if we feel comfortable enough moving to a vote on it or if there is a uniform request from the submitting folks as to how they could change it. I mean, you have to kind of step up to the plate I think one way or the other.

DR. ADLER: The point of the measure really was to try to assure favorable patient outcomes by establishing the fact that flow cytometry is an important baseline test to obtain at the time of diagnosis. Now we could certainly take it back and say that it should be done prior to treatment but the bottom line was to have it done at some point in time in that patient's clinical course.

CHAIR LUTZ: Should we vote and

then see if that leads us to request a big change or not? Is that alright?

MR. CUNNINGHAM: We just need a couple more so can you hit it again, please?

Nine high, six moderate, and two insufficient.

Still waiting on one. Everyone please vote again. Seven high, eight moderate, and two insufficient.

Eleven yes, six insufficient.

CHAIR LUTZ: So by that criteria, we continue. Correct?

DR. BURSTIN: Yes.

CHAIR LUTZ: Any further discussion before we move on to the next question two, voting two?

MEMBER MALIN: Do we have clarification on the issue of the timing of the measurement of the numerator? I mean, it is hard to evaluate validity if we don't have that clarification.

CHAIR LUTZ: Correct me if I'm

wrong but I think we are actually voting on what we have and what we see in front of us. And then if we don't like it, then they will get the idea and change it if we are comfortable. So it kind of a thumbs up, thumbs down.

MS. KHAN: So voting on reliability. We have one high, six moderate, four low, and six insufficient evidence. So we don't go forward.

CHAIR LUTZ: Okay. So if we add low and insufficient together, I think that stops us there.

MS. KHAN: Yes, right.

DR. BURSTIN: I think the question is do the developers feel like they have a sense of the potential suggestions made by the committee or does it need further discussion?

MS. TIERNEY: I guess, you know, one thing, I think there was a little confusion about the numerator time window. And I don't know if it would affect the voting

results but I just wanted to offer some clarification.

So in a program like PQRS this measure is reported in a 12-month time window, which is what is specified in the denominator time window. But the numerator time window says once during the measurement period because we would expect that a physician would report on the measure once a year. That doesn't they have to perform the flow cytometry once a year but rather that they have to report that baseline flow was done once within that time period. So I don't know if that offers any further clarification but I think that was a point of confusion earlier.

MS. FRANKLIN: Dr. Hammond.

MEMBER HAMMOND: But reporting that it has been done is not the same as having evidence that it was really done. Right? I mean, if they just report that it has been done, how do you have any evidence that you know it was, if it could have been

done ten years before? They have to show you a report or there is some data that you gather that proves that?

DR. RALLINS: Yes, so again, there are terminology codes to actually capture the performance of that particular test, flow cytometry test.

MEMBER HAMMOND: Yes, there are performance codes.

DR. RALLINS: Yes.

MEMBER HAMMOND: Are you capturing those performance codes?

DR. RALLINS: We have the ability to capture the performance codes.

MS. TIERNEY: Right. For PQRS purposes, you have to use a quality data code which is either a CPT-II code or a G code. So for the PQRS program, a physician reporting on this measure would have to report one of those codes. I think there is an expectation just like with any sort of billing data that there would be information in the medical records to

substantiate the code that has been used. And I think our testing project found information in the medical record to support that.

But for electronic purposes, we probably would specify the measure to allow for the actual looking back to see the actual performance of the test, like the procedure similar to this same issue with the last measure.

DR. RALLINS: Right. And then I would add that so looking back at the performance using an electronic data if you are using claims data, the CPT-II code is supposed to have the same strength of actually reporting the actual performance. I hope that helps.

MEMBER MARKS: If a hematologist is following a patient and is not treating a patient, you wouldn't necessarily even be stating any of this in your annual note when you see the patient who is chronically doing well and you are not doing anything for them.

Right?

MEMBER CHOTTINER: Would I state what the flow cytometry showed?

MEMBER MARKS: Right.

MEMBER CHOTTINER: Probably not.

MEMBER MARKS: Right. And to force a practitioner to go look it up is an onerous --

MEMBER CHOTTINER: But we did.

MEMBER MARKS: You did?

MEMBER CHOTTINER: Yes, and when they came through they were able to identify all of them. It just took a bit of work. So I am uncomfortable about flunking the measure on that basis. I think it makes it more onerous but I don't think it is makes impossible for less meaningful.

And if something had changed dramatically, I mean we repeat the flow cytometry so there may be a more recent or there may be multiple flows in the chart.

MEMBER MARKS: Well would

hematologists get in the habit of diagnosis CLL, comma, flow cytometry, whatever or however you report it?

MEMBER CHOTTINER: You could argue a hematologist wouldn't call it CLL unless they did the study.

CHAIR LUTZ: All right. Are we good to move on to 3080 then? So we have a nice option being given here. We still have to do public comment either way. But do we want to do public comment and then break for lunch or do we want to try to get through 0380 and then public comment and then go to lunch?

I guess the question is, how hungry is everyone?

DR. ADLER: I'll be very quick.

CHAIR LUTZ: If you guys are good, let's just give 0380 a look-see.

DR. ADLER: So 0380 is the use of bisphosphonates in the treatment of multiple myeloma. And the numerator here is all patients who received IV bisphosphonate

therapy within the past 12 months and the denominator is all patients over the age of 18 who have been diagnosed with myeloma. And the point of this is that we know there is a beneficial effect of the use of these drugs to reduce the possibility of pathologic fractures and to reduce bone pain as related to myeloma.

And the hope of the measure is to ensure the fact that patients with myeloma are receiving these treatments on an appropriate basis. And I will leave it open for discussion.

CHAIR LUTZ: I think Dr. Fields was the primary discussant.

MEMBER FIELDS: Yes and what I would say is you described the measure adequately. I think that the main striking information is bisphosphonates have been around for more than a decade and the gap in care for prescribing bisphosphonates in the patients that were in the measure was 47.4 patients for some of the patients didn't meet

the measure.

Although what I didn't understand is the next measure said 86.6 but still there is a huge performance gap. Given the fact that bisphosphonates do have evidence a prospective randomized trial in that the authors described four prospective randomized trials that described the benefits from a decrease in skeletal complications, decrease in vertebral fractures, and decrease in pain.

And then one of our committee members also reminded us and updated the most Corcoran analysis showing again the number of randomized trials supporting the use of bisphosphonates went up.

So this was actually endorsed by multiple external review bodies as a category one or a grade A or the highest level of data to support the use of bisphosphonates in these patients. So it was just striking that the performance gap was that high.

I would also comment that it is a

easily reliable measure. You got the drug or you didn't get the drug and the diagnosis was pretty well outlined. The main issues were that the literature supports the use in patients with lytic lesions and you are drawing a conclusion that all patients with evidence of any bony involvement should get bisphosphonates but I think that the author has adequately described the reasons for that.

And also one of our reviewers reminded us that bisphosphonates should be given on a monthly basis, yet it is an annual measure. So should the measure be done more frequently, say every three months rather than annually, although I think that makes it more onerous. I assume that if the provider knows that the patient should be receiving bisphosphonates then they would be giving them and I would think that that would still measure quality.

And then they also reminded us that these drugs are not without harm. There

is evidence that patients can get osteonecrosis of the jaw. My personal comment on that is that I think that the providers are well aware of that as a complication and that the patients are educated aggressively about the use of bisphosphonates and any dental disease.

So I think that when you look at some of the randomized trials, that complications continue to decrease over the years because of our knowledge about how to manage bisphosphonates in that patient.

And I will also just add one more caveat. None of the studies showed an improvement in survival or progression for survival but they did show improvement in quality of life and decrease in bony complications. So I think it is a reliable and valid measure of high importance.

CHAIR LUTZ: Does anybody in the smaller workgroup have anything to add about that?

MEMBER FIELDS: It was the only measure so far that had level one evidence.

CHAIR LUTZ: Can I ask a semantic question? When I read a recent review of some of the folks from ASCO that did the ASCO guidelines for they preferred to be called osteoclast inhibitors. Does it make a difference if it says bisphosphonates versus OIs? Because the difference was up in my face writing this review pretty heavily. Do people have a strong feeling about the need to say OI versus bisphosphonates?

MEMBER FIELDS: I think that is because there are drugs, new categories of drugs that address, you know, have a different mechanism of action. So I didn't.

In myeloma there is not data that the new categories have any validity. The only two drugs that are useful are pamidronate and zoledronate in this instance.

MEMBER PFISTER: It's been a while since I have looked at this data but my

recollection is kind of like the scenario with the ESAs that you need to think about iron that when you do the whatever the new OCAs or whatever it is, that you need to worry about vitamin D and calcium supplementation. And what strikes me is actually where there is probably as huge a performance gap is that people just give the bisphosphonate and don't do the thing that theoretically makes it work better. Did this come up at all in the small group discussions?

MEMBER FIELDS: I will confess that I was on an airplane and missed the small group discussion. So I defer to the rest of the members that participated.

MEMBER CHOTTINER: No. Our instructions were to take these at face value.

MEMBER PFISTER: Yes, because I am just thinking just to get out on the tail of it, that certainly the data in this setting that the intervention with the bisphosphonate certainly as you correctly pointed out, we

don't see that level one evidence based on the deliberations. So it sticks out.

But it also strikes me how well people actually do the intervention which is a highly expensive thing to do that you are giving every month, I am always struck how people, it is amazing to me how often they are not on vitamin D and calcium.

And so are sort of tracking on the thing that yes, you are doing it but you are not doing it well. And we are doing nothing to leverage that behavior.

CHAIR LUTZ: Bryan, I think you had --

MEMBER LOY: I just wanted to make sure that I understood the answer.

Are there other drugs with a different mechanism of action that are either indicated or acceptable off label use for this? I didn't think I heard the answer.

MEMBER FIELDS: No, their drug was approved, as far as I know just in breast

cancer and other cases. It wasn't approved in multiple myeloma.

MEMBER LOY: Okay, thank you.

CHAIR LUTZ: Yes, Jennifer?

MEMBER MALIN: I think the one concern I have with this measure is it seems like it is setting the bar for quality pretty low. To say that administering it once over the course of the year is sufficient, I mean you could still do a one-year look back but say that it had to be administered at least nine times over -- something like that.

MS. FRANKLIN: Will the developer speak to the time frame again, please?

MS. TIERNEY: So I think like many other prescription measures or drug therapy measures, I'm thinking of many in the cardiovascular realm, because it is a measure you are just trying to look at one point in time. So I think that is why the measure only looks at just at least once within a 12-month period. Certainly the workgroup when we

developed this measure discussed that this is something that is done routinely but from the purposes of a measurement, we just wanted to measure it at one point in time.

I also hear on the small group discussion, although I don't remember this from the workgroup discussion, that there is some evidence that every three months is appropriate, maybe every month. So I think it would be a little difficult for us to define a time frame if there seems to be some controversy about how often it should be given.

MEMBER FIELDS: My interpretation wasn't -- well other people were in the group. I think the question was should we just measure it more often, not should we give the drug less often, unless somebody had a comment.

MEMBER CHOTTINER: No, I think given the constraints of how we report these, we really do just fill out our PQRS forms once

a year. So it would be difficult to come up with a schedule.

MEMBER GORE: Plus even when you look at the performance reporting, even for a low bar like this, a glaring number of patients don't meet the measure. So, even though it seems like a low bar, it is a low bar that people aren't meeting.

MEMBER MALIN: An example of similar measure in cardiovascular -- You said there were other examples of measures that you used to have similar --

MEMBER BRUERA: Sorry. Within the NQF portfolio some of them are PCPI measures, some of them are others, there are many measures that looked at the patient population with coronary artery disease and the numerator is are they receiving antiplatelet therapy, are they receiving beta blocker therapy and within the time window do they have a prescription.

MEMBER MALIN: So they might have

a prescription but they are not taking it?

MS. BOSSLEY: Yes. Some other measures start looking at more drug utilization and looking at the proportion of days covered. Those are slightly different than what you see here, in part because those are from claims data and using pharmacy claims. So in part it is depending on your data source but it does vary.

CHAIR LUTZ: Yes?

MEMBER TAPAY: Hi. I just have a question for any member of the workgroup. Was everybody comfortable with the exceptions process on either the patient or the provider level?

MEMBER FIELDS: I think that the exceptions are easily documented and the providers completely -- the average provider that uses these meds understands well those exceptions. And if not, the pharmacists that are dispensing out IV medications frequently understand the renal exceptions and some of

the other kinds of exceptions. So I would think it is a pretty standard exception. I don't think that -- I didn't hear the rest of the discussions.

MEMBER MARKS: But although with those exceptions, we simply use code here or does one need to do a chart review to get them?

MEMBER CHOTTINER: Chart review. It would be things like dental issues, allergic reactions.

CHAIR LUTZ: All right, have we answered questions sufficiently well to vote? All right.

MS. KHAN: We are voting on 1a on impact. You can go ahead and start.

So we have 11 high, five moderate, and one low.

And go ahead to performance gap. We have 13 high, three moderate, and one low.

And 1c, evidence. I think we are missing two people. So if you could just

enter your responses in again.

Fifteen yes and two no. And so we are going to go on to scientific acceptability.

So voting on reliability. So we are missing two people again. And we have nine high and eight moderate.

Voting on validity. You can go ahead and start. Six high and 11 moderate.

And usability. So we have seven high and ten moderate.

And feasibility. Okay, five high and 12 moderate.

And lastly we are voting on overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? And you can start now.

And we have 17 yes and zero no. So the measure will pass.

CHAIR LUTZ: All right. So the only other thing we have to do before we get to public comment I think it was 0378 was the

one we said that we were going to give them a few minutes to piece together a little bit more information to see if we could vote on it. That was the documentation if iron stores. And so I guess the question is whether the developers have had sufficient time to answer the question, a question which alludes might now. Does anyone else remember what we asked them for? Was that a timing issue?

MS. BOSSLEY: It was the definition. So have they defined the testing? Gene, can you pull up -- can you make that bigger so I can read it? My eyesight is pretty good but not that good.

So looked at bone marrow examination including iron stain or serum iron measurement. Where did the or fall in? And do you have the answer yet? No.

CHAIR LUTZ: It sound like they don't have the answer. So I guess we can hold out for a little longer. And then maybe

should we move on to the public comment for the morning?

MS. FRANKLIN: Yes.

CHAIR LUTZ: Okay.

MS. FRANKLIN: Nicole, could you please open it for public comment, open the lines?

OPERATOR: Certainly. For public comment, ladies and gentlemen, please press *1 at this time.

CHAIR LUTZ: Is there anybody there for public comment?

OPERATOR: We do have a couple people over the phone but no one has cued for public comment.

