NATIONAL QUALITY FORUM
+ + + + +
CANCER ENDORSEMENT MAINTENANCE
STEERING COMMITTEE
+ + + + +
WEDNESDAY
MAY 23, 2012

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., Stephen Lutz, MD, Chair, presiding.

PRESENT:
STEPHEN LUTZ, MD, Blanchard Valley Regional Cancer Center
JOSEPH ALVARNAS, MD, City of Hope*
ELAINE CHOTTINER, MD, University of Michigan Medical Center
HEIDI DONOVAN, University of Pittsburgh School of Nursing*
STEPHEN EDGE, MD, Roswell Park Cancer Institute
KAREN FIELDS, MD, Moffitt Cancer Center
JOHN GORE, MD, MS, University of Washington School of Medicine
ELIZABETH HAMMOND, MD, Intermountain Healthcare
JOSEPH LAVER, MD, MHA, St. Jude Children's Research Hospital*
BRYAN LOY, MD, MBA, Humana, Inc.
JENNIFER MALIN, MD, PhD, WellPoint
LAWRENCE MARKS, MD, FASTRO, University of North Carolina School of Medicine*
ROBERT MILLER, Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins

DAVID PFISTER, Memorial Sloan-Kettering Cancer Center
ROCCO RICCIARDI, MD, MPH, Lahey Clinic Medical Center*
PATRICK ROSS, MD, PhD, The Ohio State University Comprehensive Cancer Center - James Cancer Hospital
NICOLE TAPAY, JD, Eli Lilly and Company
WENDY TENZUK, Colorado PERA

NQF STAFF:
HEIDI BOSSLEY, MSN, MNA Vice President, Performance Measures
EUGENE CUNNINGHAM, Project Manager, erformance Measures
ANGELA J. FRANKLIN, Senior Director, Performance Measures
ADEELA KHAN, Project Analyst, Performance Measures
KAREN PACE, Senior Director, Performance Measures
LINDSEY TIGHE, Project Manager, Performance Measures

ALSO PRESENT:
MARK ANTMAN, DDS, MBA, AMA-PCPI Measure Development

MARY BARTON, MD, National Committee for Quality Assurance
SEPHEEN C. BYRON, MHS, National Committee for Quality Assurance
LINDEE CHIN, MD, ActiveHealth Management
KERI CHRISTENSEN, MS, AMA-PCPI Measure Development

MICHAEL HASSETT, MD, MPH, Dana Farber Cancer Institute*
KRISTEN McNIFF, MPH, American Society of Clinical Oncology
CAROL S. PALACKDHARRY, MD, MS, ActiveHealth Management
FAY SHAMANSKI, PhD, College of American Pathologists

V.O. SPEIGHTS, JR, DO, College of American Pathologists and Texas A&M Health Science Center College of Medicine

ANDREW STEWART, MA, American College of Surgeons

SAMANTHA TIERNEY, MPH, Physician Quality Reporting Initiative*

EMILY E. VOLK, MD, College of American Pathologists*

DAVID WITTE, MD, PhD, FCAP, College of American Pathologists

*Present by teleconference

T-A-B-L-E O-F C-O-N-T-E-N-T-S

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

9:05 a.m.

MS. FRANKLIN: Hello, and welcome to the Cancer Endorsement Maintenance Steering Committee Meeting. We are looking at Phase II of this project.

And in the room I have with me -- my name is Angela Franklin, I'm the Senior Director for the Project.

Dr. Steven Lutz is our Chair. And in the room with me on the project is Lindsey Tighe, our Project Manager, as well as Adeela Khan, our Project Analyst and Eugene Cunningham, our Project Analyst.

So, with that we'll go ahead and get started with introductions and disclosures of interest around the room. And then we'll go to our members that are on the phone.

MS. BOSSLEY: About disclosures, you did that the last time, but we have several people who are new. So if you have, again, anything that is relevant to the work

before this Committee, a slightly different set of measures, please disclose anything related to that. Other than that, you can just say "no disclosures." But, again, just covering our bases since a few new people in the room.

MEMBER TAPAY: Nicole Tapay. I've changed jobs since the last meeting, so I'm actually now with Eli Lily. But I'm not aware with respect to any of these standards any conflicts.

MS. FRANKLIN: Since we started on that end, do you mind, Dr. Miller, we'll start with you.

MEMBER MILLER: Thank you.

Bob Miller with Johns Hopkins. And I can't remember if this disclosure is relevant, but I'll just say it: Research funding from Pfizer.

MEMBER EDGE: Stephen Edge. I'm Chair of the Commission on Cancer.

As disclosed originally, I've

participated on development of measures six or seven years ago but have not since.

CHAIRMAN LUTZ: I'm Steve Lutz, radiation oncologist from Findlay, Ohio.

No new disclosures.

MEMBER CHOTTINER: Elaine Chottiner, University of Michigan.

No disclosures relative to these measures.

MEMBER TENZYK: Wendy Tenzyk, Colorado Public Employees Retirement Association.

No disclosures.

MEMBER GORE: John Gore, University of Washington.

No disclosures.

MEMBER FIELDS: Karen Fields, Moffitt Cancer Center.

No new disclosures.

MEMBER HAMMOND: Elizabeth Hammond, University of Utah and Intermountain Health Care.

No disclosures.

MEMBER LOY: Bryan Loy, Humana.

I have no new disclosures.

MEMBER PFISTER: David Pfister, Memorial Sloan-Kettering.

No new disclosures.

MEMBER ROSS: Pat Ross of Ohio State.

No disclosures.

MS. BOSSLEY: So, since our general counsel is not here, I'll just ask the question that she always asks: Is there anything that your colleagues have disclosed that in any way you'd like to discuss or have additional questions on, any concerns?

(No response.)

CHAIRMAN LUTZ: Disclosures on the phone?

MS. BOSSLEY: Oh, yes. And then we have people on the phone. Sorry.

MS. FRANKLIN: Could the Steering Committee Members on the phone please give

their disclosures since last meeting?

MEMBER MARKS: Larry Marks, University of North Carolina at Chapel Hill.

No new disclosures since the last meeting.

MEMBER DONOVAN: Heidi Donovan, University of Pittsburgh.

No new disclosures.

MEMBER RICCIARDI: This is Rocco Ricciardi from Lahey Clinic.

No disclosures.

MS. FRANKLIN: Thank you. All right.

And with that, I think we'll move into a very quick overview of our evaluation process. So we'll move on.

Again, this is our Steering Committee Chair, Stephen Lutz, is here in the room with us as well as NQF staff: Heidi Bossley, our Vice President for Performance Measures, myself, Angela Franklin, Senior Director, Lindsey Tighe, Project Manager and

Adeela Khan, our Project Analyst.

As you're aware, we completed our in-person meeting for our Phase 1, at which time we had 27 measures for review and those measures primarily addressed hematology, melanoma, prostate, lung, oncology cancers as well as palliative care.

Today we begin our work on Phase II. We currently have 18 measures in front of us for review and we'll be addressing breast and colorectal cancer at this time.

The four major endorsement criteria are:

Importance to measure and report, intended to measure those aspects with the greatest potential of driving improvement;

If this criterion is not passed, the other criteria are less meaningful, so this is your must pass criteria, or one of them.

Next we'll look at scientific acceptability of the measure properties. And

the goal here is to make valid conclusions about quality. If a measure is not reliable and valid, the risk of improper interpretation in the field is great. This is also a must pass criteria.

Then, if the measures pass these two, we move on to look at the useability of a measure and the goal is to use it for decisions related to accountability and improvement. If a measure is not useful, we probably do not reach the feasibility assessment.

Feasibility is our last criterion. Ideally, we want the measure to cause as little burden as possible in the field. If the measure is not feasible, we should consider alternative approaches.

If a measure as a whole is considered suitable for endorsement, we'll evaluate the measure if it needs to be harmonized and determine if other measures in the portfolio need to be evaluated and choose

a best in class measure.

Looking at new versus endorsed measures. All measures new and endorsed are expected to meet current criteria and guidance. Our endorsed measures are expected to present data from the implementation of measure as specified in 1b of our form, Opportunity for Improvement. There also potential for reserve status if we feel like a measure the gap has narrowed, has topped out, but there's a possibility to put it into reserve status if we feel like we need to bring it up and continue to measure on it if the gap widens once again.

Reliability and validity testing. We're also looking for endorsed measures at the reliability and validity testing to be expanded unless it meets the high rating.

Useability of the measure. We want to see actual use in public reporting and other accountability and improvement programs or specific plans and a timeline for use.

For feasibility, we want to see if there were any problems with implementation or unintended consequences as the measure is implemented.

So, in front of you, you have our generic rating scale that we've been using. We're looking at 1a High Impact, 1b the Performance Gap as mentioned earlier, Usability and Feasibility.

Importance to measure and report, I think we walked through that earlier.

High impact indicators as a national health goal or priority. There's data on numbers of persons affected, high resource use, severity of illness or consequences of poor quality.

For the gap in 1b we're looking for data demonstrating considerable variation and performance or overall less than optimal performance. And we're also looking for data on disparities in care and the potential for reserve status where endorsed measures can be

assessed at this point.

Moving onto 1c Evidence, we're looking at quality, quantity and consistency of the body of evidence.

Again, individual Committee Members have rated the measures based on the evidence submitted. As part of the Steering Committee process we allow you to let us know if you are aware of additional evidence that could be presented. And we would continue to evaluate the measures on all remaining criteria.

After our work group discussions, if we're confident of the evidence presented by the Committee Members and the measure is likely to meet criteria for high impact and scientific acceptability, we'll look at that. And we could also ask the developer to provide additional evidence for consideration.

Here we have our evidence decision logic. And we've also included in your packets a quick guide that you can also reference as

we go through the meeting. And if we feel like there's an exception, if the Steering Committee as a whole feels like there's basis for an exception to our evidence subcriterion 1c, here's our decision logic.