CHAIR LUTZ: Anyone else in the room with a comment?

OPERATOR: And I do apologize. We do have someone over the phone now. We have Charles Hampsey.

CHAIR LUTZ: Okay.

MR. HAMPSEY: My name is Charles

Hampsey. I am with Eisai and we are a member of the supplier counsel. I apologize. I'm going to turn down the echo on my computer.

My comments are specific to the palliation section.

CHAIR LUTZ: I'm sorry --

MR. HAMPSEY: Going back to the measure that looks at the percent of patients on chemotherapy for 14 days before death.

CHAIR LUTZ: I apologize to cut you off. We need exactly 24 hours probably in advance of when we would be able to --

MR. HAMPSEY: Oh, I'm sorry.

CHAIR LUTZ: No, you're fine. I understand. That is for a different day, as they say.

MR. HAMPSEY: I see. With that being said, the only other comment I had would be for later in the afternoon on the oncology measures. So I apologize.

CHAIR LUTZ: Oh, thank you.

OPERATOR: There is no other public

comment over the telephone.

CHAIR LUTZ: Any other comments before lunch? Let's do it.

(Whereupon at 1:08 p.m., a lunch recess was taken.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

(1:39 p.m.)

CHAIR LUTZ: The NQF was so proud of us this morning they thought they would throw a few wrinkles in this afternoon.

So the first thing is we do need, if we can, to go back to 0378. I think the sponsors of 0378 or the presenter has an update on the request we had for additional information and documentation of iron stores in patients receiving erythropoietin therapy. I think we just had one question.

DR. ADLER: To revisit the question of iron stores, for patients on erythropoietin, to revisit the narrated detail where it states that documentation of iron stores there was some discussion of how that should read would include either bone marrow examination, including iron stain or serum iron measurement like ferritin and serum iron TIBC.

So we are actually saying where it said or serum iron TIBC, make it and. Involve both of those measures of iron stores and see

if that would meet the needs of the group here.

CHAIR LUTZ: I see. So the folks that brought up the issue initially, does that sound like a reasonable way to make up the difference?

So I guess procedurally are we allowed to change that and then vote on the change or how do we work that? I mean, we have to vote on the whole thing but are we allowed to then read it as an and instead of an or?

MS. BOSSLEY: So I guess the one question I would have is how does this impact the reliability and the validity of the measure? And I think we would need to have you provide that back and then the committee can tell me if I am wrong.

Then we can have the committee revisit that. But because you are, from the sounds of it, combining the two, I actually think it will change your information on the

performance score, as well as the reliability and validity, I would assume. It actually, I would assume, would lower performance.

MS. TIERNEY: Just to clarify, there were two options for meeting the measure. So there still would be two but the second one I think would be clarified a little bit more but certainly we can look back at our testing data to help explain how we believe the minor change would still be supported by the testing data that we have completed.

MS. BOSSLEY: Maybe the best thing to do is we will talk to PCPI and have them bring it back to the committee. And we may be able to do this either through a quick phone call or by email. Does that sound good to everyone?

CHAIR LUTZ: All right and as we head on to the oncology measures, unless someone has a reason to suggest otherwise, the point was made that they are sort of in reverse order 0381, 3, 4, and 6 as to what

would happen in real life. The suggestion was made if we start with 0386, which is staging and then quantifying pain and caring for pain and treatment summary, if we work backwards. So if we started with 0386 and worked backwards, chronologically that makes more sense. If that is okay.

So we would go through 0386 first and I believe that is the cancer stage documented and I think AMA also has this one to introduce.

DR. HAYMAN: My name is Jim Hayman. I am a radiation oncologist at the University of Michigan. By way of introduction, I have also co-chaired the oncology workgroup that developed these measures. I am the chair of ASTRO's Clinical Affairs and Quality Committee and also serve on ASCO's quality of care committee.

And I am here with Emily Wilson, ASTRO staff, and Kristen McNiff from ASCO.

Just as way of background, I think

I will give a little bit of background because all of the measures we are going to talk about in the next set came from this workgroup and then I will talk specifically to the measure.

So the AMA convened an oncology workgroup with representation from ASCO and ASTRO. That was in 2007. We had 30 members, about a third of which were radiation oncologists, a third medical oncologists, and a third were from some of the surgical specialties, nursing, a patient representative and so forth.

We developed a set of measures which were approved by PCPI in October of 2007 and were endorsed with time-limited endorsement in 2008. And so we are going for maintenance endorsement today.

Several of the measures have been used in CMS PQRS program and also in ASTRO and ASCO's practice improvement programs.

So the first measure that we are going to talk about today is 0386, which is

cancer stage documented. This is a measure where the denominator is all patients with breast or colorectal cancer and the numerator is patients who have a baseline AJCC cancer stage or documentation that the cancer was metastatic at least once during the 12-month reporting period. And like some of the other measures, this measure was specked out initially for PQRS, which requires yearly reporting.

So in terms of the issue of importance to measure and report, I think it is probably obvious to hopefully everyone in the room that this is a high-impact topic, given the number of patients. We are talking about hundreds of thousands of patients a year who are diagnosed in the U.S. with breast and colorectal cancer.

In terms of demonstrated opportunities for improvement, this measure has been included with slight modification in ASCO's Quality Oncology Practice Initiative of

QOPI. And it is important to realize these are self-selected practices. So I think that this is probably the upper range of what one might expect, 84 percent of practices were reporting on this measure. Within ASTRO's quality oncology -- sorry, ASTRO's quality improvement program, the rate of performance was 87 percent with a range of ten to 100 percent. And there is also a study that was published recently in the literature for colorectal cancer demonstrating only 40 percent of patients having reporting TNM stage.

Lastly, I want to talk to the issue of the quality, quantity, and consistency of the body of evidence. As I am sure is obvious to a lot of people and true for many oncology process of care issues, there aren't any randomized controlled trials that address this issue. And so we don't have a strong evidence base as defined by NQF. However, there is a consensus-based guideline

from the NCCN which recommends both for breast and for colorectal cancer that the patients stage be documented. And I would ask that you consider this measure for an exemption in terms of the quality of the evidence, given the fact that the potential benefit to patients clearly outweighs any risk of harm.

I would be happy to answer any questions you might have. I don't know if Emily, or Chris, or Sam have anything else to add.

CHAIR LUTZ: Okay and I think Dr. Pfister was going to be our primary discussant.

MEMBER PFISTER: So I found out this morning I was going to be the discussant. So and then I found out you reversed order just after lunch. So, it is sort of like I thought that I actually was going to prepare this kind of insidiously while Dr. Malin was presenting her data.

So I think that the prior

presentation makes my job a lot easier. You know I think in terms of to summarize the discussions at the subgroup level, I think that there is pretty much higher moderate sort of agreement that it was important. I think that the data is heavily weighted toward that it is well-documented to be associated with prognosis. The data is not so well-documented that let's say if you put in your note it is T this, N this, M0, versus let's say it is local regionally advanced and they necessarily end up at a better outcome.

But I think as far as the importance in evidence like I think that there wasn't a lot of concern about the -- You know, it wasn't above some bar.

What was actually again a little surprising to me after I got over the surprise on the call that Steve Edge wasn't going to discuss it, that I was going to, was that the reliability discussion actually generated a lot of back and forth. Actually a lot more

than I guess that I would have perhaps predicted ahead of time. And I think that -- and actually I have had a chance that I should kind of digest this a little bit since then because it was, as you will see in this summary sheet there that if you look at the proposer's submission, they talk about sort of I think clinical staging for breast, pathologic staging for colorectal. You know, there are some concerns about some of the issues in terms of being able to sort of keep that straight in the way that they would look to extract this information.

There was also issues related to how easily or accurately you would be able to get that information. I guess after I sort of digested this a little bit that the problem with the developer is really what you do is you are thinking about doing any staging and that perhaps it is really the most accurate consensus-driven staging is maybe a secondary concern. It is sort of like you are thinking

that attempting to fill in that TN&M box in some way is, in and of itself, an important thing to try to leverage that behavior. Although I think that there are just looking at a place where I routinely see staging done by me, a range of oncologist and surgeons, it is always striking me how commonly the staging there is a lot of inner-observed variability.

As far as usability, I think that again a bit surprise to me actually that this sort of ended up kind of I think at best in the intermediate category. And also I think that there is a certain sense that well it can't hurt but I guess that the two issues that came up in our discussions is that when you leverage people to get a stage in there and people tend to preprogram their notes now and they keep, like they keep perpetuating the wrong stage. And so again, I could see how that could be a harm.

The other thing is that again it is always striking to me the patient's

perception of stage. So for example, there is a certain feeling that Stage IV lung cancer is the same as Stage IV head and neck cancer. And so I will get these calls afterwards when I say oh gee you have Stage IV head and neck cancer and we cure most people like you, they are NED. And then I get this frantic call from Martha's Vineyard. I'm here with my brother-in-law and he said you are lying to me.

And so I think similarly how patients process stage when it is out for public reporting is something that while it may be again we want to leverage that information as something they should know, I think that if you look at certainly end of life care, that there is a lack of sort of it being direct as we should be certainly. But I do think that there are potential things that could happen that need to be considered.

With regard to feasibility again I think in terms of again probably felt it would

be doable but then sort of then again the moderate group.

And then I think in terms of the preliminary assessment for endorsement there is actually a split in the group. Again, it surprised me a little bit, three to two.

CHAIR LUTZ: So is there anyone else that was on that phone call in that group who want to share a little bit more of their thoughts?

So you are saying there was two that said they didn't want to pass it on. Is there anyone here that wants to admit that they were one of those two and tell us what they --

MEMBER FIELDS: I wasn't on the committee but I guess the real concern is it is Mom and apple pie that we should stage patients appropriately.

But again, when you try to get to the quality endpoint, was the pathology correct in the first place? Were the

measurements correct and consistent? So I could see shy people might say are we getting to a quality measure. But if you are talking about that much of a gap and we aren't even documenting it adequately, then I think that has to be step one. And then we have to figure out in the next iteration how do we get to quality.

And it would be interesting to hear what the rest of the discussions were about, if it was about that topic or we just didn't think it would change outcomes at all.

MEMBER PFISTER: A lot of the supporting data really had to do with again we do this all the time as sort of it really doesn't really weigh in on is do the folks actually do better. I think that it sort of makes sense that they should do better. But again, I think we deal with this in our specialty all the time.

Again, just to give you a contrary argument, you might say that if you know -- if

you look at the NCCN guidelines for example, they will take, you know, there may be 17 different TNM combinations for a different disease. But that may ultimately boil down to three different pathways. And if let's say if I said well you are in group one, group two, and group three, that is not TNM staging. You could end up in a worse place than if let's say I went down to all these individual categories.

It is certainly, in terms of analyzing the data, assuming it is accurate, it is going to be much better. It should be prognostically significant. There is a difference, you would think, between a Stage IV that is like, you know, that is kind of clearly in the IV-A as opposed to being more advanced.

You know but in terms of the link with outcome, clearly stage is associated with survival. It is how you quantitate disease extent as a way to improve outcome. And I

think that the data that is presented really has to do with that equally impacts on prognosis.

MEMBER ROSS: I mean certainly no one can sit here and say we shouldn't stage the patients. It is actually surprising to me that the number is so low. But I am not quite sure what we are getting at with this. I mean, are we really putting the burden now on each of the health systems that employs or gives privileges to those physicians to insist? Because it doesn't seem like this will change behavior.

The real issue here is not is the physician documenting the stage correctly but is that physician offering that patient the right therapy. Is the patient in the right place?

And I'm not sure how the documentation gets to the idea of what the next level of intervention is. So I don't understand this really as a quality outcome

for each individual patient. I agree it might help us retrospectively in outcome studies when we go back.

CHAIR LUTZ: Do you have a statement?

DR. HAYMAN: Would it be okay if I speak to that?

I mean I think it is critical. I mean how can you have an intelligent conversation with a patient if you haven't gone through the intellectual exercise in your own mind of assigning the patient a stage category to talk with them about their prognosis? How can you think about what might be the best treatment for that treatment, unless you consider that and also documented that?

So I mean, documentation isn't sufficient to lead to those sorts of discussions and decisions but I think it is an important step. And I think it is proximal again, based on some of the discussions we

were having this morning to better outcomes.

So is it a standard thing that should be done? Yes, but it is not being done. And I would argue that this is a first step.

MEMBER ROSS: But your assumption is is that if you get a physician who is currently not staging the patient, or at least documenting that he or she is staging the patient, your assumption is that if they have to write it in the charge, that they will have that intelligent conversation about that TNM stage with the patient. I am going to be the negative on that and say they are going to have their nurse practitioner document it in the chart and the conversation with the patient probably will never change.

So I just don't see how this changes physician behavior in that 17 percent or whatever it is that currently doesn't do it. I may be wrong about the whole thing.

CHAIR LUTZ: I'm sorry. Dr.

Alvarnas, you were going to say something?

MEMBER ALVARNAS: I think one of the things of concern, I mean, we all want to ensure that patients get access to what represents the optimal care for their particular disease but I think to try to tackle everything at once makes it unmeasurable.

So I am a systems guy from a manufacturing point of view and I think breaking down processes into its granular constituent parts that are in fact measurable, gives us a place to start and I think if somebody staged their patient or not staged their patient, that is one part of this process that we hope culminates in superior care but it is actually measurable and something that you can look at in a very granular way.

And I think ultimately over time what we would hope comes through this system of quality measures is that you have a whole

tapestry of metrics that relate to each other in a continuum but I would most certainly see this as an important first point in the measurement process.

CHAIR LUTZ: Jennifer, I think we --

MEMBER MALIN: I think I just want a clarification on the measure because I think I may have misunderstood it in the past.

So it looks like this is not limited to newly diagnosed patients. It is any patient with these cancers seen during that year. And then they should have a documentation of stage at any point in time prior. Is that the way the measure gets?

So I guess it is two things. It is just it is a survivorship population mostly. Most of these people survive and in any given practice you tend to have a lot more survivors than you do new diagnoses.

MS. FRANKLIN: Can the developer speak to that?

MS. MC NIFF: Okay, you are interpreting it correctly that it would be any cancer diagnosis. And the argument of the workgroup or the thought of the workgroup was that if a patient is coming into an oncology provider, their stage at diagnosis should be documented, even if they are several years later down the path into their cancer survivorship.

So the intention is stage at diagnosis. There is actually on specification not a distinction about clinical or pathologic stage. I think there is a bit of confusion there so it is not different for breast and colorectal cancer. But stage of diagnosis is document, regardless of where the patient is in their disease path.

MEMBER TAPAY: And then if it is progressed, you would have both in the record or not? That is the point of documentation.