For an outcome measure, there's a rationale that supports a relationship of the health outcome to at least one health care structure process, intervention or service. And then if it's a process or other type of measure, we'll look at if there's no empirical evidence, we'll look at whether expert opinion is systematically assessed, with agreement that the benefits to the patients greatly outweigh potential harms. So we can invoke the exception in that case.

So, here's some additional considerations for the exception.

The impact and opportunity for improvement; that is a performance gap must be met. There should be a strong rationale. The proximity to the desired outcome should be

that performance measures for distal structures and processes may be less likely to drive significant improvements.

If there's a measure of a more proximal process or intermediate outcome and it linkages is our outcome, it's probably not necessary.

And distinguishing between something important to do in the clinical process and things that are important to devote resources to for a national performance measure.

So as reviewed earlier, we're looking at the scientific acceptability of measures. We'll be looking at the reliability and validity. Reliability, looking for precise specifications on whether testing has been done at the data element or measure score a level. For validity we'll be looking at specifications that are consistent with the evidence. A validity testing that's showing at the data elements, a measure score showing

results there.

We'll look for justification of the exclusions, a risk adjustment, identification of differences in performance and comparability of data source and methods.

So, evaluation of the scientific acceptability is here shown to you in a graphical context. And again, you'll also have your quick guides.

I think we've run through the useability piece. Let's see, so I will breeze through that one.

And then feasibility. I think we talked about this earlier. The extent to which required data readily available, retrievable without undue burden and can be implemented for performance measurement. And there you have your subcriterion.

So when we reach the end of our review of each measure, where there's a measure in the portfolio or in front of us today that is related, we will assess both

measures to see if the specifications are harmonized or, if needed, differences in the specifications are justified.

Then we'll look at measures to see whether they're superior to competing measures. That is, they're more valid or efficient way to measure an issue or if multiple measures are justified. So we could reach that conclusion as well.

And here's our logic for related versus competing also in your quick guides. And we'll go through this logic as we go through any measures that meet this criteria. So I will move on, because I think we have a few of these. And we'll focus on that as we get to those measures.

So with that, didn't want to take up too much time there, I will turn to Dr. Lutz, who is our Chair. And we can begin consideration of our candidate measures. First measures are best cancer measures.

CHAIRMAN LUTZ: Okay. Welcome

back, everyone and looking forward through to getting through these 18.

The only thing I say in terms of procedure, obviously we have Heidi, Larry and Rocco on the phone, so if they turn up their name cards on their sides, we're not going to see them. So in between every few comments I'll just ask you guys on the phone if you have anything you want to add, because I hate to make you have to go last all the time because we can't see you with your cards up.

Going along with that, I guess Larry, if it's okay with you, I think our first one is 0219: post-breast cancer surgery irradiation.

MS. TIGHE: He may have had to jump off just for five minutes, but what we could do is ask ACS to tee up the measure.

CHAIRMAN LUTZ: Okay. If ACS is willing and able, let's do that.

MS. TIGHE: And I guess also we should explain the process to the developers.

When your measures are being discussed, if you want to join us at the side tables here, there's a microphone that you can speak into.

CHAIRMAN LUTZ: Larry, are you back now?

MEMBER MARKS: Yes, I'm back.

CHAIRMAN LUTZ: Hi, it's Steve. How are you doing?

MEMBER MARKS: Hi, Steve. I'm fine, thank you. Yourself?

CHAIRMAN LUTZ: Great. And you know, the only thing that would make the morning better is to hear your voice describing 219 for us because we are starving for it.

MEMBER MARKS: You're starving for 219.

CHAIRMAN LUTZ: I think ACS folks maybe are going to give us a little segue in and then you'll be up.

MEMBER MARKS: Okay. That's good.

Thank you.

CHAIRMAN LUTZ: Yes.

MR. STEWART: Good morning. Being my first time around here.

This is a measure that we originally submitted to NQF and had reviewed back in 2006/2007 and received endorsement. The measure itself has not been respecified or modified in any form since that original review process was undertaken.

We have taken in to account some of the comments that were made during the telephone conference call sessions and corrected some of the denominator conditions. So, hopefully, those shouldn't be of concern at this point.

I don't know what else you want us in the role of developer to comment on at this point.

MEMBER MARKS: Can you specify, did you change the business about the DL negative and DL positive?

MR. STEWART: We did three things to this measure. We removed the ER -- the hormone receptor status condition.

We also clarified, I think it was there was an over-specification in the tumor stage requirement. Both of these were just clerical process errors as we moved all of our documentation into the online forms that NQF were supporting. It was a click issue on our part, not a fundamental problem with the measure specification.

MEMBER MARKS: Okay. So there's no level of inconsistency in the denominator statement and exclusion; that's what that was, I think.

MR. STEWART: That's correct. That shouldn't be there anymore.

MEMBER MARKS: Okay. Okay. I'm happy to speak now if that's okay, Steve?

MS. FRANKLIN: Yes. This is Angela.

Dr. Marks, if you could just take

us through the importance criteria, importance to measure and report?

MEMBER MARKS: Okay. So radiation therapy post-lumpectomy for breast cancers is considered standard. Actually, in the majority of patients, and certainly in the cohort of patients that are included in the denominator for this measure, this has been demonstrated in meta-analyses to improve overall survival of these patients and most guidelines recommend this as a standard treatment for patients post-lumpectomy. And so this is important. It's not a direct measure of outcome, but it is an importance measure of quality of care. So I think it does meet that criteria for the importance measure.

MS. FRANKLIN: Thank you.

Are there any other comments from the work group members on this?

Comments from the larger Steering Committee? And we're looking at 1a, High

Impact.

MEMBER PFISTER: So just to clarify: so as the measure is now with the modifications, is it receptor status is no longer specified, and patients with T1a and T1b disease are all considered to be stage 1 category and they get radiation?

MR. STEWART: That's correct. Yes, on both those counts that's correct.

MEMBER MARKS: We have in front of us on the website that I just pulled up -- let me see if this is modified from the one we had a few weeks ago in our phone conference call.

MR. STEWART: Yes.

MEMBER MARKS: Okay.

CHAIRMAN LUTZ: Okay. I think we can go ahead and vote on that 1a. Okay.

So go on. I'm sorry. Go ahead and take us through to see it. Go ahead.

MEMBER MARKS: Well, this is actually the opportunity to go through 1a and then go through 1b and go through each of

them.

CHAIRMAN LUTZ: You might as well just go ahead and go right through, please.

MEMBER MARKS: Okay. So there is some evidence that there is evidence that there is need for improvement. There are some studies demonstrating that radiation is not routinely delivered to this cohort of patients, so there is opportunity for improvement.

I don't know firsthand the data on disparities by race. Basically, the submitters say there is data, I believe there is data that they may want to speak to that. But there certainly is data, broadly speaking, that there is room for improvement.

Going through to reliability and validity. It should be relatively straightforward to measure, since whether you're getting or not getting radiation I guess is -- there's evidence from billing codes and those sorts of things.

The question I have here for the developer is it the surgeon who is being judged on this, or the medical oncologist, whether or not they refer the patients to the radiation oncologist, or is it the radiation oncologist that could be viewed as being judged on this? If that could be clarified for me, that wasn't clear.

MR. STEWART: This measure was developed and has been implemented to hold -- to make the accountable unit the hospital or the treating facility. So, in a sense, both the surgeon and the radiation oncologist are being held to account because they presumably coordinate that patient's care.

MEMBER MARKS: You're saying it's on a facility basis, correct?

MR. STEWART: Correct.

MEMBER MARKS: Interesting. Okay. Okay.

CHAIRMAN LUTZ: Bryan?

MEMBER LOY: Thank you.

Could you elaborate a little bit or help us understand how you arrived at 365 days? I'm just wondering where that length of time came from, versus a shorter period.

MR. STEWART: So back when we originally did the specification work in 2005/2006, we did a significant amount of data evaluation looking at elapsed time between our index date being date of diagnosis and the date of onset or beginning, start of radiation therapy. We looked at that distribution with some care.

At that point in time, one of the driving considerations was that these measures be developed in such a fashion that they could be equitably applied across as broad a spectrum of institutions as possible. And so one of the areas of sensitivity was picking or identifying a relevant time in which you would expect most patients to start their radiation therapy. And in looking at a number of cut points, we determined that 365 days or one

year from diagnosis was appropriate, because we had to take into consideration other intervening treatment modalities that may be administered post-surgically, and there are other potential reasons for delays in the sequencing of therapy for these women. And so 365 was identified at that point as a reasonable metric for timing of onset of radiation therapy.

MEMBER MARKS: At the time the clock starts at the time of diagnosis, there often can be several weeks if not a month or two until the patient is done with their lumpectomy, they're having a re-excision, node dissection and what not.

MR. STEWART: And there's also the possibility that there is a chemo regimen that could follow that surgical event.

MEMBER MARKS: Right.

MR. STEWART: And so pushing the radiation date out made perfect sense at that time.

CHAIRMAN LUTZ: And there any other questions or thoughts, anyone else on the phone, either Heidi or Rocco, anyone have any questions for the developers?

MEMBER DONOVAN: I don't have additional questions, no.

CHAIRMAN LUTZ: Okay. We're going to move on to a vote that quickly? All right.

MEMBER MARKS: We're going to be setting the trend for the day.

CHAIRMAN LUTZ: Well, you could be a hard act to follow, Larry, we don't know.

MS. KHAN: Does everyone have a voting clicker? Okay.

Well, when the clock starts, you can press the button.

So we're going to be voting on 1a impact. It addresses a specific national health goal or priority or the data demonstrated a high impact aspect of health care. So you're going to vote one for high, two for moderate, three for low and four for

insufficient evidence.

MEMBER EDGE: When does the clock start?

MS. KHAN: Right now. You can start now. We have high impact for this measure.

MS. BOSSLEY: We can actually stop it. The big issue now is the percentages and we usually do numbers. Is it a quick fix that you can do? Okay. We'll calculate it later. Clearly, it's high. And then several moderate.