MS. MC NIFF: This doesn't assess whether you look for -- whether there is

documentation of the patient's current disease status. It's looking at stage at diagnosis.

CHAIR LUTZ: Karen, I think you are --

MEMBER FIELDS: So I guess to reiterate, I think like you said, if we aren't even doing the fundamental documentation then we have a huge problem. But in both of these diseases that they chose, treatment is -- patients are stratified and treatments are different. So it is not just for prognostic indications for you patients.

And quite honestly, it is a way to get to overuse and under use of treatment as well. In breast we know that now we actually give more limited radiation. We give more limited, less chemotherapy in the very early stages. So I think that the developers chose two diseases where there actually has been some dynamic changes over the last couple of years as far as how we would stratify them for treatment.

It is just unfortunate. I mean, the hard part is a measure still -- none of the measures we are going to talk about today can really get to that real quality like maybe we document it but what other interventions happened. And I don't know where we -- I assume that when we develop these measures we have to start somewhere. That is my only observation.

CHAIR LUTZ: All right, Larry I think you were next.

MEMBER MARKS: I think this is a central component of doing our jobs right. I mean so much, so many of the things we discussed this morning was here is an idea but it applies only to this stage. Well if the stage isn't documented or isn't considered, it is a problem. And you are right, the physician who is primarily taking care of the patient may know what the stage is when they see their patient. But think of all the times the patient was up at in the ER or at their

primary care doctor and they pull out the note and they don't know what the stage is. They make decisions without having all the information.

So I think it is really central and a bunch of the things we will talk about later, this afternoon and tomorrow, you know, prostate cancer, bone scan yes or no, particular stages. You know, 3D radiation therapy in particular situations dictated by the stage. And NCCN is all over this document as justification for a lot of these guidelines. So it seems to me and that is all based on the stage on as well. So I think this is very fundamental to the point almost that we could consider why limited to only breast and colorectal. This might be just as valid in all or many diseases.

CHAIR LUTZ: Elizabeth, I think you were --

MEMBER HAMMOND: Yes, that is one point that I would like to make. There is a

lot of data that staging is critical in all kinds of cancers. And it does define treatment for many of these stages.

I think you really have to admit -- I admit that I am a pathologist but it is very critical that you have the pathologic stage of disease. Pathologists across the country are uniformly being told that they must do this. This is a never event for them. They aren't all doing it and so this is a very important thing. I think it should be a measure for all cancers and it should be pathologic stage as well as clinical stage but for sure pathologic stage because that is the only time you really know what is tumor and what is not. So I think it is a very important measure and it represent sort of the floor before we can go farther.

CHAIR LUTZ: Well, I'm sorry, Dr. Loy, I will get you next.

But I was going to say in answer to what you said, Dr. Ross, you and I practice

in a state where technically legally every outpatient center is supposed to have a stage on the chart if they are treating cancer. And very few do.

I can go on the less cynical end, though. I think people that are practicing well, oftentimes by virtue of having to stage, start to realize wait pathologically I still have a question about that and I need to communicate with so and so. So it is almost like the internet. It can be used by bad people for bad things. It is used by good doctors. So I see your point but on the other hand I think it almost has to be a baseline.

MEMBER LOY: And I'm recalling our discussions when we were having the small workgroup and I believe one of the things that we were faced with is what was already mentioned earlier today I believe by Dr. Miller and that is that we were presented with evidence from NCCN which, by definition was characterized as 2A or low level and we really

didn't have a good sense of so how stringent do we need to hold to these criteria in here or do we need to be considering now that we know we have got an exception, considering an exception. Because we really didn't have an outcome to link this to to be able to say this evidence supports the use of this as an indicator.

I don't think anyone would disagree, I don't want to put words in the small workgroup's mouths but I for one as part of that committee, would say that is a desirable first step as already has been mentioned. But I don't think if we were held to the standard of saying the evidence supports that. I don't think we have that evidence.

MEMBER PFISTER: You know, maybe it was one of the last measures we got to that day so maybe we are getting cranky but you know, we sort of breezed through the importance thing and so forth. And then I was

just looking at what was submitted by the proposers. You know, what it is is again when you look at the quote from the NCCN guidelines and then it started coming up about this pathologic versus clinical staging. And then it seemed that the specification, the metrics seemed to be, well it didn't really matter, just something.

Then you got down to well gee how are they going to electronically get this. And then it went down to and then how accurate it is going to be. And you know, then the data issue came up. And it is sort of like it is again in these metrics, there are lots of things that make total sense. These are clearly part of what we do. And I think it is how well they plug into this framework which I think provides a framework for this discussion but doesn't necessarily fill in the holes of missing information, which are often, I think, homogenized when we do our discussions when we vote, not that we necessarily got any more

evidence to fulfill that criteria.

So there was little question that it was a very basic thing to do. It is just that there were other issues that kind of came up, depending on how rigorous you wanted to be about the other criteria that we needed to apply.

MEMBER FIELDS: So most of the data, though, is going to always, I mean, there is no way to do prospective randomized trials about whether you staged or didn't stage a patient. So by the nature of just this very fundamental how do you document something, it is only always going to be retrospective data. But then there is still retrospective data about outcomes were different based on stage.

So I think -- I don't think that in a measure like this we could ever have prospective randomized trial data but that doesn't mean that there is not tons and tons and tons of retrospective data that still

gives you some quality and some benchmark to start with.

CHAIR LUTZ: I think Dr. Gore was waiting.

MEMBER GORE: My only comment was just speaking to the importance of this measure is that often times non-surgically treated patients who are clinically staged, these data also populate cancer registries which are an important source of quality work. So even if you can't make a direct link between that doctor documenting a stage and how they interact with that patient, that data populates cancer registries which do tend to have unknown stage listed for up to 20 percent of the patients which kind of corroborates the performance data that they contribute. And I think highlights another role that this measure plays in just kind of the broader quality care agenda.

CHAIR LUTZ: Elizabeth was next.

MEMBER HAMMOND: Yes, to answer

the question about the presence of this information, the accessibility of it, it is a standard of practice for all cancer reporting that has been made by the College of American Pathologists in their cancer protocols. In the Commission on Cancer it has been made as a requirement for the documentation of a hospital getting a cancer accreditation status. It is required for the National Cancer Institute Qualification for Cancer Centers that the stage be documented. And it is in all pathology reports. It is supposed to be in all pathology reports.

Cancer reporting is also required across all the United States by the tumor registries and they prefer to get the information as information that is directly recorded as being the T&M stage.

So I think that it is an accessible measure and it is not being done and we really need it to be done, again, for all cancers, not just these two.

CHAIR LUTZ: Nichole, I think you may have been next.

MEMBER TAPAY: Thanks. I just wanted to echo a lot of the supportive comments that have been made and just add that from a patient and survivor's perspective, the time of diagnosis is when you become a survivor. It is also the time as someone who lost her mom to ovarian cancer and although this is not for ovarian, I can speak to if you don't know the stage, you don't know when you might want to ask for a second opinion. It is an incredibly disempowering moment. Some people do have somebody with them. Sometimes they don't.

And so in addition to the broader outcome and study issues that are there, there is the personal outcome that can be especially critically and highly metastatic types of cancer and so I would just concur that it is important.

CHAIR LUTZ: And Heidi?

MEMBER DONOVAN: So I just wanted to speak a little bit to the discrepancy in the scores and sort of where that came from on my end because I think as a new reviewer here, I came to the initial discussion taking a very narrow view of evidence, focused really very, very specifically on the measure, as well as the reliability and validity speaking to the measure specifically.

So I was a very tough scorer on all of these. And I am reassured to see the discussion that goes on within the group to talk about let's talk about how we can broaden this out when we don't have those kinds of direct relationships that we are looking to see in the evidence that is provided by the measure sponsors.

That being said I think we constantly need to remind ourselves that as these become endorsed measures and are in practice for a period of time, that it is important that we begin to draw relationships

between the specific measure and other measures of quality outcomes. And I think one of my, one of the things I was disappointed with as we were going through this is that the reliability and validity measures, especially the validity measures were really, really still depending purely on face validity. And these measures have now been in practice for three years and it is time that we started to see whether or not a measure like staging is associated with important outcomes. So is it associated with appropriate treatment for a patient? And I think that we need to start seeing that.

So I am still willing to say that there is enough gap in performance at this point that we ought to keep documenting this. But I don't think the next time this comes around we should be able to say well it is really important that everybody has staging because we will have the data to move on.

CHAIR LUTZ: Bryan.

MEMBER LOY: And I think she just nailed the dilemma that we had in our workgroup and that was is that I think there is agreement around the importance of the measure and that it was essential as a first step, as has already been stated. But in terms of being able to link it to a quality outcome and having the data there to be able to assess the criteria that we were asked to, we could, many of us could not make that claim.

And I don't know where that would lead us as a voting member. Because I think at some level, we have to at least understand what our limitations are in our vote versus whether or not we have to have an exception-based process.

MEMBER PFISTER: Again, I think you know Heidi shared her -- Just to get through the spectrum of comments, the other reviewer is not here, but I will summarize kind of the flavor. Let me just check to see.

Yes, suitable for endorsement, no. But it was that the comment was, problems with specification measure no information on impact that document a stage improved outcomes compared with assessment whether the care versus if the care would be a way to assess whether appropriate care was done for the staging. They were pessimistic about how easily it would be obtainable electronically with the potential need of chart review.

And those are a couple of highlights. They acknowledged that it certainly made sense that this would be an important thing to do.

MS. FRANKLIN: Can the developer speak to the comments about the importance?

Do we want to have any response from the developer on the issues around the evidence.

MS. TIERNEY: So I would say that the information that we included in the submission forms to support the measure comes

from clinical practice guidelines and specifically the NCCN guidelines. It is typical of our methodology to use clinical practice guidelines to support the development of measures. So we have provided the documentation available to us from the NCCN about the quantity, quality and consistency, which was admittedly limited but we tried to include that. And I think that the NCCN guidelines do mention in the verbatim statement they do have some mention of the link to the outcome, particularly for patients with breast cancer. So I think there is some evidence there that was included in the guidelines. I don't know if anyone has anything else to add.

DR. HAYMAN: I would just add just to echo I guess my opening statement that I think that the potential benefit outweighs the harm and I think that this is a situation where one has to have an exception to the quality and quantity of the evidence. It just

seems appropriate to me.

CHAIR LUTZ: Bryan, did you have anything else? You're fine. I was just making sure we didn't skip you.

Yes, Larry?

MEMBER MARKS: Maybe it's a bad analogy but if we systematically didn't have the right sex and age of the patient in the chart, we would say gee, that is malpractice and this is not that different.

Yes, the data is there but we are seeing a follow-up patient, seeing a patient and you don't easily have the stage, you are wasting time. You are looking through the chart figuring out what the situation is and then maybe you are making a right or wrong decision. So I think it is sort of it is a vital sign almost.

MEMBER MILLER: This is just a general question about going back to the question of clinical versus pathologic staging. You know, analogy is even though the

asterisk says stage is the stage at diagnosis, I wonder how often that is misinterpreted and whether we have, and this is maybe a question for the developers, but whether we have any information about is the AJCC stage truly the AJCC stage that is listed in a lot of these reports. Because not infrequently I see patients and I will say from my own institution sometimes that clearly it is not Stage IV breast cancer. It didn't start as Stage IV breast cancer. It started as Stage I breast cancer.

And so I worry. It goes to the reliability question which I guess we will get to. I'm not saying this is the deal breaker but I worry a little bit about that. And I just didn't know if anyone had any additional info on that.

MEMBER HAMMOND: That is one of the big problems with all cancer reporting and we are actually working on that in reporting groups across the country in pathology because

you need to have summary information ultimately on patients. And how do you get that and deliver that to the clinician in such a way that they can understand what is going on. So that is a big problem.

The stage migrates with time. And because we don't have an integrated system of data gathering that we can't really always do that. So typically it needs a stage at diagnosis but there is a real effort going on to summarize or integrate all cancer reporting ultimately. We are not there yet. We are just beginning that journey.

CHAIR LUTZ: Yes, Heidi?

MEMBER DONOVAN: So I think maybe to clarify also I don't think you are asking for an exception really. I mean to say that he evidence isn't there and that we want to make an exception. I mean for me what is sort of clarified is that the evidence exists at this problem -- well that accurate staging and treatment by accurate staging has a tremendous

impact on outcomes. The question is whether this is a valid measure. And I think we have to separate those two things.

I mean, from the discussion and sort of thinking more broadly about evidence not just around the specific measure but the question of staging, to me it feels like the evidence is there but the question about validity still remains. The question is, has it been out there long enough for us to understand it?

DR. HAYMAN: You know, my understanding is that NQF has definitions to rate the quality of the evidence. And to have the evidence be rated highly, you need to have multiple randomized controlled clinical trials. And we don't have --

But you have to have evidence. Right? And what we have is a consensus-based guideline. And so you know, to meet that maybe I'm not understanding your process but I am not here to argue semantics but when I

looked at your guidance on evaluating evidence, the NQF publication on this, there is a rating of evidence and that does not rate a consensus-based guideline highly.

But I still think that this is very important. And so that is why I brought that up.

CHAIR LUTZ: Bryan, I think you were --

MEMBER LOY: Yes, I was just prompted to think about yet another issue. If we look up two years, three years from now and we have somehow gained ground in meeting the measure, meaning we have improved the documentation of getting the stage, then I guess I would ask myself what will we do with that information? Will we be confident that the stage has been accurately documented to the extent that we would say we moved the quality needle in the right way. And I don't know that I could answer that question. I guess it goes back to Dr. Miller's comment

about reliability. But what to do with that information I think still remains somewhat of a question in my mind.

Any thoughts that you care to share around that concern?

DR. HAYMAN: You know, I guess this is a measure for public reporting. Right? So your health system isn't documenting what the patient's stage was at diagnosis. And the hospital down the street, you are getting it at 50 percent. The hospital down the street is getting 100 percent. You know, if you had to decide where to send your mother, which hospital would you recommend she go to.

So I think there is, and this speaks I guess to the issue of usability, is this data usable to patients, the payers? I think it is.

CHAIR LUTZ: Elizabeth, I think you were next. Do you still -- okay. Heidi? Heidi do you have anything else? That's

okay. I don't want to skip anyone.

MEMBER ALVARNAS: I guess we are confusing or at least I think we are confusing two issues, which is one whether or not a metric is worth measuring and whether or not it is granular. And then the whole other thing is the strategic plan for how you use these metrics to forward the care of patients. And I hate to so load a metric having to carry the weight of a strategic plan for advancing the state of the art that we sink a metric that is good.