So we're going to vote on importance to measure, the performance gap. 1b, performance gap, the data demonstrated considerable variation or overall less than optimal performance across providers and/or population groups and disparities in care.

So we're going to again vote one high, two moderate, three low and four insufficient. You can start voting.

So we have 86 percent for moderate, seven percent for high and seven percent for low.

And voting on evidence. Again, if it's a health outcome with a rationale, you're looking at the quantity, quality and consistency of the body of evidence. So you're going to vote one for yes, two for no and three for insufficient evidence.

MEMBER MARKS: I'm sorry. We're voting on, is this for health outcome?

MS. KHAN: You're just voting on the evidence piece.

So we have 93 percent for yes and seven percent for insufficient evidence.

So we can move on to scientific acceptability.

MS. FRANKLIN: Okay, Dr. Marks, if you could --

MEMBER MARKS: Yes?

MS. FRANKLIN: Okay. Hold on, sorry.

CHAIRMAN LUTZ: You went through so quickly and efficiently they thought there was still more to discuss. We're still voting.

MS. KHAN: So looking at reliability. We're looking at the precise specifications and the testing. We'll vote one high, two moderate, three for low and four for insufficient evidence.

Dr. Ricciardi, if you could send your vote in.

So you have 71 percent for high, 29 percent for moderate.

MEMBER MARKS: I do have a question, this is Larry Marks, for the developer, if I could right here. What is the threshold for this? Because certainly there are patients who are 65 with comorbid conditions where it would be reasonable not to do the radiation. So is the expectation that this would be 100 percent, or is there a way of excluding patients from the denominator who are deemed not to be medically appropriate for radiation?

MR. STEWART: We have not chosen to include any comorbid condition consideration

in this measure. We have simply followed the randomized clinical trials evidence that established an age cutoff at under 70.

MEMBER MARKS: Thanks. And what is the threshold of expectation or is that sort of dropped? Did you figure it out?

MR. STEWART: Well, quite independently, through other processes, the Commission has recently established performance thresholds for this measure across its 1500 programs where we are anticipating, we're expecting at least a 90 percent threshold to be met, understanding fully that there are a vast majority of institutions that will easily exceed that expected rate.

MEMBER MARKS: Okay.

MS. KHAN: And moving on to 2b, validity. That includes the specifications are consistent with the evidence, they're looking at the testing, exclusions, risk adjustment, meaningful differences and comparability between data sources.

So again one high, two moderate, three low and four insufficient evidence.

So you have 53 percent for high, 40 percent for moderate and seven for insufficient evidence.

And moving on to usability. We're looking at meaningful and understandable use for public reporting and accountability and is it useful for quality improvement.

So, one high, two moderate, three low and four insufficient information.

We have forty percent for high and 60 percent for moderate.

And moving on to feasibility. The data generated during care electronic sources, susceptibility to inaccuracies and unintended consequences have been identified and data collection can be implemented.

So again, one high, two moderate, three low and four insufficient information.

So 53 percent high and 47 percent for moderate.

And now voting on overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? You're going to vote one for yes and two for no.

And we have 100 percent agreement on yes, and the measure will pass.

CHAIRMAN LUTZ: All right. So next we move on to 220: adjuvant hormonal therapy. I think Joseph Laver on the phone is the one who is going to direct us through this, give us the synopsis.

I guess I should ask. Joseph Laver, are you on the phone?

(No response.)

MS. FRANKLIN: We'll go ahead and have -- well, we can move on to the next one in the process. I think Dr. Laver did say he was going to join us. We're just a tad early. So we can go on to the next one.

CHAIRMAN LUTZ: So the next one, Pat, I think we're doing needle biopsy to establish diagnosis.

MS. FRANKLIN: First, could we have the developer just give us a quick overview of 0221?

MR. STEWART: The brief overview here is the understanding, at least of the surgical community, that having a pre-operative needle biopsy prior to surgical treatment of women with breast cancer is a necessary prerequisite to understanding the disease being managed.

I think we discussed some of the nuances about this measure on the telephone conference call, and I think the commentator from the panel will raise some of those summary findings and we can address those as we move forward.

CHAIRMAN LUTZ: Okay. Pat?

MEMBER ROSS: This measure is very straightforward. It is a process measure looking at the needle biopsy to establish diagnosis prior to surgical excision or resection. As you know, the ACS Commission on

Cancer is the steward.

I think that there is value here, because of the data that has shown the needle biopsy is at least as accurate as surgical biopsy. And the value, the importance really goes to what impact it can have on improving quality of care, on improving quality of the surgical procedure and there may even be some cost/benefit, cost/effectiveness components to it as well.

I think the developer does a great job in elucidating all of the components. There's one question on the disparities by population group, which I think they've raised the issue that age, race/ethnicity, geography as well as details about the individual providers all account for the disparities, which I think are probably significant regionally.

And I think this is -- the evidence is observational studies. I think that this is something that is of value and

will be easy to measure.

One of the limitations is the fact that this is not a technique which would be available everywhere. There is a user component to it in terms of successfulness accomplishing the task. But I think that it is something that will be straightforward, it will be easy to measure and it will in fact impact the quality of care for the patients requesting it.

CHAIRMAN LUTZ: Was there anyone else in the subgroup that had the phone conversation about this that wants to chime in?

Okay. Elizabeth?

MEMBER HAMMOND: On the phone I raised two questions. One was whether or not this measure is valid in rural areas where needle biopsies may or may not be appropriate? And second, should the measure be stratified by cytologic versus needle biopsies which have different value in this sort of setting?

MR. STEWART: I think I can answer both of those questions.

In response to the first, we understand the sensitivity around rural settings. Unfortunately, the Commission on Cancer has accredited programs where we essentially have our implementation forum. About one percent of our programs are placed in purely rural counties, and about 12 percent of our programs are in urban non-metro counties when we look at the distribution and geographic placement of those. So it's hard for us to comment explicitly on the question of rural settings.

In contrast, however, we do have access to services and resource data from these institutions. Eighty percent of our programs have diagnostic imaging available to them, and the other 20 percent provided by referral.

So even in locations where these sorts of procedures are not readily and

immediately available, patients are referred to institutions or settings where that's provided to them.

The second point you raise about cytology versus core needles is a very subtle distinction. Unfortunately, the Cancer Registry data sets that we work with routinely confound those two and we don't make them distinct and separate. And this has been the primary concern of ours and has delayed our implementation of the measure into the field. So we're sensitive to that and that's largely why we have maintained this measure over the past four or five years but not implemented across our settings because of the way the data are organized that we work with on a routine basis.

CHAIRMAN LUTZ: Karen?

MEMBER FIELDS: So, I would like to comment from the surgeons in the room about core biopsies because that would still be our gold standard that we want to move to, so why

wouldn't we create a measure that works towards getting to that end point?

CHAIRMAN LUTZ: Stephen?

MEMBER EDGE: I would actually argue that we should not make any effort to make a distinction between cytologic versus stereotactic core biopsy. The vast majority of these procedures are done with stereotactic core biopsy in 2012 as opposed to, perhaps, 1998. And if a specific center is very experienced with fine needle aspiration and uses fine needle aspiration, I would see no problem with that. I think those of us who are expert in breast cancer in the field recognize the potential limitations of fine needle aspiration with insufficient material or a lack of cytologic diagnoses. But if the program is very experienced, I would not hesitate to endorse that program's use of fine needle aspiration.

I think the benefit of getting that additional granularity of information is

outweighed by the benefit of getting the information that people are doing needle biopsy in the first place. So, I would actually argue against concerning ourselves with this nuanced distinction in a quality measure.

CHAIRMAN LUTZ: Karen?

MEMBER FIELDS: One more issue about all of these measures is the data is from 2007 and 2008 for all of us to use for these measures. And I wondered if we saw any improvement or increased acceptability, because I do think that the general knowledge about needle biopsies before surgery has increased in that time period. So, did we have any data to compare or any trends, because I think that helps us to understand if this is also a valuable measure?

MR. STEWART: Yes, there was a paper published last summer following a presentation at the Surgical -- at SSO the prior March that described increased patterns

in preoperative needle biopsy for this cohort of women. And I can find that citation and forward it to the NQF staff.

MEMBER EDGE: Is that the --

MR. STEWART: It's Dr. Williams' paper. That paper is looking at the National Cancer database. It's referenced in your materials from 2003 to 2008. So it doesn't really address Dr. Fields' question.

MEMBER FIELDS: My question is: it's so much more a part of the diagnostic workup than it was even at the time that this measure was first proposed; do we still have a problem? That was my question, because we're endorsing a lot of measures here and I'm trying to decide if there's a national problem, do we see any evidence of improvement? That was my question.

MR. STEWART: I think, for better or for worse, all these data systems suffer from some degree of lack of currency. So in 2008/2009 -- for me in my world, 2010 is as

current as I see things and can assess them. And I don't have that data at my fingertips right now.

CHAIRMAN LUTZ: Can I ask a question similar to what Dr. Marks asked in the last one? Is this meant to be a never event or is this meant to be something where someone deviates greatly from, you know the norm that it's an issue? Because one of the reasons I ask is the week that I started looking over our current set of these measures, I had a patient who had a core biopsy, it was negative. The surgeon, in their experience, said, this doesn't add up. They excised and it was cancer.

And so if it was a never event, this really takes that option of, boy, it still doesn't add up, I want to know and then do this.

I mean, this isn't an "if you ever do it, you're in trouble" measure is it?

MR. STEWART: No.

CHAIRMAN LUTZ: Okay.

MEMBER EDGE: Steve, in the case that you just cited that patient would be coded as having had a needle biopsy. I believe that's true.

CHAIRMAN LUTZ: Is that true? Would the patient have been coded --

MR. STEWART: If the result of the biopsy was negative, if the procedure was actually performed, we would have to recast that event.

CHAIRMAN LUTZ: Oh, good. Okay.

MR. STEWART: But not sensitive to the outcome or assessment of that event.