I mean I think this is actually a useful metric. It doesn't answer every question. It doesn't guarantee that somebody is going to get optimal care but it provides us with a starting point, I think.

As a strategic plan you would like to build upon these three years. Look at the data in a really rigorous fashion to figure out what the next set of metrics that advance the state of the art are. But that is beyond

the scope of this particular metric. And again, I would hate to weigh down this discussion having to come up with the whole strategic plan aspect.

CHAIR LUTZ: Well and unless someone disagrees, I would say after a very healthy discussion, I mean they gave us this nifty little voting tools, we could always go ahead and see what is what, if you guys are okay to move ahead.

MS. KHAN: So 1a on impact. Oh, no. One sec. Okay, you can go ahead.

So 14 high, two moderate, and one low.

1b, performance gap. You can go ahead. We have 13 high and four moderate.

And rating the evidence at 1c. You can go ahead. I think we are missing -- oh, there we go. So we have 12 yes, two no, and three insufficient evidence.

So we are going to go on to scientific acceptability.

Looking at reliability. We have five high, nine moderate, one low, and two insufficient.

And moving on to validity. So we are missing one person. If you all could just enter them again. Oh, there we go. And we have two high, 13 moderate, one low, and one insufficient.

And usability. Ten high and seven moderate.

And going on to feasibility. We have seven high, nine moderate, one low.

And overall suitability for endorsement, does this measure meet NQF criteria for endorsement?

So we have 17 yeses.

CHAIR LUTZ: All right, so if we continue in our reverse order of the oncology measures, I think the next would be 0384, which is oncology pain intensity quantified. I think it is also an AMA presentation and Dr. Pfister is the first discussant.

MEMBER ALVARNAS: So this is 0384. This measure is also from the AMA-PCPI ASCO ASTRO oncology workgroup.

The denominator for this measure is all patients with a diagnosis of cancer who are receiving chemotherapy or radiation therapy. So we just focused on patients who are under treatment. And the numerator is a patient visit in which pain intensity if quantified. And we left that sort of a little bit open-ended in terms of how that could be quantified, either using a zero to 10 scale, a categorical scale or a pictorial scale.

In terms of importance to measure and report, I think it is pretty obvious that this is a high impact area, given probably again, oh I don't know, it would probably be a million patients maybe each year who are undergoing treatment with chemotherapy or radiation therapy who have cancer in the U.S.

In terms of opportunities for improvement, this is a measure that is

included in ASCO's QOPI program, so that is a practice improvement program. They have one component where they ask is pain intensity quantified by the second office visit with a performance rate of 87 percent with a range of 23 percent to 100 percent, and that is looking at over 21,000 patients.

Again in -- Oh, I'm sorry. It was also included as part of ASTRO's PAAROT program. Again, that is another practice improvement program with a lower performance rate of I'm sad to say 57 percent. And this measure has also been part of the PQRS program and the performance rate in 2009 which as Samantha said earlier is the only year that we have data available for, was 67 percent. And unfortunately, they don't provide us any information about the variability.

In terms of again getting to the issue of the available body of evidence, again there are no randomized controlled trials looking at quantification of pain during

treatment. And so this measure is based on two consensus guidelines, one from the NCCN and the other from the American Pain Society. And they are consistent in their recommendation that pain be quantified as part of routine care.

So again this is a situation where I think the potential benefit to patients being asked if they have pain and not only if they have pain but quantifying that pain clearly outweighs, the benefits clearly outweigh the harm.

And so we would ask or recommend that you endorse this measure. I don't know if anyone else has anything else to add.

CHAIR LUTZ: Dr. Pfister?

MEMBER PFISTER: So I think that from importance point of view, I certainly think there was agreement among the group that it was moderate or higher. I think that there are gaps in the evidence, as Jim noted.

With regard to reliability, again

it was felt ultimately to be the majority felt that it was moderate or higher. Again, because it sounds like you just have to use a scale but not being that exclusive about what that scale is, that might have some bearing on -- if you are a proponent, you will say it is the first step. If you are looking to be critical, you would say trying to do comparisons, you need to have some harmonization there to sort of fully and reproducibly see what impact you are having.

As far as usability, again most felt it was moderate or higher. Feasibility moderate or higher. And the majority of the subgroup recommended endorsement.

CHAIR LUTZ: Anybody else from the working group that dealt with that?

All right, anybody in general? I'm sorry.

MEMBER RICCIARDI: It seems like an important measure but one wonders if there is any association between measuring pain and

actually changing pain management. Is there any outcome associated with measuring that process measure?

MS. FRANKLIN: Yes, developer?

MS. MC NIFF: I would point out that this is paired with the next measure we will talk about which has to do with a plan of care for pain. So you must report on both of them together.

CHAIR LUTZ: Does that make sense? Okay. Any other questions? I'm sorry, Karen?

MEMBER FIELDS: Just a comment. This was actually one of the few guidelines or measures that we saw that actually noted literature to support a disparity in access for the patients, which obviously should be one of the focuses of improving measures and measuring quality.

CHAIR LUTZ: Okay, good points. Anyone else?

Moving on to voting that quickly?

All right.

MS. KHAN: So 1a on impact. We have 16 high and one moderate.

Looking at performance gap. Eleven high and six moderate.

Rating the evidence. I think we are missing one person. So we have 16 yes and one no.

So we are moving on to scientific acceptability. There are seven high and ten moderate.

And looking at validity. We have six high and 11 moderate.

And moving on to usability. We have one person missing. So we have ten high and seven moderate.

And feasibility. We have nine high and eight moderate.

And overall suitability for endorsement. Does the measure meet NQF criteria for endorsement?

We have 17 yes. The measure will

pass.

DR. BURSTIN: Just one comment, there were two pain measures that recently went through our palliative care project about pain assessment and pain screening. So we will bring that for your discussion tomorrow because granted the patient population may be slightly different but the harmonization should at least be done in a standardized way.

CHAIR LUTZ: All right, so if we continue next will be the paired pain, it is basically plan of care for pain, also an AMA and then I think Jennifer will be discussing after they give us the setup.

DR. HAYMAN: So this is 0383 and I apologize. This is a paired measure. I didn't mention that earlier.

Again from the oncology workgroup, this measure had been endorsed in 2008. The denominator for this measure is all visits for all patients with a diagnosis of cancer who are receiving chemotherapy or radiation

therapy and report having pain. And then the numerator statement, to be in the numerator patient, the patient visit must have a documented plan of care to address pain. And that plan of care can include prescribing opioids or non-opioid analgesics, psychological support, patient and/or family education, referral to a pain clinic or something as simple as reassessment of pain at an appropriate time interval.

I want to point out that when the workgroup was developing this measure we had a lot of discussion about whether the denominator should include patients who report any pain or patients who report say moderate or severe pain. And the feeling was that the consensus was to be more comprehensive than not because of the fact that the range of options in terms of a plan of care is quite broad. So if someone has mild pain and they are undergoing treatment with chemotherapy and radiation therapy, the next time you see them,

you know, the plan could be to reassess at the next time you see them.

And so that was I think, and Kristen could give her impression as well. So that was why the decision was made to go in that direction.

In terms of the issue of impact, I think we would all agree again for the reasons that I mentioned earlier that this is a high impact area.

In terms of opportunity for improvement, there is a slight modification of this measure that is part of ASCO's QOPI program and in that setting, the performance was 78 percent with a range of 12 percent to 100 percent. So a pretty wide range.

I'm embarrassed to say that for radiation oncology, we are again behind our colleagues in medical oncology so we had a performance rate of 61 percent with zero to 100 and then in PQRS in 2009 the performance rate was 91 percent.

And again just to speak to the quality of the body of evidence, again, this is a process of care issue where there aren't any randomized trials. And so again this is a measure that is based on consensus-based b guidelines from both NCCN and the American Pain Society.

I want to emphasize, too, that the NCCN guidelines also address the issue of mild pain. So again, that was justification for including those patients in the denominator.

These two guidelines are consistent in their recommendation for developing a care plan for pain. And again, this is a situation where we think that potential benefit to patients clearly outweighs the harm.

So we would recommend endorsement. Thank you.

CHAIR LUTZ: Jennifer, what did you and the smaller group think?

MEMBER MALIN: Sure, we had a

pretty engaged discussion on this one in our group. And if you look over the summary sheet you will see that I think the ratings were pretty diverse, which reflects that discussion.

I think the concerns that were raised about this measure, you know, there was a whole-hearted endorsement of the importance and the impact. I don't think there was any question with that. The concerns were raised because the denominator includes all patients, even if they have a pain score of one, you know, mild headache when they are talking to the nurse and they report one. And you as a physician talk to them about it. It turns out it wasn't a big deal. That would still, at least according to the measure specs, require a plan of care.

And then secondly, the numerator is equally broad. So the way it is described, if someone who has had severe or uncontrolled pain for three weeks, documenting that you are

going to reassess pain in the next visit, should pass the measure specification.

So there was a lot of concern expressed just about the breadth of this particular specification of this measure.

CHAIR LUTZ: Anybody from the working or the small group have anything to add? Anyone in the bigger picture, bigger group?

MEMBER MARKS: Can you clarify what it means to address the pain? I forget what you call the -- I mean, how broad is that and how do you score that?

Do you say patient has mild pain in your subjective section and then down in assessment and plan, mild pain, comma, follow-up. Would that be in the realm of acceptable the way you capture it?

DR. HAYMAN: Yes.

MEMBER FIELDS: So can you go over again the groups, the authors' groups discussions about why to do all levels of pain

so that we can understand it again? Because even the QOPI measures when we respond to those, it is must moderate and severe pain and they needed a pain intervention.

And I don't think anybody disagrees with it but certainly the problem is different providers might be getting the information and interpreting the information and the physician provider is the one that is responsible for the information and coming up with the pain plan.

So just summarize again for us why we chose all levels of pain.

DR. HAYMAN: So you are really, even though I consider myself relatively young, you are challenging the capacity of my memory to think back five years ago in terms of those discussions.

But I think it was basically this idea that any pain potentially for someone who is under treatment -- or this is just limited to patients who are under treatment. So we

are not talking about follow-up. We are not talking about consults. We are talking about patients that are actively under treatment, that any pain that they might be experiencing is worthy of consideration in those specific circumstances.

If I remember correctly, we had some members on the committee on our workgroup who had expertise in palliative care and symptom management and they felt strongly about that. And so we were trying to be respectful is my recollection of their opinion.

You know I think the point is well taken that maybe it is not unreasonable to consider limiting this to a certain group of patients and that was the direction that ASCO chose to go in for their quality improvement program. But I think that was the rationale. I don't know if Kristen or Emily or Sam if you remember anything else.

MS. MC NIFF: Well I would just

add to that I mean we actually the measure in QOPI predated the specific specifications for the PCPI group and there were, I mean, 40 people or something involved in this discussion, a huge number of people, and they were able to definitely argue persuasively and convince their colleagues that this should be broadened out to any patient who reports any pain whatsoever and that would be the best denominator for the measure.

So I mean, hours of conversation about this. It was not a quick thing. And ultimately the group's consensus was to use the broader.

MEMBER FIELDS: See, I don't disagree at all that we should always try to intervene and treat it appropriately but I guess the way some of the ones that we are going to review tomorrow described this was it is not an always or -- You know, it is intended to move toward perfection rather than 100 percent compliance with that. And maybe

that is the statement that needs to be in it. Because we can't set ourselves up for something that is impossible if we are talking about I stubbed my toe on the way in. And I'm not suggesting it would be that trivial. I completely agree that we need to address it. It is just that sometimes we are going to be asking the providers to do more documentation about minor problems and is quality going to go up on the lower level. And I would love to hear Dr. Bruera's comments.

MEMBER BRUERA: Thanks very much. I think to a certain degree these perfect some of the NCCN previous errors that some people might have concern about because there was this pain more than seven. You have to admit the patient to the hospital, put an IV on them. And a lot of pain VII patients are golfing so they say, you know, after I finish golfing you admit me and put me on IV opioids because it was a bit of an over-managed process. In this case, your action plan might

simply be I'm going to talk to this patient. I am going to counsel this patient. There is a plan to deal with this and that might be perfect. It is just the acknowledgment of the presence of a problem and a plan to deal with it rather than a prescribed way of treating the patient that failed at NCCN for being absolutely non-evidence based.

So that linking a number from the previous guideline to putting an IV and giving somebody a shot of something was absolutely a huge problem, particularly in this epidemic. But this, I think addresses it wonderfully in the sense that you have a plan. That's perfect.

MEMBER FIELDS: I guess I am just thinking about usability later if your hospital gets scored because you failed to address pain in a high percentage of patients.

But I hear you. It sounds to me like the goal is to move it towards -- there is a lot of different ways to address pain.

And it is just down the road it is hard to -- it is a very hard endpoint for the providers to meet in the end. So I think -- but it is an important one. But I just think I can see it being difficult when we get to public reporting and things like that.

MEMBER BRUERA: One supplementary comment. I completely agree that the idea of coming up with a number and scoring a number would be a terrible mistake. So that is why I think this is good in the sense it does not tie these numerical reporting. It ties that we have knowledge that there was a problem and then you plan to do something. Because a lot of people complaining of ten out of ten pain are somatizing their suffering. And a lot of the people who are complaining of ten out of ten are coping chemically and they need pain killers. So if you happen to have your cancer in a rough neighborhood, you are going to get punished. This protects you from that.

So I think that is what I think

would be the nice part.

MS. FRANKLIN: Heidi and then Jennifer.

MEMBER DONOVAN: I guess the only thing I would add to that and I completely would reiterate what Eduardo has said, I think that we could get into a lot of situations if there were exclusions where we would be questioning why we were excluding. So the first thing that comes to mind as an example is somebody who is well-managed on pain medicine who comes in with mild pain. And we certainly want to be following up with those patients assessing and having a plan of care.

And I think once we start thinking about how we might exclude, we are going to come up with a lot of reasons why we shouldn't.

MEMBER MALIN: You mentioned that there was another measure from the palliative committee. Do they have the same denominator and numerator or how similar or different are

the measures?

DR. BURSTIN: The measures were slightly different because they apply specifically to patients with advanced cancer but obviously a subset of these folks could be advanced cancer.

Dr. Bruera and Steve were both on these as well.

CHAIR LUTZ: My recollection was that it was either patients admitted to hospice or have had a palliative care consult. So think it actually is a different -- it is a small subgroup.

DR. BURSTIN: I mean, I actually think it is more applicable to the prior measure because it is really about is there a standardized way to do the assessment? So the assessment sort of approach shouldn't be different because you are in hospice or palliative care or an outpatient in treatment. This, I think, is a little bit different. And again, the way we usually proceed is they

review the measures on all the criteria and then we, if the measure is deemed suitable, we will put it side-by-side with the others and see if there is some harmonization work to happen.

CHAIR LUTZ: Heidi, I think you were next. Did you -- Okay.

Larry?

MEMBER MARKS: Quick clarification. What does it mean to be a paired metric? You have to use one and the other? Do you get two points? Is it a double credit? Is it one dependent on the other? Help me out here.

DR. BURSTIN: Basically measures are paired when people believe looking at one of those measures in isolation doesn't give you the complete picture and you really need to see the two together. So they should always be reported together.

MEMBER MARKS: Do you get two points for it or do you get one point for it?

DR. BURSTIN: We don't do the scoring so I don't know. I mean, essentially I think you would still get two measures submitted under PQRS.

MEMBER MARKS: But in terms of the procedural thing, if we vote this one down, does it make the prior one automatically go down because they go together?

DR. BURSTIN: No, you would have to have that discussion.

CHAIR LUTZ: Any other questions or comments? Are we good to vote? We might as well do it.

MS. KHAN: So 1a, impact. We have 15 high and two moderate.

And 1b, performance gap. We are one vote short. We have 12 high and five moderate.

And we're voting on 1c, evidence. We have 15 yes and two for insufficient evidence.

So we are going to move on to the

liability. We have four high, 12 moderate, and one low.

We are going to look at validity. We have three high, 12 moderate, one low, and one insufficient.

And usability. We have six high, nine moderate, two low, and zero insufficient.

And feasibility. Four high and 13 moderate.

And overall suitability for endorsement, does the measure meet NQF criteria for endorsement? And we have two people missing. So we have sixteen yeses and one no. So the measure will pass.

CHAIR LUTZ: All right. Then moving on to the fourth, treatment summary communication in radiation oncology. Again, we will have our submitters submit first and then I believe that Heidi is going to be our first discussant.

DR. HAYMAN: So this is Measure 0381. This is looking at treatment summary

communication just for radiation oncology. This was again a measure that was developed by the oncology workgroup and had endorsement from NQF in 2008, which was time-limited.

So the denominator for this measure is looking at all patients regardless of age who have a diagnosis of cancer who have undergone either brachytherapy or external beam radiation therapy. And to be in the numerator, patients must have a treatment summary in the medical record that was communicated to physicians involved in the continuing care of the patient and to the patient in a timely fashion within one month of completing their treatments.

The summary needs to include the dose delivered, an assessment of how well the patient tolerated the therapy. So any acute side effect that they might have experienced during their therapy, whether or not the treatment goal was achieved. So in other words, did the patient finish therapy or not?

And then a subsequent follow-up plan for that patient.

In terms of the impact of this area, it is estimated about two-thirds of all cancer patients undergo treatment with radiotherapy sometime during the course of their illness. So I think we are talking about hundreds of thousands of patients per year for which this would be relevant.

In terms of opportunity for improvement, several components of this measure were included as part of the ASCO/RAND National Initiative for Cancer Care Quality. NICCQ is the acronym that that study went by. And again they were looking for dose delivered and the site treated. So just a couple components only for breast cancer in this particular study and found only a 50 percent performance rating.

And then within ASTRO's Practice Improvement Program, PAAROT, the average performance rate was 92 percent with a range

of zero to 100 percent.

And as part of our testing of this measure's validity and reliability we also assess performance on this measure and had a response -- I'm sorry -- a performance rating of about 89 percent. So I think that there is room for improvement.

And then in terms of again the body of evidence to support this measure, as with the prior three, this process of care measure doesn't have a randomized controlled trial to support its use. It is based on a consensus-based guideline from the American College of Radiology. They have guideline, a technical standard on the practice of radiation oncology in general and recommend that this information be conveyed in the treatment summary.

In terms of linkage between this process measure and outcome in terms of care coordination, I would argue that providing this information in a timely fashion not only

to the physicians who are caring for the patient but to the patient themselves is an important outcome.

And so I would recommend that you endorse this measure.

CHAIR LUTZ: Heidi, what did you guys think?

MEMBER DONOVAN: So there was a little bit of discussion around this measure. We, just to start out, also got hung up on the evidence that was brought to bear on the measure. And again because it appeared to be based purely on opinion or consensus from ACR and the guidelines themselves were much more broad than this specific measure. So that is really where we got hung up and so we didn't do a lot of further discussion.

Some of the other things that did come out there was in terms of importance to measure, there was some concern by some panel members that this is something that has been done for a very long time, although I think

that in terms of writing a summary, that has been common practice but the question of involving patients is quite different.

I think there was also some discussion in terms of the specification of the numerator. We had quite a bit of a discussion with that and where we ask about what exactly was the reliability assessing. Was it just the CPT-II code or was it really going back into the charts and identifying whether physicians or advanced practice nurses or clinicians were accurately documenting the code based on what was in the record and I think we were all satisfied that that was the case.

Let's see and then I guess the one other issue that was brought up was related to the gap. There was some concern about whether the citation for the only 50 percent of patients had a documented summary of treatment may not have been an accurate representation of that article.

And I think that is primarily it. This is one where I think most of us around the table would definitely say that this is a great step towards getting care coordination, something that is really important and bringing the patient into that care coordination is very important. So I think that this is definitely worth discussing further.

CHAIR LUTZ: Does anybody else in the small group have comments? Okay, the whole group? Larry, your card is up. I was looking forward to an insight there. Do you want to give us one?

MEMBER MARKS: I think it is a good thing. On some level it is a vital sign, almost. It is what we did to the patient. What is missing here though is the site. You have the dose but it doesn't specifically say the site. I presume that is implied.

MEMBER MALIN: Actually that is interesting you say that because the reference

to the NICCQ study was actually the percent of people who had a treatment summary so the denominator was having a treatment summary that included the dose and the site. And the reason for failing was most often that site was missing.

CHAIR LUTZ: Anybody have anything else or any questions?

MEMBER ROSS: So to the radiation oncologists, is there a convenient way that for example is there an epic version, is there something in the electronic medical record that will make this easy for people to accomplish or not to get that treatment plan out?

DR. HAYMAN: So a related effort, ASTRO has a Health Service Research Committee and they are in the process of undertaking a project to standardize, create some templates if you will, around reporting of the treatment summary. And that is something that ASCO has been actively pursuing as well for medical

oncology.

And so to the extent that this information can be standardized, I think really will --

MEMBER ROSS: That would certainly make this easily achievable. It could potentially be onerous for some people I would think.

DR. HAYMAN: There is a tremendous penetration in radiation oncology of several of the software vendors. So we have two companies that control 90 percent of the market and trying to -- I mean, there have been discussions underway about how to link those systems to Epic so that that information could be downloaded. It speaks to feasibility.

MEMBER MARKS: It's currently being done. So ASTRO is going to come out with this is what the complete structure should have but that doesn't mean that there is the electronic tools are there to do it.

And it is currently being done in most places, even without the electronic tools. Culturally it is viewed as something we are supposed to do. It is the equivalent of an op note.

MEMBER ROSS: I understand but I am looking for ease of doing it.

DR. HAYMAN: It's getting better.

MEMBER PFISTER: I think that one thing that is worth -- on the call I think this was actually the first measure. So kind of like if you look at our experience today what happens to the first measure, that that is always not a good place to be.

But I think what is the -- as I think you have gotten a sense from the discussion is that what the supporting data is, there is a spectrum of forgiveness in terms of like how you look at it. And when you are talking about something like pain or you are talking about something like staging, you know, you kind of go with the flow.

In the workgroup it is worth

emphasizing that virtually everyone wrote the evidence is low here supporting it. So I think that it would seem to me in looking at this measure that that is, as much as on its face it seems to be a very important thing, it is certainly analogous to the chemotherapy treatment summary or operative treatment summary that if this is something that the group is looking to -- I do think that that review of the available evidence is accurate. And I think this would be something to consider whether you need some sort of exception to move it forward.

MEMBER FIELDS: I assume that the measure was brought forward because there is a subgroup of rad oncs that don't necessarily think that a treatment summary adds to the patient care. Now I am a hem onc so I can't imagine not describing the treatments that we have given to our patients and having them be aware of that.

But I don't -- Again, this is mom

and apple pie. We should be documenting how we treated the patients and I'm -- So are there any other reasons why we would find barriers to this? Because I know that there is a subset and it has just always been surprising to me that people, some rad oncs didn't think this needed to be documented.

DR. HAYMAN: Well I think that some treatment summaries, you know, list a dose. They list the site that was treated. They list the start date and the end date and that is it.

You know, the workgroup felt that that wasn't sufficient. That you know, it was important that it be timely. You know, everyone is busy sending out a note three months after the patient has been treated isn't probably going to be that helpful. In fact, there was discussion about what the right time interval should be and that sort of gets to the issue of feasibility.

I mean, there is some data that I

have seen as the median next contact after a patient finishes treatment with radiation might be as short as a week. And so but I mean you have to be feasible.

And then you know, including the patient as well was felt to be an important component of this during the discussions in the workgroup. And then also the issue of how well the patient tolerated the treatment, what their follow-up plans are, whether they completed treatment as planned. All those components were also felt to be important. So that is why they were all included in the measure.

CHAIR LUTZ: Well I mean, actually in our practice is not a never event where necessarily someone is going to die but in 15 years, I have never not done one. And so there is always someone that needs the information immediately thereafter. So it is sort of a how could you ever justify not, I guess.

Thank you.

MEMBER FIELDS: I think we try to make those flow sheets and get the nurses to fill them in. I'm just kidding.

But I agree with you. I think that all of us, everyone that treats patients with antineoplastics should be documenting it better. And I agree. It is not -- I'm sure that the med oncs would probably have this same kind of discussion in order to go forward. So I think we think we document it with our flow sheets but it is sometimes hard to get to the data in the usable form then.

CHAIR LUTZ: I think Nicole you were --

MEMBER TAPAY: Hi, sure. I just wanted to respond to Dr. Ross and then also provide a little bit of a broader comment. There are some efforts underway, some public-private partnerships and the NCCS is part of one with UCLA. In other words, actually developing. We have an electronic treatment

summary for post-treatment and working with some private partners on that and in the course of that effort have been reaching out to Epic and some of the other groups and finding that some of the major HIT vendors are in the process of creating these. Others are slower but they are at least thinking about it. But it is definitely out there in the space right now. But obviously the specifics of what is being mentioned, I think this is why this is potentially a really timely thing is that could feed into the specifics as they are developing it.

And then just to echo I think what Heidi said in terms of how this feeds into the care coordination effort, this comment definitely goes beyond the radiation oncology as to the treatment plan issue. But a lot of you around the table are here because you are the best practices, that is what you do. That is your expertise naturally. But the findings of the Institute of Medicine and others are

that this is not happening all the time and it is in fact those findings that have led to some legislation that ASCO and us and others have been pushing on the treatment plans happening. And so there are findings out there that it is not occurring. And if that would be helpful to the group to see, I mean, those reports are available.

CHAIR LUTZ: Bryan, I think you are next.

MEMBER LOY: Yes, I just heard it mentioned but I am not sure I heard the response. Was there an intent to include the site on this measure?

DR. HAYMAN: To be honest with you, I can't remember if there was discussion about that. I am sure we could, you know, potentially because it seems to me like a relatively, I don't want to speak to the AMA staff but a relatively minor modification that that potentially could be included without -- it is always hard because we bring these as

they are, but I think that that is something that hopefully could be addressed.

MEMBER LOY: And I might, let me just add, I know they are related but not the same, would there be a reason not to include stage?

DR. HAYMAN: Not that I can think of.

MEMBER LOY: Okay.

CHAIR LUTZ: Joe?

MEMBER ALVARNAS: My question is one more based upon curiosity. In the ASTRO PAAROT program, do you know what the baseline data were for the use of these summaries and do you know of any outcome changes that were achieved beyond the scope of compliance with it?

I asked more out of curiosity because if we are trying to put a punctuation on the meaning of these metrics, it would be nice to see what was achieved through a program that kind of reinforce these

behaviors.

DR. HAYMAN: So PAAROT is a relatively new program and so I think other than the data that I mentioned, I'm not sure I have much to add at this point in time.

MEMBER PFISTER: I know that the Committee scrutinized the data that was available really carefully and I think that it is clear that this potentially does impact in a significant way on one of the IOM priority areas, which is coordination of care. But at this point it is a theoretical impact whether it truly impacts on things in a way that you would expect. It is sort of, in terms of the distinction between good intentions and actually the proximal relationship with this to what happens down the road is something that is really not addressed by available data.

MEMBER GORE: And building on that, I just wonder if this is just another

example of something that we all agree is good clinical care but maybe not a priority for performance measurement because of that lack of a link that you are talking about. And so I just wonder if this falls under that same umbrella similar to the melanoma measure.

MEMBER DONOVAN: I think the difference is is that the emphasis on trying to get more than just a treatment summary and that it is a treatment summary. It is a documentation of response to the treatment and advancement toward treatment goals. And probably more importantly a plan of care which doesn't really get specified but hopefully is a first step in sort of realizing the Institute of Medicine's desire to get what are late effects that need to be watched for, you know, what sort of follow-up should be done. So I think that is where -- I don't think that is even recognized.

And then that other piece of bringing the patient into the conversation,

which I think is critical, which I don't think is current practice.

MEMBER GORE: So maybe more analogous to the recall measure, where it is not just simply sending a report. It is invoking a plan. So that makes more sense.

CHAIR LUTZ: Okay, anyone else or should we proceed to vote? All right, let's vote.

MS. KHAN: Looking at 1a, impact. You can start now. We have seven high and ten moderate.

Looking at performance gap. We have four high, ten moderate, one low, and two insufficient evidence.

Moving on to scientific acceptability and reliability. Oh -- evidence. Sorry.

Okay, looking at evidence. We're one person short. You have ten seconds.

So we have eight yes, two no, and six insufficient evidence.

(Laughter.)

MS. KHAN: Oh, we don't? Let's try that again. All right, you can start now. There's that one last person again. There we go.

We have nine yes, one no, and seven insufficient evidence.

MS. FRANKLIN: So we go forward. So it passed. I mean, narrow but it passed. So I think you should keep on going to scientific acceptability.

MS. KHAN: So looking at reliability. Oh, shoot. Okay. So you have seven high and ten moderate.

And then looking at validity. One high, 14 moderate, one low, one insufficient evidence.

So going on to usability. We have six high, ten moderate, and one low.

And feasibility. Five high, ten moderate, and two low.

And lastly overall suitability for

endorsement. Does the measure meet NQF criteria for endorsement? You can go ahead and start.

We have fourteen yes and three no, so the measure will pass.