MEMBER EDGE: But you can expect that between 10 and 20 percent of women who have biopsy will have to have a surgical biopsy. There are technical reasons why you can't do a core biopsy; the lesion is very peripherally located and cannot be located on the mammogram, it's very deep within the breast, or it's a very small breast. So there

are technical reasons why a stereotactic needle biopsy cannot be done, and somewhere between 10 to 20 percent of women will probably have surgical biopsies. So this is not one where you can set up a 100 percent or even a 90 percent.

CHAIRMAN LUTZ: Bryan?

MEMBER LOY: I'd direct this back to I guess the surgical expertise in the room, and that would be: are we somehow creating a measure that is promoting the use of a biopsy when the surgeon believes that those results are not going to inform the ultimate decision to excise?

MEMBER EDGE: The answer is no, but there are a substantial number of cases where you do a core biopsy, particularly for microcalcifications, where the core biopsy will show a specific benign lesion, but we know from published literature that the sampling issue means that there is cancer in the surrounding tissue in somewhere between

five and 20 percent of the cases. So when atypical ductal hyperplasia is identified, that's somewhere on the order of five to 20 percent, depending on which paper you read, those women actually will have either in situ or in a few cases invasive cancer in the surrounding tissue. And so the standard is to proceed with surgical excision even though the biopsy is technically benign. That's probably the circumstances of the type of case that Dr. Lutz was outlining.

Dr. Hammond, do you have any comment on that?

MEMBER HAMMOND: No. I think that's accurate.

CHAIRMAN LUTZ: David?

MEMBER PFISTER: So just that I am clear when we go to measure this, let's say the person has their diagnostic evaluation elsewhere. And, for whatever reason, they don't do a needle, but they do get tissue so they do an incisional biopsy. But then they

end up getting their treatment done somewhere else. And then I'm at that somewhere else place and now I'm managing the breast cancer. And there would appear to be little reason to do anything before I do the surgical procedure because I clearly have tissue, but while I might have personally pursued that diagnoses differently, it is what it is. And then when they go to evaluate my performance based on how the numerator and denominator are defined, how will it be tracked when you have care divided in two different settings? Do you see what I'm saying?

CHAIRMAN LUTZ: I agree. I mean, I think for the first one we voted on today and maybe several others we're going to have today it's an issue of the system is not as well defined in some geographic areas as it is in others.

MEMBER PFISTER: Because I think that it has -- earlier was probably about the rural factor, but I think when you

particularly get to larger rural centers, lots of times the diagnoses will be made for better or for worse in terms of the process by which it was arrived at elsewhere and then where the recipient of what was kind of done at that time. And so it's unclear to me how the numerator and denominators as defined is going to distinguish cases where you are often the get-go in terms of how the person is evaluated versus ones where part of its clearly been elsewhere, you inherit a certain amount of information and then you kind of make the best of the situation even though it may not have been how you would have proceeded in the first place and how this measure actually evaluates that.

MEMBER EDGE: Well, this issue of attribution is quite difficult in many of these measures. I'm not sure, were the developers asked to specifically comment on the issue of attribution in any of these measures?

CHAIRMAN LUTZ: I'm not sure they were asked to.

MEMBER EDGE: I don't remember reading through that there's a specific issue of attribution. Maybe that's a shortcoming of the way that we asked the developers to do this.

CHAIRMAN LUTZ: Right. And I think one of the things we've learned from being on the Committee is we end up looking for any unintended consequences. So this comes up a lot, because this is one of the recurring concerns.

Before I forget, anyone on the phone, anyone have any thoughts to add, anyone have their card on their side on the phone?

MEMBER MARKS: I think just because -- I was thinking of this from before -- the radiation question from the last item very similar, right? The surgeon went to a biopsy from a surgery how do we code that to get to the liability get to the issue? It's

a huge problem; I didn't realize that.

MEMBER PFISTER: I understand what you're saying but I see that as a slightly different permutation in the sense that there, I think there's little argument that something's going to get done and that the measure is evaluating whether radiation is done within a certain period of time.

Here, the person who would ultimately potentially would be subject to measurement based on this metric is going to potentially modify how they might proceed based on information they inherit. And I guess, at least in my mind, it seems to be a slightly different issue of attribution.

MEMBER MARKS: I recognize that this is different, but it's similar as well, right? But if one queries the database from that facility, you know not having record of a prior needle biopsy, so for that case that facility might be deemed not in compliance when indeed the patient did have a biopsy.

MEMBER EDGE: But the way the Cancer Registry is now structured, however, that Registry would say that the patient did or did not have a needle biopsy and it would say where the original biopsy was done. It would say the original biopsy was done at the reporting institution or was done at another institution and would have a date when the biopsy was done.

MEMBER MARKS: Oh, okay. Is that captured in these registries?

MEMBER EDGE: Can Mr. Stewart comment on that question?

MR. STEWART: I'm sorry. The person on the phone, the question was what again?

MEMBER MARKS: I was asking whether registries do indeed capture that information about a prior biopsy.

MR. STEWART: Yes. Yes. So there are a couple of considerations here.

One is that Cancer Registries by a

whole set of other rules and regulations are obligated to have tracked down that information if that's available.

They also have the ability to distinguish the combination of where certain events took place. And this is something I have not looked at for this particular measure. But we can distinguish between patients who were diagnosed elsewhere and treated at the reporting institution or diagnosed and treated at the reporting institution to understand what the relative balance or dynamic of that data look like to understand if the denominator needs to be fine tuned around those sorts of considerations, if that begins to address the concern on the table.

CHAIRMAN LUTZ: Well and Heidi points out, I think that the denominator statement says diagnosis and all or part of first course of treatment performed at the reporting facility. And so maybe that would

leave --

MR. STEWART: I think that does address the question from the other side of the room where --

CHAIRMAN LUTZ: Right.

MR. STEWART: -- we're only looking at patients whose entire encounter for the diagnoses and management of their disease happened inside the walls of the reporting institution and we don't have a problem with patients moving between hospitals here.

CHAIRMAN LUTZ: Karen?

MEMBER FIELDS: I was just going to comment earlier but it's an extension of that. Perhaps the wording in all of these needs to be, you're reporting your analytic cases where you have all of the responsibility for tracking down, and then you're attributing it to that -- you're not attributing it to any one person but you're tracking down the analytic cases for which that institution takes responsibility. Because even if you're

going to say part or all of their initial therapy, patients move around and it would be very difficult to get this data if you didn't say something in all of these like analytic cases.

MR. STEWART: I think you'll find in the measures that we'll talk about today, this one and then one tomorrow around colon disease where we know that it's either basically a single-modality intervention that we're trying to capture and evaluate, we close those parameters to make sure that it's all happening within the reporting institution and that's our accountable organization or agency.

When you move into the multi-module therapies such as the conservation surgery and radiation measure that we just discussed, we're not sensitive to the fact that we want to look at only analytic cases within a reporting institution. We're concerned about the continuity of care for a patient, and so we're patient-centric in that

sense. And we're very ecumenical about making sure that if surgery is done in institution A and radiation is done elsewhere, both institutions are being watched to be accountable for the continuity of that care for that patient.

CHAIRMAN LUTZ: I was just going to add one aside. It might be too far astray, but one thing this doesn't help control, and I've seen this in three geographic areas and heard about it in others, are places where surgeons overdo their diagnosis.

So I actually have worked with -- there are surgeons who do an FNA, it's positive. Then they do a core. Then they do an incisional. Then they do an excisional. Then they do a re-excision. Then they do a sentinel lymph node biopsy. Then you do an external lymph node dissection. And so I think one of the things you have to keep in mind is that surgeon is doing great with this. They are doing 100 percent. They will always

have some -- you know, it sounds funny, but actually you know a busy practice in Memphis, small rural area in Ohio I've seen this and I've heard about it from friends around the country. It's not -- again, we practice usually in bigger centers where we see good care. There's a lot of things -- first in reading through this, I thought well there's a lot of folks that may look good when they're not.

Dave?

MEMBER PFISTER: I am a little confused by that discussion prior to your comment. The way that the numerator and denominator is currently specified, any further descriptions, say, that, let's say it's limited to people that were -- you know, had everything done at one institution. That is not the case. It's as specified as it is, which would mean that people that were diagnosed at one place but then managed elsewhere are all part of this denominator.

Like there's no further descriptor analytic cases, only the institution cases, et cetera.

MR. STEWART: No. If you read the denominator statement it says diagnosis and all treatment at the reporting facility. There's a linguistical trick here. In my world, an analytic case is more than just that, it may lead to other characteristics. This is actually the subset of what I consider to be an analytical case.

MEMBER PFISTER: So you're saying that the diagnosis -- so what you're saying is that the diagnosis --

MR. STEWART: Both the diagnosis and the treatment have to have occurred at the reporting institution.

MEMBER PFISTER: Okay.

MEMBER FIELDS: It says first course of treatment. So that means just the surgical treatment?

MR. STEWART: No. First course treatments means everything to manage that

diagnosis until the time of recurrence or disease progression.

MEMBER PFISTER: But in most circumstances, Steve, that would be surgeon, right, in terms of first course of treatment?

MEMBER EDGE: Yes.

MEMBER PFISTER: Like, I would say 95 plus percent of the time surgery is going to be the first thing.

CHAIRMAN LUTZ: Anybody else on the phone have a comment?

(No response.)

Any other discussion or we moving on to vote? Looks like we're voting.

MS. KHAN: So voting on 1a, impact. Again, addresses a specific national health goal or priority or the data demonstrated a high-impact aspect of health care. So one high, two moderate, three low and four insufficient.

So we have two high, 13 moderate and one insufficient evidence.

And moving on to performance gap, the data demonstrated considerable variation or overall less than optimal performance across providers and/or population groups. One high, two moderate, three low and four insufficient.

So you have three high, 12 moderate, one low and zero for insufficient.

And going on to evidence. It's one for yes, two for no and three for insufficient evidence.

And that's 14 yes, one no and one insufficient evidence.