CHAIR LUTZ: All right, I think we have reached the time for a break and we will have to give the NQF staff a couple of extra minutes because they have to erect a large statue to a group that is really on schedule. We are actually exactly to the minute.

(Whereupon, the foregoing proceeding went off the record at 3:31 p.m. and went back on the record at 3:48 p.m.)

CHAIR LUTZ: It looks like the first one is going to be 1854, Barrett's esophagus and CAP protocol. And I think Dr. Loy is the one taking a look at that.

MEMBER LOY: That's me.

Oh, I'm sorry, the developer first. I apologize.

DR. VOLK: Thank you for having us

here today. Sorry, I'm getting used to the microphone. My name is Emily Volk. I am a private practice pathologist in San Antonio, Texas and I work in the Baptist Health System there. It is a five-hospital system.

I am with Fay Shamanski from the College of American Pathologists and Dr. Michael Cohen, from the University of Utah, who is an academic pathologist.

MEMBER LOY: And similar to some of the themes that we have had earlier today, in our general comments I would point out that our workgroup evaluated this and said yes, it is desirable but trying to make the link of the evidence to the outcome was a struggle for us. So certainly it was desirable to see the documentation. We saw that it was a good first step trying to link the evidence to an outcome in terms of quality was a bit more of a challenge for us.

The other area of interest in our discussions were that we were curious why we

would not go to the next step of trying to figure out whether it was high grade versus low grade dysplasia that would be required in the measure that was of interest in our discussions.

And then finally we recognize clearly that although it was desirable, we saw that many of the criteria that we were asked to evaluate have yet to be determined because we didn't have the data.

So all of that led us to an ultimate place of saying we could not recommend but I think as we have deliberated today and gotten a broader understanding of what may be acceptable, I think that is certainly open for additional comment.

As I review through 1854, I think I have already talked it through the numerator versus the denominator, the biopsy reports having Barrett's esophagus in the denominator looking for a mention of dysplasia. We thought it might be more desirable to have it

graded versus just present, versus absent, versus indefinite.

I think the workgroup in terms of the importance of the measure and report concluded that it was split. There was a yes, this is important but could potentially become more important if there was a little bit more definition into what the dysplasia, the grade of the dysplasia was.

And turning our attention to the evidence basis, I believe again we didn't see that we had it when we recognized that this was a new measure and that impacted many of the criteria that we had in terms of the acceptability of the measure properties. You will find that no in the usability was on the medium low or insufficient; feasibility fell into a similar category or similar spectrum of medium, low, or insufficient. And again based on that criteria, led us to a place saying that we felt like we, based on the lack of being able to have data to support the

findings, we were not able to recommend an endorsement on this particular measure.

Now having said all of that, I think we have come to a different place today understanding that this very well may be a first step in being able to accumulate the necessary data to be able to better define what the value of this measure might be. And I will stop there.

MS. FRANKLIN: And I just wanted to add this is also a measure that we are looking at that is eligible for time-limited endorsement because of the untested nature of the measure. It is also in the PQRS 2012 program. So as the Committee discusses it, please keep that in mind.

MS. BOSSLEY: -- add a little bit more about how you will vote perhaps, and part of the discussion because you don't have testing, reliability and validity testing.

So here we would ask for you to look through is it precisely specified and are

the specifications in line with the evidence. Those are really the two questions that you can answer at the moment. And then when they come back with the testing results and we will go through a review against the reliability and validity. So this really would be just a yes/no on those two questions.

MEMBER LOY: And I might ask my fellow small group members if I missed anything.

MEMBER PFISTER: Though I think was the only measure which we discussed which was in this special status. So I think that -- so it made it a little different.

The one -- You know, I'm not a pathologist. I know though that if you take lung cancer pathology and they have done these interobserver variability studies, you know, and that there can be a decent amount of disagreement, even between like sort of low grade, high grade type stuff.

And the one thing I saw with this

measure to kind of following some of the things down the road in terms of arms and things like this is that once you sort of get that there is dysplasia there, obviously it is important to know because it triggers other streams of events.

The question is is that if you end up sort of putting in like dysplasia without any sort of descriptor or whatever, the sense I get, I'm not a gastroenterologist, is that if it is mild you just kind of finesse it, keep an eye on things. If it is more, you are a lot more interactive but to what extent you potentially go down this over treatment pathway and in part the challenge cause because of observer variability associated with appropriately classifying the dysplasia in the first place.

And I guess my question for the proposers is like when you look at observer variability for this among pathologists sort of what, you know, how that looks.

DR. VOLK: I'd be happy to address that. The measure is solely based on reporting of the presence or absence of dysplasia. It does not cover grading of dysplasia; however I believe it is implied that pathologists would be encouraged to use the standard grading system, low grade, indeterminate, high grade.

The interobserver variability with high grade dysplasia is actually quite good and it is high grade dysplasia that is the sharp end of the therapeutic stick, if you will, in determining whether or not mucosal resection or more drastic intervention is required.

The anecdotal data from experts in a variety of practice settings gave us an expert consensus opinion and there is one, although limited study in 2008 that concluded that greater than 30 percent of pathology reports lacked critical information with regard to Barrett's esophagus and dysplasia.

There are also two studies in the pathology literature, one from 2003 and one from 2008 of Q-PROBES and Q-TRACKS studies from the College of American Pathologists that conclude that statistically significant dissatisfaction exists by clinicians with the quality of content for surgical pathology reports.

So although the expert opinion is that documentation is significant, whether or not there is dysplasia high grade, low grade, or indeterminate, will impact the care and treatment plan.

This measure does not address interobserver variability. And it was not designed to do so.

MEMBER ROSS: So I think this is a good measure because it is important for us to improve the quality of the path reports on this particular topic but it seems like so many other issues we talked about today like about the melanoma having the skin exam within a year or what is the pain plan, so many of

the things addressed a prospective event for the patient that implies there is the appropriate next step of care that is going to happen. This measure would be so much stronger if it included what the recommendation was for that patient with dysplasia, whether it was surveillance endoscopy, endoscopic ablation or surgical resection. I think the measure, as it stands, doesn't have a lot of oomph to it at all.

DR. COHEN: Let me try to field that one. I think pathologists are always in a quandary in trying to recommend what kind of therapeutic interventions ought to be and therefore generally we are reluctant to do so. A lot of these things are discussed at case management conferences. I suspect as a thoracic oncologist you are probably familiar with tumor boards and the like. And so a lot of these patients are dealt with on a case-by-case basis where they are discussed. But I think it would be distinctly unusual in almost

any pathology report where you would expect a specific recommendation for therapy.

MEMBER ROSS: Right. I'm not saying it should come from the pathology report but it should come from that patient's medical record whether it be the biopsy was obtained -- someone did a biopsy. So there was an interventionalist who did a biopsy. And that combination of the pathologist and the interventionalist, whoever it is, gastroenterologist, general surgeon thoracic surgeon, whatever, has to have a plan of what they are going to do with that information. And somehow recording that plan would make this so much stronger.

DR. COHEN: I think overall I absolutely agree except we have been asked to design a pathology-specific metric to improve patient care. And so something like what you are proposing with respect to integration or I think one of the words you used quite often today is harmonization is how you would truly

impact the overall care of individual patients.

CHAIR LUTZ: Karen, I think you were --

MEMBER FIELDS: So can we hear a little bit more about the natural history? Because the thing that was confusing to me was the description of the controversies in the data that some patients regress, some patients progress. But I don't really know the esophageal literature very well.

MEMBER ROSS: So I think so about 40 percent of those who develop high grade dysplasia will go on to develop an invasive carcinoma.

So at one point in time, even if you have high grade dysplasia and you don't do an intervention on the next biopsy, there may be low grade. So I do think knowing the next step is really key because the natural history is still, it is known to some extent but it is still being evaluated and the abundance of

treatment options is so good right now that we ought to start -- and we don't know which ones better.

So it would be great to get that information because industry is driving a lot of the interventions right now. Industry drives some things are indicated with low grade. Some only have indications at high grade. There are some real controversies. It is a quality issue.

DR. VOLK: If I might offer a few more statistics, there are approximately 20 million patients a year in the United States who have described symptoms of gastroesophageal reflux disease, which is considered one of the precursor states for developing Barrett's esophagus. Of those patients, a million patients will develop Barrett's esophagus.

Those patients with Barrett's have an increased risk of adenocarcinoma, as you all know, of at least 30 times. When patients

are diagnosed with adenocarcinoma of the esophagus, they have a five-year survival right now of about 15 percent.

So the key to helping these patient survive is to diagnose this lesion before it becomes adenocarcinoma, when it is in the high grade stage or even potentially the low grade stage. So I mean, this is a cancer that is responsible for two percent of the cancer deaths in the United States and early detection is the only real meaningful intervention that is available.

CHAIR LUTZ: Elizabeth, I think you were next.

MEMBER HAMMOND: Yes, I strongly agree with the thought though that it would be very useful to discriminate between the high and low grade dysplasias because I think that the treatment plan -- When we have been talking here today about different things, we have been focusing on those interventions that drive treatment in different directions

as being a sort of baseline thing that we are going to start with. And the treatment options for people in low and high grade are very different from each other and so I would wonder if the developers couldn't modify the measure to include both high and low grade.

DR. VOLK: You know, again we are asking for obviously time-limited endorsement on a measure that is currently being used in the PQRS process. So I don't think that we would -- I mean, I think we are taking this input very seriously and I think the measure in the future could potentially be modified but it is my understanding that we can't change the measure today. This measure was approved by the AMA PCPI Committee by the Physician Consortium in January of 2011.

So again, we understand that this measure is only up for time-limited endorsement.

CHAIR LUTZ: Karen?

MEMBER FIELDS: My real question

more is just understanding the natural history. And if we don't have a body of data that can give us as much of that information, isn't this more of a national high priority trial or study? Shouldn't it be some sort of registry kind of study in addition so we could actually understand that a little bit more?

Because I agree that then unless we are going to include some therapeutic questions in the future, then we can't really get to quality as much.

And my other question -- my other statement though is of course if we have a preventable disease and esophageal cancer can be potentially preventable just like doing colonoscopies and getting rid of the polyps decreases your morbidity and mortality and improves your survival, I don't have a problem with us endorsing a measure that has real meaningful input. It is just I just needed to understand the natural history and it looks like the data is quoted from the Netherlands

and maybe we need to have a high priority registration trial or something to get more data as well.

DR. VOLK: The data from the Netherlands was not about the natural history. The data from the Netherlands was about the content of pathology reports that was lacking.

The natural -- I mean, there is, the data that I was referring to, there is data from the Netherlands about the natural history and it seems that this is a clear case of precursor Barrett's esophagus low grade to high grade to intramucosal, to invasive carcinoma, not unlike the natural history of what we see in the colon for colorectal --

MEMBER FIELDS: I guess that is what I always naively thought. This is the first time that I have ever seen data about it regressing or reversing. So I don't know that I understand the disease very well.

I know that when I have reflux, I run and take some Pepcid so that I am not

going to get Barrett's esophagus but I don't know --

DR. VOLK: Some things that are defined as regression, too, may actually just be representing sampling variability, too.

CHAIR LUTZ: Elizabeth?

DR. VOLK: We certainly see that in IBD with dysplasias associated with Crohn's disease and mucosal ulcerative colitis.

MEMBER PFISTER: You know, following that breakdown you gave, I think this is kind of getting at what Karen was talking about is that you said 20 million had reflux. Of those, one million have some sort of I guess Barrett's. And then of that one million, how many develop esophageal cancer?

DR. VOLK: That's a small percentage. It is about 0.5 percent. However, those patients with Barrett's are at significant increased risk for development of adenocarcinoma and with each severity of dysplasia become more at risk.

MEMBER PFISTER: Because I think the natural history here is sort of the critical piece of information because I think that it goes to whatever you do that measures your leverage behavior. So the question is do you leverage behavior in a kind of a productive way or in a way that is at least risk neutral?

And so just following the thought process with the people I have who have Barrett's, certainly, they don't get less endoscopies. They get a lot of endoscopies and they get on the endoscopy train and so whether it is sampling, whether it is whatever. And there is a lot of downstream diagnostic testing that comes once you get kind of that is what it is.

And so I can see how with the parallel with colorectal cancer certainly makes sense. You know, you are looking at probably a more common disease. You are looking at randomized data. You are looking

at the role of polypectomy and things like that. But it is just something I think when you are considering a measure like this, you need to kind of kick around because this is something that is clearly going to lead to a ton more diagnostic testing.

You know, just because it is late in the afternoon I will share a joke and lighten up the proceedings. Well I guess it is recorded. But this thing about the unintended consequences of what you do, a few years ago I was going overseas to Austria. And so my older daughter says oh, went to the library and got German tapes. And my other daughter kind of sees what my older daughter did; she goes to the library. She gets Japanese tapes. And so, you know, she is connecting on the fact that gee, you are going over to sort of a language but it sort of ended up being a different way you wanted to go.

And so I think you need to

consider what are the other consequences of what the metric is going to lead to.

CHAIR LUTZ: Pat?

DR. VOLK: If I can comment to respond to that. I would say that by informing the clinician clearly in the pathology report about Barrett's esophagus and certainly the pathology report doesn't drive the endoscopies per se but the biopsies that come to pathology then should have reports that are complete, including whether or not dysplasia is there. And I don't disagree with appropriate grading if it is there. That way, patients are put on the appropriate endoscopy train, if you will. And so you don't have a patient going on a track that would have him receiving unnecessary too frequent endoscopies, if in fact they have low grade versus low grade dysplasia or no dysplasia at all.

So I think clear communication of whether or not dysplasia is present or absent

would actually help reduce the number of unnecessary endoscopies that you are concerned about and understandably so.

CHAIR LUTZ: Pat, can I ask a quick question? We have used the comparison between colon cancer and esophageal cancer a couple times. And after all these years, we finally have a study that says if you do colonoscopies it can change survival. My sense was that for this progression from dysplasia to cancer and esophagus cancer, we have less data than that by far to know whether we are really impacting. Is that a fair way to phrase it?

MEMBER ROSS: No, I think that is true and I was going to say that it is not really analogous because we know that colonoscopy as a screening tool is effective. The real question is do the 20 million people with reflux all need an upper endoscopy? I mean if you really want to take this back to what is a national objective that we could

help because we are not, right now, we don't have specifics -- Well we do have guidelines but we have guidelines predominately for driving when once we have a biopsy. We don't have the guidelines for before the biopsy.

DR. MYLES: This is Dr. Myles. Can I make a comment?

MS. FRANKLIN: Yes.