So going to reliability. We're looking at precise specifications and the testing. Again, one high, two moderate, three low and four insufficient evidence.

And four high, ten moderate, two low and zero for insufficient evidence.

Looking at 2b, validity. Again, looking at specifications are consistent with the evidence, testing, exclusions, risk

adjustment, meaningful differences and comparability in data sources.

So one high, two moderate, three low and four insufficient evidence.

Can we have everyone press their button one more time?

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

And we moving on to usability. We're looking at usability for public reporting and accountability and for quality improvement.

So, one high, two moderate, three low and four insufficient information.

Can we have everyone do it one more time?

Four high, 10 moderate, two low and zero insufficient information.

Going on to feasibility. We're looking at the data generated during care electronic sources, susceptibility to

inaccuracies and unintended consequences are identified and data collection can be implemented.

Again, that's one high, two moderate, three low and four insufficient information.

One more time. Again, the receiver is actually over here, so if you want to point your clicker over here. I think it's fine. I got them all.

So we have three high, ten moderate, three low and zero insufficient information.

And overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? One yes, two no.

Dr. Laver, are you on the line now?

(No response.)

So we have 12 yes and four no. The measure will pass.

CHAIRMAN LUTZ: All right. So just

double checking, Dr. Laver is not on yet, right? Okay. Then we will skip forward to 559. We'll have the developer frame things for us and then Jennifer just came on in because she had a desperate need to tell us more.

MR. STEWART: So is this the combination? 559?

This is a measure looking at multi-modal management of appropriately staged hormone receptor negative breast cancers for women under the age of 70 with the expectation that using diagnosis date as the index reference point that combination chemotherapy be started or initiated within four months or 120 days of diagnosis.

I don't know that there was much commentary or requests for clarification during the telephone conference calls. I would like to have the commentator pick it up from here, and I'll be happy to answer questions as they arise.

MEMBER MARKS: I'm sorry, are we on 559 or 220?

CHAIRMAN LUTZ: We're on 559. The person who is going to present 220 is not on the line yet, so we skipped forward to 559.

MEMBER MARKS: Thank you.

MEMBER MALIN: So I think this measure, you know, is probably one of those measures that has reams if not the most data behind it. It's one of the ones with the most data behind it in terms of evidence that it improves patient outcomes.

I think clearly it's important, this is high-impact. I would say it's been in use for a long time. There's ample data on its reliability and validity, feasibility and usability.

I would say probably these are more kind of general concerns, the necessary concern specifically about the measure is that at this point it's pretty dated. It's not necessarily -- you know, we should probably

strive to have measures that keep up with the current nuance in breast cancer treatment and providing good breast cancer care is more than just providing chemotherapy generally. And so, you know, I would encourage the developers to think about ways to maybe improve upon this going forward.

And then the corollary of that is, I think, because this is such a generic mom and apple pie measure, most of the data out there suggests at this point that there's not a lot of gaps in care related to this measure.

Any questions?

CHAIRMAN LUTZ: Bob, you had your card up early on this one.

MEMBER MILLER: So, my question is the verb "considered." How is considered tracked in the medical record?

MR. STEWART: So the registry coding systems allow and provide opportunity for the capture of information describing the fact that physicians or attending physicians

responsible for the patient's care did one or a number of things. Either documented it in the medical record that the treatment or the chemotherapy in this case was appropriate but there were other extenuating circumstances, patient's overall other health condition. what not, that recommended care was simply not-- you know, the standard of care was simply not recommended for those reasons.

Also, they do capture indications that that consultation occurred and the patient or their guardian declined the therapy that the physicians recommended to them and so forth.

So, there are probably about three or four different ways that a generic umbrella of considered is captured and reported through these systems.

MEMBER MILLER: So are those elements coded in some standard fashion?

MR. STEWART: They are. They are.

MEMBER MILLER: Okay. Because I

guess that would be my concern, is: how do you really know if something was considered? If wasn't documented, it wasn't done. You know, I'm just thinking of my own practice, you know I don't code this way, but I can see easily how a decision was made not to give chemotherapy after an extended discussion. If it's not abstracted properly from the written or the electronic medical record, you're not going to see that. So I wondered about just about the consistency of application. But I understand your explanation. I wasn't on the small work group on this one, but do you present data that shows that the consideration you said has tested, that it's reliable?

MR. STEWART: We do that in two ways. One is that we actually indicate in our report-back mechanisms to the hospital what their quote/unquote "considered rate" happened to be so that they can identify themselves as whether they were either low or high outliers in that regard.

Secondly, during the accreditation site visit we actually have peer reviewers examine selected medical charts and we actually target nonconcordant and charts where it's indicated that considered therapy was not actually given so that we can verify that that was actually documented in the medical record. So we do external objective validation checks of that reporting information.

CHAIRMAN LUTZ: Karen?

MEMBER FIELDS: So would an appropriate exclusion criteria be patient declined? Because that's not one of the exclusion criteria.

MR. STEWART: No. If the patients are advised that chemotherapy is recommended for their condition and they decline it, that case appears in both the numerator and the denominator. We're interested in making sure that clinicians and medical systems are cognizant of this particular standard of care and are documenting the fact that even if the

patient doesn't actually receive or have the chemotherapy administered, that they had made the choice not to do so. We want to make sure that the physicians who are responsible for that patient's care are quote/unquote "doing the right thing at the right time" even if the patient subsequently declines.

MEMBER FIELDS: And do you also capture lost to follow up, I assume, then too?

MR. STEWART: Lost to follow up in the sense of?

MEMBER FIELDS: Well, declining in some of these populations is lost to follow up because the women that would be likely to decline might seek alternative therapies, you might not have that --

MR. STEWART: I don't think it's that nuanced. The data that are reported through the registries simply signal administration or lack thereof. And if it's not administered and there's evidence in the medical record for why that wasn't done, and

it fits the appropriate considered criteria, that's how it appears.

The fact that the patient may go elsewhere for alternative therapy or intervention isn't something that we would pick up as a matter of course.

MEMBER PFISTER: As Larry was saying, this has been sort of a heavily vetted measure. So there's like, you get vetting fatigue after a while. So at the risk of saying that, how do you know that they didn't get crazy combination chemotherapy?

MR. STEWART: We don't. We distinguish between single agent and multiagent. But what that combination happened to have been is not something that's been standardized to this data collection mechanism.

MEMBER PFISTER: Because, you know, clearly there are things which would be viewed as kind of fairly mainstream and acceptable combination chemotherapy to give

here. I know when I was involved in a practice guideline in lung cancer several years ago that there is in fact wrong combination chemotherapy to give. In fact, people seemed like they did worse with the wrong combination chemotherapy and it seems -- you know, again, you might say well, gee, 95 percent of the time they're getting a reasonable thing so it's going to come out in the wash. But it seems that at least what drugs they get that that should be -- you know, that should be accessible information electronically. And I'm just thinking about like raising the bar in a measure like this that's been heavily endorsed. You know, I think raising the bar a little bit would be a reasonable expectation.

CHAIRMAN LUTZ: We'll do Elaine and then Karen and then check on the phone.

MEMBER CHOTTINER: Okay. Going back to this process of looking at exclusions. I think that what you're describing is very

cumbersome and to rely upon people going back to the chart and pulling out reasons why patients didn't get chemotherapy is very difficult, especially if this is going to be incorporated into one of the PQRS measures it would be difficult to report the coding. And I think it would be much better if you do have a category for patient refusal or comorbidities or something that would give us an easier way to pull that information out.

CHAIRMAN LUTZ: Karen?

MEMBER FIELDS: How do you capture neoadjuvant therapy and staging then?

MR. STEWART: We capture dates of service so we know whether or not the chemotherapy is being provided neoadjuvantly. And we also capture both clinical and pathologic staging information. So I think we have those considerations accounted for.

MEMBER FIELDS: That's fine. Because staging is no longer pathologic staging.

MR. STEWART: No. This is no longer pathologic staging.

CHAIRMAN LUTZ: Okay. Heidi, Larry, Rocco, anyone on the phone?

MEMBER MARKS: Yes. I'm sorry. I stepped away for a few minutes, and maybe this was addressed. Is the goal again 100 percent, because the same issue applies about the comorbidities and what not?

MR. STEWART: Again, consistent with my earlier comment, the Commission is setting a bar of 90, knowing that there will be some flexibility in the way that we look at these data, but we will expect institutions to be able to demonstrate at least a 90 percent concordance knowing that 100 percent is likely but not always going to be observed.

MEMBER MARKS: Do we know 90 is a national number for this one? Also the radiation one, for that matter. What percent of patients have comorbidities that would prevent the delivery of radiation or chemo?

I don't know the answer, but maybe someone does.

MEMBER MALIN: Also, I would think 90 would be kind of a low bar. This isn't receipt of chemotherapy, it's consideration of it. So it should be close to 100 percent.

MEMBER MARKS: Yes, that's true.

MEMBER MALIN: It means you didn't do your job if you didn't consider it, at least.

MEMBER MARKS: This is less stringent than the radiation one where it was actually delivery of radiation.

MEMBER EDGE: I think there is a couple of differences here. A couple of points here.

First of all, Larry, this measure, unlike the radiation measure, the patients with comorbidity, as Mr. Stewart outlined, are included in the numerator as having received concordant care. If the doctor said, "I understand that this person would generally

receive chemotherapy but because of these comorbidities they should not," and they are considered concordant and would be in those patients who would be positively considered for this measure.

MEMBER MARKS: Okay.

MEMBER EDGE: The second issue is that, again, I believe the developers were not asked to set a threshold measure for us to consider, nor were we looking at attribution.

As Mr. Stewart said earlier, the Commission on Cancer has separately, for the purposes of its accreditation program for cancer programs, has set a standard of 90 percent and if centers fall below that, they have to develop a written action plan and demonstrate to us on our site visit surveys that they have acted on it. But those have no bearing on our deliberations here, is my understanding.

And I would agree with Dr. Malin it's a relatively low bar, but again it is

completely separate from our discussions here. The Commission set that as a place to start to say we need to meet this standard, and there was a lot of discussion of whether it should be 85 or 90 or 95 or 100. But since it's never been done before to set this kind of standard on a national level, we started at 90.

But that really has I think -- that level of expected concordance has no bearing on NQF discussions because I think the developers were not asked to present that kind of information.

MEMBER MARKS: All right. Thank you.

MS. BOSSLEY: This is Heidi. Maybe I should add a little clarification as to exactly -- you're right. We don't specifically ask for benchmarks. And it's been something that the committees have tried to determine should there be.

I do think it's interesting when

you look at the reliability results here, you do provide some data from '07 and '08, and that may help to answer some of the questions. And it looks like cancer programs back then in the 75th percentile had performance of 100 percent. So it at least gives you a sense of where everyone is.

It appears to be, again, that's four years ago, but fairly high. So I'm not sure that a benchmark in this instance actually would be needed because it looks like it's actually high. But I think you all need to talk that part through. Based on the data you're seeing, it is rather high. There is some variation, but again I think that's the question in my mind that probably should be answered.

CHAIRMAN LUTZ: Anyone else have comments or thoughts?

MEMBER DONOVAN: I do have some questions about the reliability data that was presented. So performance ratings that are

so high and reliability testing that, to me, it doesn't look like it really addresses the extent to which people are able to accurately extract information on this consideration variable. It seems impossible that we can weigh performing services more than issuers' reliability than their performance. So, that's one question.

And then the other question is: is there a precedent for how to handle sort of longstanding measures that seem to need to be upgraded or made more current, you know, when the previous measure was viewed as sort of a, as everybody said, mom and apple pie sort of measure that now seems to be sort of a measure that may start achieving and not really capturing current practice? That's a strong statement, I don't mean not capturing current practice, but not nuanced enough to catch whether the chemotherapy administered was appropriate.

MR. STEWART: And so in order of

the two questions, the response to the first question is that, from all of our work and evidence, the institutions with low-lying performance rates tends to be a reflection of completeness of information in their registry systems. And so what we've discovered is that as we put these measures into play, institutional completeness and accuracy of data have increased as institutions have paid attention to the fact that they're being watched. It's the classic Hawthorne effect. So I think I'll stop my answer at that point.

And then secondary, I think you're quite right. We suspected this at the outset that a number of the measures that the Commission and the College put forward to NQF that are being discussed again here were pretty straightforward. And in some cases, they remain that way. I think some of the suggestions for how to push the edge of the envelope and raise the bar and add additional levels of possible specificity to these

measures are probably well worded, but they'll take some time to fully assess and understand how best to do that.

CHAIRMAN LUTZ: Okay.

MEMBER DONOVAN: Has there been a precedent where there has been a formal request that the bar is raised prior to the next review or the sense that it's, you know trying to close the measure and sort of request formally that, you know, this measure be stopped and then a new one be proposed?

MR. STEWART: Is that a question for the developer or a question for NQF?

MEMBER DONOVAN: It was a question for NQF.

MS. BOSSLEY: So this is Heidi.

It's a very good question and you actually have both options on the table. So I think we should vote once you're done discussing it, see if the measure passes as it is against all the criteria.

You can put forward

recommendations on what you think you would like to see the next time around if this measure does pass the criteria. Or, it is your choice if this measure doesn't pass, endorsements removed and then there will be an opportunity hopefully in the near future that they can bring forward another measure that addresses some of the concerns in the areas that you would like. So, you have both options.

MEMBER LAVER: Can you update us, which measure are we talking about?

CHAIRMAN LUTZ: We're on 559.

MEMBER LAVER: Okay.

CHAIRMAN LUTZ: All right. Any other suggestions or thoughts? It looks like we're going onto voting.

MS. KHAN: So voting on 1a, impact.

MEMBER LAVER: So I'm not in front of a computer, so I have to have a computer to vote or --

MS. KHAN: You can just say your votes over the phone and we'll put them in for you.

MEMBER LAVER: Okay.

MS. KHAN: Do you have a vote on 1a, impact?

MEMBER LAVER: So are we doing it by phone call or --

CHAIRMAN LUTZ: No, for you. It's high, moderate, low or insufficient for impact on 559.

MS. KHAN: So we have eight high, seven moderate and one insufficient evidence.

MEMBER LAVER: I vote by pushing the buttons or how?

MS. KHAN: Dr. Laver, you can just say high, moderate, low or insufficient over the phone and then we'll capture that for you.

MEMBER LAVER: Okay. Moderate.

MS. KHAN: Okay. So it's tied eight high, eight moderate and one insufficient.

Voting on performance gap. Again, it's high, moderate and low or insufficient evidence.

And Dr. Laver, did you give us your vote?

MEMBER LAVER: I'm looking through the pages. And this is the same measure, right?

MS. KHAN: Yes, it's performance gap. Same measure.

MEMBER LAVER: Okay. I vote two.

MS. KHAN: Okay. Thank you. So we have one high, 12 moderate, three low and one insufficient evidence.

And moving onto the evidence, we're going to vote one yes, two no and insufficient evidence.

And Dr. Laver, you can just say your vote whenever you're ready.

MEMBER LAVER: Three.

MS. KHAN: So we have 12 yes, three no and two insufficient evidence.

And going on to reliability, you're going to vote one high, two moderate, three low and four insufficient evidence.

MEMBER LAVER: I'll vote two.

MS. KHAN: Can we have everyone press their number again?

So we have seven high, eight moderate, two low and zero insufficient.

Voting on 2b, validity. It's one high, two moderate, three low, four insufficient evidence.

Dr. Laver?

MEMBER LAVER: Two.

Can I ask you a question while everybody's voting? Did you discuss already the 220?

CHAIRMAN LUTZ: No, we waited just for you. We're actually going to do that next.

MEMBER LAVER: Okay.

MS. KHAN: So we have seven high, eight moderate and two low.

Gong on to usability. We're going to vote one high, two moderate, three low or four insufficient information.

And, Dr. Laver?

MEMBER LAVER: On which one now?

MS. KHAN: This is usability.

MEMBER LAVER: Three.

MS. KHAN: We have six high, six moderate and five low.

And going on to feasibility, one high, two moderate, three low or four insufficient information.

And Dr. Laver?

MEMBER LAVER: I vote three.

MS. KHAN: So we have three high, nine moderate and five low.

And overall suitability for endorsement, does the measure meet NQF criteria for endorsement? Yes or no.

And, Dr. Laver?

MEMBER LAVER: I'm debating here. So give me a second.

MS. KHAN: Sure. Whenever you're ready. So we have ten seconds left on the clock. Did you want to put a vote in?

MEMBER LAVER: Okay. I would say yes, one.

MS. KHAN: Okay. So we have 14 yes and three no. So the measure will pass.

CHAIRMAN LUTZ: All right. So we're onto 220. So we will have our developer present first and then move on to you, Dr. Laver.

MEMBER LAVER: Thank you.

CHAIRMAN LUTZ: So if the developer is ready?

MR. STEWART: So analogous to the measure we've just discussed, there are many of the same sorts of components and considerations at hand.

This is a measure that examines adult female breast cancer patients with hormone receptor positive disease and appropriate midstage diagnosis for whom we

would expect hormone therapy to be either recommended or administered --

MEMBER LAVER: Could you speak up?

MR. STEWART: -- within a 365-day time frame. I'm sorry.

Similar to the measure we just reviewed with respect to adjuvant chemotherapy, this measure examines adult women with appropriately midstaged breast cancer who are hormone receptor positive with the expectation that tamoxifen or third generation aromatase inhibitor be administered or considered within 365 days of the index date of diagnosis.

I don't think I have anything more to comment on with respect to the numerator and the denominator criteria. There were some comments raised during the phone conference call. I'll be happy to address those during the discussion as they arise.

MS. FRANKLIN: All right. Dr. Laver, if you could lead us through your

discussion of the measure.

MEMBER LAVER: Okay. Again, as I said previously, I am a pediatric oncologist so it was a stiff learning curve for me to look into breast cancer.

I reviewed the literature and there's a tremendous body of literature with high evidence and quality data that treating within 365 days is beneficial and improves survival and improves quality of life. So I for one supported the measure. I think it's a well-thought one. I think it's feasible to do. I think measuring quality of care, this is a parameter that should be measured.

I'll stop here.

CHAIRMAN LUTZ: All right. Is there anyone on the conference call about this that had anything to add?

MEMBER MARKS: Just a question about the stage, the same business about the Stage I versus II, were there some inconsistencies similar to one of the other

metrics because T1 -- I guess not. I'm not seeing that.

MS. FRANKLIN: We have a response from the developer.

MR. STEWART: I think in the denominator statement we are clear that it's a AJCC T1c for Stage II or Stage III --

MEMBER MARKS: Okay. Yes. I'm sorry, a different one I'm thinking of. Thank you.

CHAIRMAN LUTZ: Any other questions that come to mind? I think the developer has more to add.

MR. STEWART: So just to bring closure on the commentary from the telephone conference call, a question was raised whether or not we had considered the exclusion of pregnancy or planned pregnancy from the denominator of the measure.

MEMBER LAVER: Yes, I remember that.

MR. STEWART: So I promised to

look into that. First let me just caveat. There's no way we can anticipate planned pregnancies in our data sets, so that's neither here nor there.

We did look at a diagnosis of the cohort of patients in the denominator of this measure constituted just over 110,000 women, of which we identified 63 who had a secondary diagnosis code in some way related to pregnancy or pregnancy care, which constitutes one half of one percent of the denominator. Whether or not that constitutes sufficient specificity concern to exclude those women or not, I would invite comment on.

I would only go on to observe that half of those women actually did eventually show up in our data set as having received hormonal therapy for their breast cancer. So it's not clear to us at what stage in their pregnancy they were when the original diagnosis occurred, but it's plausible that post-delivery hormonal therapy was

administered to those women as would be appropriate, I presume.