DR. MYLES: I'm a pathologist at the Cleveland Clinic and I think that the natural history of Barrett's is well understood. I mean Barrett's progresses through a series of stages from Barrett's to dysplasia to invasive cancer. You know, the five-year survival for invasive cancer is 15 percent and patients with identified high grade dysplasia, 13.5 percent of those patients per year will progress to invasive cancer. So it is important that the pathologist identify dysplasia in the specimen.

In fact if you do identify

dysplasia, that triggers a treatment change in the frequency of surveillance or more. Whereas, if you have repeated endoscopies with negative dysplasia findings, your frequency of endoscopy will decrease and the guidelines do state that.

I think that identification of dysplasia is important and when the measure was developed, it is not controversial whether dysplasia is a precursor to cancer in esophagus. What the controversy is as was stated, is whether patients with reflux need to get scoped. That is where the controversy is. The controversy is not if you have Barrett's whether you need to get scoped or if you have dysplasia whether you need to get scoped.

We would certainly be open in the future to considering altering the measure to include grading of the dysplasia but why that wasn't included originally, that is a little bit more controversial.

What is not controversial and what the measure states is whether dysplasia is there or not. That is not controversial. Thank you.

MEMBER PFISTER: I have one other question while you are on the phone. Going back to the question I asked before, when you have the 20 million with reflux, one million with Barrett's and then let's say you biopsy those million with Barrett's, what is the dysplasia breakdown?

DR. MYLES: I don't have that number off the top of my head so I can't answer that question.

CHAIR LUTZ: Okay, do we have any other questions or issues? I think were you about to give us more?

DR. COHEN: Certain kinds of dysplasia do regress. There is a well-defined percentage.

CHAIR LUTZ: So this is usually when I ask if we want to vote but we had a lot

of different discussions in different directions. Is there any further pathway you would like to follow on any of those?

MEMBER FIELDS: So your original statement was more about your group had a lot of controversy and voted one to four not to approve it because you didn't know how to interpret the science but everyone felt comfortable with the concept? Just so we could understand when we weigh that.

MEMBER LOY: I think if I were to reflect back on our calls and our discussions, it was largely around we didn't feel like we had the evidence in order to say that I think there were two studies that have been cited. We didn't feel like we had the evidence to be able to say conclusively this met the quality and quantity that would support the link between the documentation of ungraded dysplasia to a health outcome. But we did acknowledge that it would be desirable to collect and document that data versus not

documenting it.

So I think that is where we went to.

MEMBER PFISTER: You know, and I think also clearly in retrospect we are still kind of having the mindset from the prior measures and this was only one that went into this kind of candidate, you know preliminary measures. You know, I think that there was perhaps a higher bar than would have necessarily been appropriate given the different status.

CHAIR LUTZ: Jennifer?

MEMBER MALIN: I think one of the distinctions between some of the other measures and maybe the measure developers can provide this is that -- and I think the summaries weren't as robust as they could have been for some of the other measures. But for stage even if documentation of stage doesn't have an outcomes link, there is well-recognized links between the documentation of

stage and what your next clinical process is going to be. So there is a link to the intermediate process. I guess the question is here and I don't feel like I have gotten a total sense. Is there a clear link between what the dysplasia is and how that is going to effect the course? I mean, I heard that well maybe you would get a few less endoscopies a year if they were less low but I didn't hear kind of definitively like if someone is low grade dysplasia that they no longer have to be screened anymore.

DR. VOLK: There is a clear link, actually. And in fact, if a patient has a diagnosis of Barrett's esophagus and has two consecutive years of negative for dysplasia screening, they can be taken on to a much less frequent endoscopy schedule. So yes, and these guidelines are outlined in the American College of Gastroenterology guidelines for a diagnosis, surveillance, and therapy of Barrett's esophagus.

MEMBER ROSS: So two quick comments. The first is that in some ways this is like the staging because they were asking for -- actually we have had three of them today that in my mind are the same. So should a radiation oncologist or should any doctor communicate with the other doctors on the team? Well you would be silly not to say yes but we now made that a quality measure.

Should a patient with a malignancy be staged? Well yes, but now the third one is should a pathologist do an accurate interpretation of an esophageal biopsy? I mean, are we going to say no?

So to some extent, at the simplistic level we have asked the same question three times, which is if you are a doctor, should you do the right thing every time. And that is what all three of these measures are.

So I think that yes, this is a good thing to do but what we should try to do

is make it as robust as possible. On all of these measures they should be as robust as possible. Otherwise, why are all of these smart people sitting here and people in the next set of committees trying to make something out of what it isn't?

CHAIR LUTZ: Well then I wonder it had been mentioned about the low, maybe the bar is too high. This may be the one of those three where since it is something that is going to be time-limited, there will be data to come back and the other two are pretty much set. This one actually has a chance to then become more robust with an exception.

MEMBER ROSS: I think we should move this one forward but we need to broaden in. Why are we only interested in staging breast and colorectal? You know, when you hammer everything that looks like a nail, I want every lung cancer to be staged appropriately. Right?

We just need to, I think we need

to look at what we are doing.

MEMBER MILLER: So just a point of information. Did I understand someone to say that because this is a 2012 PQRS measure we can't change it?

MS. BOSSLEY: So this is always the dilemma when you have got a measure that actually is -- I don't think you have started testing yet. Correct? But it is being reported on actively in the PQRS program.

So and part of this will be up to the developer to determine whether or not they would be willing to make changes. I don't think you want a completely new measure because it is a completely new measure and everything that they have provided to you in importance changes. But if there are things that you think that would make this stronger so that they incorporate that into their testing, they make the changes to what it is now and it goes into the testing, then I think you should discuss it because I do think this

question will be revisited when it comes up with the testing at some point, I would assume.

So if there are things that don't change it completely, I would think they were on the table if the developer is willing.

MEMBER MILLER: Yes, so I think others have said it but I will say it also. I think the greater dysplasia ought to be in the measure, bottom line.

MEMBER HAMMOND: You know, I think maybe other people than just me around this table have the same frustration and that is, there is this PQRS measures which if they are already out there, if we make good suggestions that probably should be incorporated in measures to make them stronger and better and more likely to do what we need, those don't have to be taken into consideration. We have no option to help developers get us measures that are really useful. I mean, is that true?

What is the recourse? If we have

some good ideas, say something comes up here that is really what we should do, in this case we can tell the developers and they don't have the testing so it might work but what about all those measures this morning where we had ideas? What happens with them, nothing? We just have to wait?

MS. BOSSLEY: It's a really good question. So today I think one of the things you should do and you will get asked tomorrow to identify gaps. But this is part of one of the challenges we have identified and actually developers have come back and asked us to find a way to redesign our process to allow that feedback to come earlier. Very similar to what you are looking at now, we actually are working on redesigning the process to allow measures to come in reassessed as a concept against the importance criterion and then give developers 18 months to go test it using the feedback that is provided by the Committee and then bring it back fully specified to be

endorsed. But that is not yet available right now. We are going to pilot it hopefully this summary and then implement fully if it is approved next year.

So today you don't have that, unfortunately, so we are in that middle ground at the moment. So I would give them the feedback and we will see if they are willing to make the change and then I think you need to vote on the measure depending on that today or if we need to give them a little time later.

And then in the future, hopefully we are hoping to solve this issue.

CHAIR LUTZ: So if I am hearing correctly, are we asking the developers if they are willing to split it off into high and low grade? Is that what we are asking as a group? Okay.

Jennifer?

MEMBER MALIN: I was going to say I don't think the voting necessarily has to be

contingent on them doing that. I mean, we will get a chance to reassess at the end of the time period.

MS. BOSSLEY: I do think though what would show, you want it updated to show what you voted on because that will then be what is on the website and go out for comment and everything else. So we would ideally want that in the changes that I think you have just heard they are willing to do it. So your vote would then be assuming that change is made. We will circle back and share it with you but that is how I would recommend you move forward.

DR. SHAMANSKI: Can I ask a question? So the measure as it is currently written is what is in the PQRS now and we can't change that now. And it is not likely we will be able to change it before 2013 and get it in the system. So in order to continue to use it we need to know if it is usable as is with the idea that we could change it in

the future. We can make it better but we also need to know about this measure as it is, too.

MS. BOSSLEY: So I think you can get the changes made for 2013, unless the timeline has changed since the last time I knew it. But there is always a discrepancy between what is in a public program and then when it is maintained or updated by the developer, that is kind of how -- it is imperfect but I think what would be endorsed would be what the committee is asking if you have agreed to it and there would just be hopefully a short discrepancy with what is PQRS. That would be the hope.

MEMBER HAMMOND: Can I add this is not really a direct question about this particular -- but it is about this measure. Think how much stronger this would be if we could have a combined measure between pathologists and gastroenterologists that said was the report correct and did the gastroenterologist act on the recommendations

appropriately.

Are there any strategies out there to try to combine measures between groups of physicians or are they all just specialty related?

MS. BOSSLEY: So ideally, measures are as broad as possible to be applicable to any specialty or any person who is caring for patients. These just tend to be more narrow slices because there are only a few people who, you know, there is one specialty that really does this.

There are measures and I am trying to find, that I think are coming forward from AGA and there might be a Barrett's esophagus measure in there. So again, one of the things we can do is always show as that measure comes forward and if it is reviewed, we can show that there is a suite of measures or several measures that should be used together when looking at a patient, more of a patient-centered piece.

I just don't -- Developers have tried to bring forward some measures on Barrett's esophagus before and it has been very challenging, given the evidence and everything else. So I am looking. I can't remember if we have one or not.

CHAIR LUTZ: Okay so just to be clear for my sake, if we voted now are we voting on it as it is with the promise that when it is allowable it will be more divided into high and low grade dysplasia? Is that sort of the process we are voting on what it is now. Is that correct?

MS. BOSSLEY: I would actually recommend that you vote -- Assuming CAP is willing to make the change, I would recommend you vote on it with the change. There will be a difference between what is in PQRS but that is actually quite common right now.

But what you would want to see is that measure if it is picked up by other groups using what you have recommended as well

as show the change in PQRS. So assuming CAP is willing to do it, I would recommend that you take it with the changes and vote on it.

MEMBER HAMMOND: Could somebody tell us what we are -- exactly how we vote for this again or are you going to do that when we vote? Because somebody went by that really quickly. This is different voting than what we just did. Right?

MS. FRANKLIN: That's right. For this vote you would be just voting on whether the numerator and denominator in the exclusions are clear and precise and then you would also be looking on whether the measure focus is supported by the evidence.

MS. BOSSLEY: Right. We have a slide specific to this time-limited measure on scientific acceptability. We are all set.

CHAIR LUTZ: Are we good to move on to those now to the voting part?

MEMBER MARKS: Could somebody state what they mean the numerator is exactly?

We are voting on something that is not written down. So I just want to make sure we are all on the same page.

CHAIR LUTZ: Actually it would be best if the developer used the words they were comfortable with.

DR. SHAMANSKI: I'm not sure what that would be yet. I think we would have to go back and figure that out based on what you have recommended. But essentially I think the statement --

DR. VOLK: Actually I think what we could do is say the numerator statement currently says esophagus biopsy reports with the histologic finding of Barrett's mucosa that contain a statement about dysplasia (present, absent, or indefinite) and then perhaps we could put, comma, if appropriate grading would then be reported.

Would that be acceptable?

MS. BOSSLEY: So we will make sure you see the language again one more time but

it is a good idea to clarify it before you vote.

DR. VOLK: Thank you the opportunity to do that.

CHAIR LUTZ: Do you guys want to wait to vote then? I mean, just -- You're okay to go? All right, let's go.

MS. KHAN: So 1a on impact. We have six high, ten moderate, and one low.

Looking at performance gap. We are one short. Oh, there we go. Two high, 12 moderate, one low, and two insufficient evidence.

And then 1c evidence. So we have 11 yes, two no, and four insufficient.

So this is specific to untested measures. The foundation for reliability and validity, the measure specifications, numerator, denominator, and exclusions are unambiguous and likely to consistently (1) identify who is included and excluded from the target population; (2) identify the process

condition or event being measured; (3) compute the score and reflect the quality of care problem seen in 1a and 1b and the evidence cited in support of the measure focus in 1c.

Again, you are voting one for yes and two for no. We have 16 yes and one no. So I believe we move forward. Right?

So looking at usability. We have three high and 14 moderate.

And feasibility. We have eight high and nine moderate. And lastly, overall suitability for endorsement. Does the measure meet NQF criteria for endorsement?

So we have 16 yes and one no. So the measure will pass.

CHAIR LUTZ: Okay, and I think we have one more for today. 1790: Risk-adjusted morbidity and mortality for lung resection for lung cancer. So we will have our submitting folks discuss it and then who is that?

MS. FRANKLIN: From STS.

CHAIR LUTZ: From STS and then I

think Dr. Ross is going to be our first discussant after they are done.

MS. FRANKLIN: Is there anyone on the line or in the room from STS?

MS. REESE: Yes. Hi, this is Vadie Reese from STS.

MS. FRANKLIN: Could you tee up the measure for us, tell us a bit about the measure?

MS. REESE: Okay, can you give me one moment? We should also have our surgeon leader, Dr. Cam Wright. I just want to make sure he is on.

DR. CAMERON WRIGHT: Can you hear me? Hello?

MS. FRANKLIN: Yes, we can hear you.

DR. CAMERON WRIGHT: Oh, I'm sorry. Okay.

So this looks at a very common problem obviously, lung cancer, about 200,000 deaths per year. And for those lucky 25

percent of people who have early stage disease, lung cancers offers the possibility for a cure and is the standard of care.

And there is a fair variation in the outcome of perioperative morbidity and mortality after elective lung cancer resection. And we developed a measure in the STS looking at elective lung cancer resections in patients older than 18 and that is the denominator. And the numerator is patients who have an elective lung cancer resection older than 18 that have significant serious complications and they are outlined in our measure application. But those include re-intubation, need for tracheostomy, ventilator support greater than 48 hours, ARDS, pneumonia, pulmonary embolus, bronchopleural fistula, bleeding requiring reoperation, myocardial infarction, or operative mortality.

And we developed a risk adjustment model based on preoperative risk factors and centers and have published it. And we now

report it at an outcome measure every six months to the centers who report to us.

Although the vast majority of surgeons who participate in this database are thoracic surgeons, several years ago we did open it up to general surgeons as well. Currently, about 20 to 25 percent of lung resections in America are done by general surgeons, whereas 80 percent are done by thoracic surgeons. The number done by general surgeons is declining every year just because of the modern specialization of surgery but there is that number. But we do allow them to participate in our database and a number of them do.

They also obviously have the option of participating in the ACS-sponsored NSQIP database, which does allow entry of pulmonary resection as well. But we believe ours is far superior.