MEMBER LAVER: Well, do people have to report pregnancies in the same database so you can have an idea of how many were on tamoxifen and got pregnant or you can capture this in data if you target it?

MR. STEWART: These data are reported to us as secondary diagnoses or conditions that exit at the time of the index disease diagnosis, which was prior to the breast cancer.

MEMBER LAVER: I see. But not somebody being two years on tamoxifen and then reported, right?

MR. STEWART: No.

MEMBER LAVER: So this would be tactical measure.

CHAIRMAN LUTZ: Karen?

MEMBER FIELDS: The measure is just that they started and were given tamoxifen or aromatase inhibitors. So

obviously our suggestions for improvement would be how do we measure that they got the prescribed course and they got the right duration of course, and they got the right kind of anti-estrogen therapy based on their menopausal status. So those are, I think, the shortcomings of the measure, but obviously there was a huge disparity already, we have a disparity issue, so we aren't there yet, but I guess at the end we should also make recommendations about improving the quality of the measure.

MR. STEWART: So the question of menopausal status was extensively discussed when the NQF originally reviewed this measure five years ago. The conclusion was that the feasibility of determining menopausal status was very low, and so there was a decision made to basically include all comers in this measure and not distinguish around that fact. It's just a shortcoming of not just our data set, but probably many others that could be

used to assess this.

The second question about care compliance, if you will, is not one that we measure directly. But even in associated work where we've had a chance to look at claims data sets and what not, you know we can tell the fact of prescriptions being written and filled. It's also clear that there's some elasticity, if you will, in patients continuing to fill those scrips over time. And those sorts of data enterprises to look at concordance or patient compliance over time were very difficult to think about from a feasibility perspective. You know, where we had simply chosen to focus on the fact of, you know, at least initiation or the prescription being written for the patient to fill. And using that as our indicator for compliance with the standard of care.

MEMBER FIELDS: And we will discuss this and make recommendations, but there's also another measure this afternoon

that's the same endpoint. So how do we deal with that? Because it actually has some different exclusion criteria.

MR. STEWART: If I can comment quickly.

So I've had brief conversations with the other developer of that complementary measure and we'll see if we can address your concerns this afternoon when the conversation comes up.

MS. BOSSLEY: Right. So this one is a facility, the other one that you'll look at is clinician. So those would be viewed as, I would think, related. They're not competing, because they do have different levels of analysis. The question will be: are they harmonized. And it sounds like there's discussions already.

So part of what I think the feedback you should provide is exactly where you think the harmonization should occur and we'll walk through that once -- we'll do a

table of the two, and I think that will be helpful. And then, again, go back to the developers and see what they can do. But it's a very good question.

MEMBER LAVER: Race can be extracted from the electronic medical record, right?

MR. STEWART: Yes.

CHAIRMAN LUTZ: Anybody else on the phones have anything to add? All right. Are we moving on to vote?

MS. KHAN: So voting on 1a, impact. Again, it's high, moderate and low or insufficient evidence.

MEMBER LAVER: Laver, I vote one.

MS. KHAN: So we have 14 high, three moderate, zero for low and zero for insufficient.

Moving on to 2b, performance gap. High, moderate, low or insufficient evidence.

MEMBER LAVER: So basically if you vote low, there is no performance gap?

MS. KHAN: Yes, that's correct.

MEMBER LAVER: Did I get it correct that the data show 3.5 outlier, 3.5 percent?

MS. KHAN: Andrew?

MR. STEWART: I'm sorry, I don't have that full set of documentation in front of me. So the 3.5 percent that you cite are hospitals -- are the proportions of hospitals that we have applied this measure to where they lie at a significantly low performance rate. You know, beyond a standard deviation or some such from the mean.

MS. KHAN: Did you want to put your vote in?

MEMBER LAVER: Yes. Three.

MS. KHAN: Okay. Thank you.

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

And looking at 1c the evidence, you're going to vote yes, no or insufficient evidence.

MEMBER LAVER: Laver, I vote yes.

MS. KHAN: So you have 16 yes and one no.

Moving on to reliability. High, moderate, low or insufficient evidence.

MEMBER LAVER: It's Laver. I vote high.

MS. KHAN: Can we have everyone press their clicker again? One more time. No. All right.

We have 11 high and six moderate.

Moving on to validity. High, moderate, low or insufficient evidence.

MEMBER LAVER: This is Laver. High.

I will have to step out for a few minutes.

MS. KHAN: All right. Thank you.

MEMBER LAVER: So I can tell you I vote high and yes on all of the coming ones.

MS. KHAN: Okay. Thank you very much.

Can we have everyone press theirs one more time, please? There we go.

So eight for high and nine for moderate.

Moving on to usability. So we have ten for high, six moderate and one low.

And looking at feasibility, again high, moderate, low or insufficient information.

We have seven high, ten moderate, zero low, zero insufficient information.

And overall suitability for the endorsement, does the measure meet NQF criteria for endorsement, yes or no.

We have 17 yes, zero no so the measure will pass.

CHAIRMAN LUTZ: I think that based upon the strong start that Member Marks gave us, we made it to the break a little bit early.

MEMBER MARKS: Thank you.

(Whereupon, at 11:03 a.m. off the

record until 11:25 a.m.)

CHAIRMAN LUTZ: And a request was made if we could find who is still on the line from the Committee that's going to be voting. I know, Larry, you said you're free in about five minutes.

Rocco, you still on?

MEMBER RICCIARDI: I'm still on.

CHAIRMAN LUTZ: Okay. And Heidi? All right.

So, I guess Rocco will be our lone holdout after Larry steps aside and unless Heidi comes back on. All right.

So the next one we have is 1857. I think it's the HER2/neu. I think ASCO is going to be the one that's giving us the framework and then Stephen Edge is going to give us the perspective from this Subcommittee that looked at it. So, I think ASCO is presenters first.

MS. McNIFF: Thank you.

So the first measure you'll be

reviewing 1857 is of course the three related: HER2 testing and appropriate use of these measures that ASCO has submitted for breast cancer.

We did submit a few updates in response to the work group calls. And Dr. Edge pointed out to me that one of his recommendations he did not give me, which is to change the title to make it clear that we're talking about the adjuvant setting. And so if that's all right with NQF staff, we can certainly make that change.

MS. FRANKLIN: Sure. We will open the measure for you.

MS. McNIFF: Thank you. Happy to do that to clarify.

I did want to make one general statement that is relevant to all of the three measures we'll be reviewing this morning and also right after lunch, and that is that we recognize and understand the comments that we heard about the high performance demonstrated

by QOPI data. Did look further at the data as requested by the folks on the work group, and you know confirmed that the QOPI data do show a little bit of variation, but both the mean and the median are high and practices are clustered to the extreme cortile. And this is similar to some of the other measures that have been reviewed this morning.

We want to reenforce that QOPI is a selected group and they're participating voluntarily in a quality improvement program. So this group is likely not reflective of care overall nationwide.

We would ask that you consider these measures in the same way you thought about some measures brought initially for consideration for accountability use in the past, and that is to see what happens when they are used outside of the QOPI system in an accountability way and we can see whether there is variation within the wider communities with wider use.

At the point of reconsideration of the measure, of the maintenance review, within a few years we show that there is not variation, then we would absolutely agree that the measure should be retired from accountability use. But at this point we're suggesting that the approach be taken, we see what happens when these are implemented nationwide.

CHAIRMAN LUTZ: Okay. So, Stephen is this your --

MEMBER EDGE: So as people recall, this is to measure the appropriate nonuse of trastuzumab in receiving adjuvant therapy for stage I T1c and zero or stage II or III breast cancer. And the concerns that we raised were related to what Ms. McNiff just discussed with the very high performance on QOPI and really the absence of data from other practice settings besides those volunteers who choose to participate in QOPI. And whether those are or are not high performers, I don't think ASCO

has demonstrated, although I could be wrong on that.

And beyond that, I think the importance is certainly clear that women should not receive an extensive and toxic therapy when there really is no indication. There is a clinical trial now looking at the use in HER2 negative breast cancer, but the clinical trial's exclusion is included in the measure.

The measure properties are certainly acceptable.

The useability will require probably chart abstraction at the hospitals, cancer registries, collect immunotherapy. And Mr. Stewart and I'd have to comment as to whether trastuzumab is considered chemotherapy or immunotherapy in the cancer registry system.

Is trastuzumab considered chemotherapy or immunotherapy in the cancer register system? I don't recall the answer.

But the registry still does not code the exact name of the drug, so it will require significant chart abstraction unless there's ability to get electronic health record data abstracted automatically, which is probably quite some years away on a national basis, or if there's an ability to obtain administrative claims from payers. Certainly Medicare could do that if this was implied in the medical population.

It's certainly a useable measure. It's feasible, though it would require some of the things we just talked about, and I don't think there's any measures here. So I think they're largely addressed the concerns that we had regarding the claim.

CHAIRMAN LUTZ: Okay. Thank you.

Bryan?

MEMBER LOY: I just wanted to go back to the comment you made about demonstrating performance gap that was back on this page just a moment ago. I'd like to

understand a little bit better what your finding in your data. Is that largely the variance and the opportunity for improvement largely a lack of documentation or is it in fact those folks receiving trastuzumab that re HER2 negative?

MS. McNIFF: So while we don't know necessarily. But the position is that there needs to be documentation in the record about HER2 status. I mean, actually it would be better if the three measures were flipped in order. And if the documentation is not there, then the treatment decision should not be made.

So, there is the possibility in Dr. Hammond's office during the work group call as well, certainly that the HER2 testing was done. But ASCO's position is it needs to actually be documented in the medical oncologist's record before the decision to give or not give trastuzumab is made.

MEMBER LOY: To the earlier

comment made if you've got folks receiving therapy in the face of HER2 negative result, that's a different problem versus documentation problem. So, I don't know if that needs to get resolved or not, but it just feels like two different levels of severity.

I appreciate your reaction to that.

CHAIRMAN LUTZ: Robert?