And our data is audited. We have an independent agency that audits a randomly

selected pool of participants for all important data measures, including major complications and mortality. And our agreement rates are over 95 percent. So our data, we believe is quite accurate.

And even though it is somewhat of a select group that participates in the STS database, people who are very early adopters, very interested in quality, there is still substantial variation that is statistically significant between the best providers and the worst providers. And we view this as just furthering our goal of pushing quality forward in cardiothoracic surgery. I know all of you are familiar with the STS adult cardiac database and those measures. And we do plan as our next major initiative in the next three years to move this to public reporting just like we did a little over a year ago with our CABG measures for public reporting.

And I think I will stop there and let people ask questions and comment as need

be.

CHAIR LUTZ: I think Dr. Ross led the small group discussion on this.

MEMBER ROSS: Thanks. So Cam, it's Pat Ross. How are you?

DR. CAMERON WRIGHT: Great!

MEMBER ROSS: Good. So we discussed this in our workgroup and unlike so many of the others we talked about today, this is a true outcomes measure that we will be voting on, as opposed to a process measure.

And it is late in the day and I don't know, David, is it good to be the last one or not? I have heard you say it both ways. It's good to the first and then it is good to be the last. So I am taking you at your word that it is good to be the last.

MEMBER PFISTER: Well so that is the difference between being a surgeon and medical oncologist. So you are going to be a lot faster.

(Laughter.)

MEMBER ROSS: So Dr. Wright has done a great job detailing the numerator and denominator. And I think it underscores the real importance of this, which is the fact that this is the most common operation done for resecting lung cancers and there is tremendous variability in the outcomes.

Institutions and surgeons who utilize this database and the data which comes back to them can actually use this as an almost real-time quality improvement measurement and process. And I think there will be a lot that the individuals will learn. I'm very supportive of this becoming a measure.

I have a couple of concerns that came out through the workgroup and you can see the comments in there. And the first is these are obviously self-reported. It is an election to participate in the STS database and we have already heard that we are not going to collect at least 25 percent of the

data and probably that will go down. But at this point, we don't get that.

So that you wind up with a database right now that is populated by centers that are hopefully motivated to deliver good product. So you really kind of, you are looking at comparing best in class is what you would hope and it will drive the bar. But it is true that if you are not participating in the database, your data won't be reported. So the metric falls short in that one area.

Otherwise, I think that this is something that our workgroup was worthy of endorsement and hopefully the group at large will agree with that assessment.

CHAIR LUTZ: Anyone else from the smaller workgroup want to add to what Pat said?

Karen, you want to dive in?

MEMBER PFISTER: I totally agree.

MEMBER FIELDS: I just wanted to

ask, so this is sort of a service, a quality of service indicator but it is not getting to the interdisciplinary oncology care question, which is was the right procedure done for the right patient.

So are there other plans in the end to add that like adequacy?

MEMBER ROSS: Well in some ways it is a surrogate for that because the fact is that a number of these patients are part of a multimodality or multidisciplinary care and the perioperative outcomes do reflect whether patients have had chemotherapy or radiation therapy prior to surgery. So I do think that the concept of multidisciplinary care is built into this and can be abstracted from it specifically as the stratification.

Would you agree, Cam? Is that correct that we could stratify patients who had induction treatment from the database?

DR. CAMERON WRIGHT: Yes, that's one of our preoperative variables is induction

therapy.

MEMBER ROSS: Yes, so I think it does get to that point.

MEMBER FIELDS: And just also there is a body of evidence that I don't know how it has been validated but that suggests that treatment by a thoracic oncology-trained surgeon outcomes are different and partly it is because of the adequacy of the dissection, the adequacy of the lymph node dissection and some other kind of things. And this really looks at morbidity and mortality which is an important endpoint because that meant you had a well-trained thoracic oncology surgeon.

But just the other data that is frequently cited includes adequacy of the other variables.

MEMBER ROSS: So actually you can stratify three groups of surgeons who do thoracic surgery. There are the general surgeons who do thoracic surgery as a part of their training. Generally that is an older

group of surgeons. The second group of surgeons who do cardiac and thoracic surgery. And the third is the group of thoracic only. And I think there is evidence that continues to be presented at the meetings in abstracts and publications that shows that the outcomes follow those three stratifications.

DR. CAMERON WRIGHT: And if I can just jump back in there --

MEMBER ROSS: Please, go ahead.

DR. CAMERON WRIGHT: -- and just say that you are getting a little ahead of us in terms of this adequacy of lymph node dissection, for example, and proper staging.

And indeed there are multiple publications that suggest that dedicated general thoracic surgeons do a better job of both staging and lymph node dissection and also have lower perioperative mortality and actually have improved survival. In our next three years, we plan have to a publicly reported measure which will include that and

we are going to come back to you all with that measure, which has those process measures within it. It is going to be very much like the adult cardiac database publicly reported measure, which has a combination of process measures and outcome measures.

And to me, the best measures have combinations of both. But this is a step in that direction. And this is a huge step but we are going to progress.

MEMBER FIELDS: So that answers my question because I think it would be a missed opportunity. We are getting to the low hanging fruit which is morbidity and mortalities decrease but then long-term survival and outcomes.

MEMBER MILLER: So I guess I was wondering about the specificity of this. And I think what troubles me a little bit is it looks like it is basically any adult patient over 18 getting any type of lung cancer surgery by anyone who is in the database.

So I just wonder, what do we hope to learn from this? I mean, I know we are looking for sort of patterns of care but just to play devil's advocate for a second, you could pick a measure like this for any type of surgical procedure done by anybody and I just want to understand why was it so broad. Was it because I think the physician on the line may have said this, you are going to be more specific with process measures later on.

But I guess that troubles me just a little bit. If you got to the trouble of making this a measure and collecting the data and reporting on it, do you think you are really going to learn enough to go to step two?

MEMBER ROSS: Oh, I think you absolutely will. I think that until you start to look at the patients, look at the outcomes, look at the details as you stratify them, you don't know. And you could pick this for any. You should. I mean we should have a measure

for colectomy and we should have a measure for every surgical procedure you want to come up with. I agree.

MEMBER MILLER: Why not pick fewer causes of morbidity, then? You have listed seven or eight different causes of morbidity.

MEMBER ROSS: So these are all captured within the database. So this is data that is currently being collected by everyone who participates. It adds no -- it is not additional workforce, if you will. The data is already there. I think this is a chance to get it endorsed by this venue.

MEMBER PFISTER: One comment and one question. I think that the big public health problem, if you look at the current NQF list measures considering that for solid tumors that surgery figures as the prime cure modality for most of them, it is actually there is an enormous under-representation of surgical measures in the NQF kind of group. And so I think any direction here is

definitely, I think, important.

I can understand about the self-reporting thing but I guess the one question that I have is that I would think that a lot of the patterns of care data and things like that often were based on surgical procedures because you can kind of track them through coding pretty easily. And so I guess that looking at this, I would think that from administrative data, that you should be able to track the procedure, track readmissions, track a lot of these things that you are looking at without sort of doing the self-election that people participate. And is it the risk adjustment isn't felt to be -- and I would think that there is probably risk adjustment that you could do off the billing data as well. Is there some that the risk adjustment is better doing it this way? Is there some reason not to do something which would be not a self-selection for providers to participate but actually that you need to

participate?

MEMBER ROSS: So the first is that the self-reporting is not an issue because the auditing shows greater than 95 percent accuracy. So I think that it is not that there is anyone gaming the system. I think their data, 21,000 cases evaluated over three years with excellent consistency.

So as far as the second, it is risk stratification, which I think adds an enhancement to this. You can get pure morbidity and mortality -- pure mortality off of any national database but that doesn't help you in terms of stratification by the perioperative variables or the type of pulmonary resection.

MEMBER PFISTER: I guess, but did I misunderstand though that -- So everyone participates but it is self-report. It is not that you selectively participate.

MEMBER ROSS: Some institutions participate and all surgeons at an

institution's data will be entered.

MEMBER PFISTER: But you can't mandate an institution to participate.

MEMBER ROSS: Correct.

MEMBER PFISTER: So I guess the other thing is it may be that the risk adjustment isn't that you would do administratively is not going to be as good as what you have. But I mean when they report the CABG data, for example, I mean don't they risk-adjust that using like a tool that is out there? Like it is not like they just say alive or dead in 30 days. They do something to adjust for case mix.

MEMBER ROSS: Right.

MEMBER PFISTER: And so I would think there must be some kind of, maybe not as perfect as this, but it is offset with the fact you are getting a complete denominator; as opposed to well certain institutions are saying that I want to participate. Like do you have any idea which percent would not

participate?

MEMBER ROSS: I don't but Dr. Wright may. Cam, do you know how many centers currently enter data into STS database for thoracic?

DR. CAMERON WRIGHT: You know we don't know the true denominator of institutions that do this. We currently have 220 institutions who participate. Every year we grow by 10 to 20 institutions. This database has only been in existence since 2003. It gets bigger every year.

I believe when we have a publicly reported measure, we are going to drive many, many more people to participate because it will be we want to prove that we are just as good as you type thing.

And we also have to remember that administrative data, while it might be a little bit easier to collect is not nearly as good as clinical databases like the STS database or like the NSQIP database. We have

to remember there is multiple publications looking at both the STS cardiac database compared to administrative and NSQIP data compared to administrative, that there is an approximate 20 percent error rate with administrative data, which impacts the results. And our data is much better. It does require, you know, you have to sign up and pay your $500 a year but it is much more high quality data.

MS. REESE: Also, in 220 participants and more than 750 surgeons, general thoracic surgeons.

CHAIR LUTZ: We'll go Jennifer and then Larry.

MEMBER MALIN: Is there enough specifications so someone could use the measure without participating in the database?

MEMBER ROSS: Say that again. I'm sorry.

MEMBER MALIN: I mean right now --

DR. CAMERON WRIGHT: There would

be no risk adjustment.

MEMBER MALIN: So there is no -- Is that typical that measures are linked to just one specific database and way of collecting it?

MS. BOSSLEY: So as is the case with other measures similar to the ones that use the STS database, or there is the College of Surgeons, there is other ones as well, the measure should be specified to the point where you could use them.

So any other way. So if it involves risk adjustment, the risk adjustment should be clear enough that if anyone else wanted to take that information and had a pot of data could run it that way. The chances of someone else doing it, I don't know but it is always possible.

That is what you would want. You would want the specifications precise enough that if anyone else wanted to take the measure and use it, they could. Does that help answer

your question?

DR. CAMERON WRIGHT: And we have published this model. And the risk model is published with all the risk factors with the odds ratio. So if you had a calculator and a computer, you could calculate your risk. But you know, we have the DCRI, the Duke Clinical Research Institute do it because it is a lot of data crunching. But is published. It is in a public domain. All the intercepts and odds ratios are in the paper.

MEMBER MARKS: I think this is a great metric. This is the best one by far. I'm not supposed to compare them but this is a real health outcome.

What we heard this morning were things like is there an op note. Did you guys check off an FEV1 before you operated. All good things but this is did the patient live or die. This would be analogous is my patient alive three years after my radiation, which we are not talking about. So I think a lot of

the criticisms that we heard, I think they are valid criticisms but they are just as valid against all the other stuff we heard this morning and this one is much farther to the right side of where we should be trying to go. So I commend surgical colleagues for doing this and pushing us forward and setting the bar pretty high for the rest of us.

MEMBER ROSS: I appreciate it. I think this is a great opportunity for us.

CHAIR LUTZ: Given Dr. Marks' strong recommendation, should we head toward a vote now?

DR. CAMERON WRIGHT: Yes, please.

MS. KHAN: Okay, 1a impact. We have 17 high.

And performance gap. We have 11 high and six moderate.

And looking at evidence. We have 17 yeses.

And looking at reliability. We have eight high and nine moderate.

And looking at validity. We have nine high and eight moderate.

And looking at usability. We have 15 high, one moderate, and one insufficient.

And feasibility. We have ten high and seven moderate.

And your overall suitability for endorsement. Does the measure meet NQF criteria for endorsement?

We have 17 yeses and zero no, so the measure will pass.

CHAIR LUTZ: All right, well done.

Sorry. Public comment. I saw the 5:00 hour and I just got all excited. I'm sorry.

Well we certainly want to know if there is any public comment, absolutely. Anyone on the line that needs to make a comment?

OPERATOR: As a reminder, that is *1 for public comment. And Charles Hampsey, your line is open. Your line is open, sir.

DR. CAMERON WRIGHT: Oh, am I supposed to say something?

OPERATOR: Charles Hampsey?

DR. CAMERON WRIGHT: No, no, no. My name is Dr. Cameron Wright. Can you hear me?

CHAIR LUTZ: Yes, you're fine. We are looking for folks not involved in the process who are listening in. You are good.

DR. CAMERON WRIGHT: Yes, okay.

OPERATOR: And yes, we can hear you now.

MR. HAMPSEY: Thank you. I just had a comment with respect to Measure 0383 and 0384, those are the paired measures.

We are generally very supportive of the measure, however we did have concerns about the descriptors for the measures and some of the exemptions in terms of the focus of the measure in that it targets intravenous chemotherapy. So that if patients were to be on an oral chemotherapy a physician would be

precluded from reporting that.

And we believe that this measure should include all modalities of care as well as -- so that patients could be considered if they had pain, regardless of the type of therapy that they were on.

And just I know that there seems to be some reliance on the infusion codes in the measure but we note that there are a number of measures from PCPI which do include oral chemotherapy, such as 0385 which is a chemotherapy measure for colon cancer, 0387 for hormonal therapy, and even the cancer staging measure that was discussed earlier. And also some of these other mechanisms do rely on CPT codes but there is registry reporting and electronic health records.

So it is just our hope that in the future with some of these other data collection methods that this measure could be broadened to include all patients regardless of the type of modality of treatment.

Those are my comments. Thank you.

CHAIR LUTZ: That was a good addition. Good comment.

Any other folks on the line to make comments?

OPERATOR: Not at this time.

CHAIR LUTZ: All right, well done.

So the only two announcements I think NQF wanted us to say we can leave our name tags where we are because we will sit back in the same seats.

And then they have asked if we can, since we are so efficient, if we could get together at 8:00 tomorrow instead of 8:30 for a starting time.

MS. TIGHE: I'll bet everybody's got problems with their flights coming at the end of the day. So the earlier we can get going the better.

MS. FRANKLIN: And you can leave your voting clickers, too.

Nicole, are you still there?

OPERATOR: Yes, ma'am.

MS. FRANKLIN: Oh, we have completed our meeting.

(Whereupon, at 5:04 p.m. the foregoing proceeding was adjourned.)