MEMBER MILLER: So without being repetitive here, I just want to speak to the values of parsimony with these measures. And, you know, I guess this one just strikes me as one that doesn't make sense. That as a practitioner I just can't see this happening very often. And it's not a very data-driven answer, I understand, but you know if I as a breast oncologist ever did this knowingly, I mean I can't imagine anything more egregious-- a few things, I suppose. But second, I can't imagine that I'm going to slide this by too many payers. And second, and again I know

that's not what the measure is all about, but again just speaking to parsimony, I can think of a lot of interventions in oncology that should be never events, but I would just submit my judgment. I'm not sure this makes it if we have to pick and choose, we can't put every measure. It's more opinion than data.

CHAIRMAN LUTZ: Karen?

MEMBER FIELDS: I have two questions, one for the sponsor or the steward, which is again I know you started out and explained this to us. But of that range, 80 to 100 percent with a 99 percent mean do we have numbers, we have ideas about what the 80 percent means? Is it five patients or is it thousands of patients? Because certainly we should not give a drug -- I mean, I thought that that gap was very wide and we shouldn't give a drug to patients that shouldn't be receiving it, so making it important to measure. But I agree with Dr. Miller, the payers are going to capture this so does it

need to be a quality measure? Because the payers, it's such an expensive drug, it's such a well known indication at the moment that the payers will do the quality monitoring for us in a different way. Because nobody's going to dispense that drug without evidence that you're HER2 positive.

MS. McNIFF: So the payers in the room may want to comment, but we certainly heard different things. That is not what we heard from payers. We have not heard that same story, you know that this would never happen and that the payers would prevent this from happening. So others may want to comment on it, because it's not certainly my area.

In terms of what do the bottom practices who are scoring look like? They do tend to have small numbers. So I don't have in front of me what the end for each one of those sites are, but yes we do start to get down to the size records or a small number in some of those cases.

CHAIRMAN LUTZ: Okay. Jennifer and then Stephen.

MEMBER MALIN: I think from the payer perspective and whether or not, you know all payers review this post-hoc or not, which I don't think is routinely done because it's very expensive to do that kind of review, I still think it's a different issue. I don't think we should mix what's sufficient for reimbursement with what we consider good quality care. And in this case there may be a lot of overlap, but it's not always going to be the case. So I think if we think it's good quality care, we should focus on that.

I have to say, you know I'm of two minds with this measure. I share the thoughts that Dr. Miller expressed about -- you know, I mean basically I mean at least in -- you know places I've practiced over the last ten years it's routinely obtained in every specimen I have. I don't think I've seen a case where it hasn't been there. So, you know

it would be hard to imagine it not being done.

On the other hand, you know I think a number of the measures that we've looked at so far today also have gaps that are negligible, if at all and we endorse them. And I think that this is at least moving with the science and focusing on more targeted therapy. And so we'll hopefully encourage, you know thoughtful consideration of submitting new measures.

MS. McNIFF: And can I just make a comment in response to Dr. Malin's comment?

From the measure developer perspective another related comment to Dr. Malin's statement is that we often -- you know, if you look at the three of these measures together as a measure of testing and then appropriate use, the measure developers are often criticized for only looking at under use and not providing the complete picture of whether the overuse of the drug is also -- not representing the fact that overuse of the drug

is also a quality problem.

So, you know again if looking at from the quality perspective you're able to identify whether the testing was done, and in this case more importantly documented in the medical record and then look at the treatment decision whether under use or overuse is an issue.

CHAIRMAN LUTZ: Stephen, do you have anything?

MEMBER EDGE: Just to Karen's comment. I think that most payers at the current time do not collect information on the result of HER2 testing and therefore would not be able to actually apply this measure directly or audit this. They would have to do a special audit.

I know that very early on in the use of trastuzumab one large payer did audit 200 cases and found that trastuzumab was administered to something on the order of 12 percent of people to whom it was administered

had not had a HER2 test done. Now this was 2005, '06, '07.

I actually was working with a medical director at one of Jennifer's companies, Wellpoint Ohio Blue Cross/Blue Shield and relayed that information to him. Again, this was very early in this time frame.

I understand that they implemented sort of you had to provide certification that they had a test that was positive or they wouldn't cover trastuzumab, and that had to be submitted within a few weeks of starting the trastuzumab. But I believe they stopped doing that because they found that comportance was so very high and it was not worthwhile. But that result is hearsay.

I don't know if you have any comments about that?

MEMBER FIELDS: I've heard of that.

MEMBER EDGE: Yes. But that would have been four years ago that it was stopped.

Because the comportance was so high.

CHAIRMAN LUTZ: Can I ask two questions, both of which might be nitpicky and you can tell me to move on. But related to that one, from a pathology perspective I mean there are sometimes when the initial biopsy will come back with the pathology, you know 2 plus, recommend FISH. Then you send it for FISH. And I notice on here I mean when you say, Robert, you can't imagine anyone would do this, I actually know about an oncologist who if you had a 2 plus there was equivocal literature go ahead and give that medication unless someone said you'd better send it off to get the FISH. And so I'm asking, I mean that's not on here. So where does positive in the circumstances --

MEMBER HAMMOND: Well, there are clear guidelines that have been published between ASCO and the College of American Pathologists saying that exactly under what circumstances the test is positive, equivocal

and so on and what extra tests have to be done, when FISH has to be done, what the thresholds are and so on. But there is considerable confusion about that still in the literature.

I mean the HER2 Panel is readdressing that issue right now, in fact.

CHAIRMAN LUTZ: And the reason I'm asking unless I'm reading this wrong, this is just negative --

MEMBER HAMMOND: Negative.

CHAIRMAN LUTZ: -- negative.

MEMBER HAMMOND: Right. Right. Bob can comment but I think the default if people think the patient really is the remaining equivocal, clinicians will use their clinical judgment to define whether or not the patient should get trastuzumab or not. And they are not an absolute exclusion from treatment at all.

MEMBER EDGE: That's correct.

CHAIRMAN LUTZ: Karen?

Go ahead.

MS. McNIFF: I'm sorry, can I comment?

CHAIRMAN LUTZ: Yes, please.

MS. McNIFF: If you look at the definitions in the measures, and this follows all the measures, we have the exact definitions from the ASCO/CAP Guideline to provide the users of the measurers to identify what is positive, what is negative and what's equivocal. So the instructions here specifically lay out positive and negativity and equivocal. We know that that is an issue interpreting correctly, so that's provided as part of the measure sets.

MEMBER EDGE: That helps. Thanks you.

MEMBER FIELDS: Well, I was going to say that's the measure that we're going to do this afternoon, too.

MEMBER EDGE: Okay.

MEMBER FIELDS: So we should have

done it in the right order: Did you measure it, did you measure it correctly, did you give it when you were supposed to, did you not?

But I would say that I would think that the payers and the way they scrutinize this varies in different parts of the country. Because out West where there's a much more heavily managed care market, you need to send in data in order to prescribe to the patients in the managed care setting a lot more than out here. So there's probably much more regional variation than we understand about this the way that payers are approaching the meds. And I think, yes, it's going to change. So that's why having just been out in a place where it was very scrutinized and if we change more in the country over the next couple of years, it's going to become a non-important measure.

So, it may be we'll always measure it and have some data, but it just seems that -- I just wanted to comment.

I also think that no woman that's HER2 positive or negative should get Herceptin outside the clinical trials. So the 80, that range is very disturbing that there's some places that are giving it to inappropriate patients.

CHAIRMAN LUTZ: Bryan?

MEMBER LOY: Just to round out the payer comments. So from a payer perspective I would just say I agree with Karen there's a lot of variation, but I'd also be quick to add there's a lot of change on the horizon for us as well. So if I'm thinking about the broad spectrum of payers, whether they be regional plans, small plans, larger commercial plans or some of the government payers, many of those folks really don't look at preauth at all and others when they're looking at claims data, they have no idea what the result is. And when you start to look at some of the preauth processes that are out there today, it's more of an attestation rather than a, you know,

show me what your FISH result was.

So I think we're in a changing environment. I think folks are now looking for mechanisms in a nonintrusive way to get lab results as part of a record to be able to have a longitudinal view of the patient. Because the other thing that we haven't really talked about is there's a gap, and I think someone alluded to it earlier, but you know sometimes these tests are ordered routinely and then other times when they can't find the result, they're asking them at a point in time, retesting perhaps in some instances, and you may not have had that member on the plan during that time. So, I think there's a lot of noise in the system that we need to at least be thinking about when we contemplate reliability.

MEMBER FIELDS: Right. Right.

CHAIRMAN LUTZ: Can I ask a second potentially nitpicky question and it'll come up in a couple of the other submissions? If

we picture someone who is not in the streaming or doesn't read this whole thing, just reading the measure title, I've been taught when I go to examinations you don't really get as much play from something not administered because you're already sunk, you're going to punish someone for not -- in my head I keep thinking of the word "appropriately." That medicine is appropriately not administered because you have not within -- I don't know, just for someone who is not sophisticated and doesn't know exactly what's right or wrong if they're coding something and not administered, oh they didn't administrated, it should be appropriately not administered or the patient -- or something where it's more of a positive statement. Because the measure should be positive and then you can fall under it versus something where you are correctly not doing something. I don't know. I'm sorry, it may be nitpicky, but it reads confusing to me.

MS. McNIFF: But we are happy to

change the title to be more clear. And the not is -- because we report it both ways --

CHAIRMAN LUTZ: Right.

MS. McNIFF: -- with the different directionality, so one says one thing, another one says not. But to stand alone absolutely--

CHAIRMAN LUTZ: I don't know if anyone else agrees, but I read it back and also see where somebody will look at it and go "Well, they didn't do it. They didn't know. It's inappropriate or something." I don't know what the word is, but--

MEMBER PFISTER: Steve mentioned that they're actually doing clinical trials now where they're giving Herceptin to this population. And I thought you said that there's a clinical trial exclusion, but I didn't clearly see it in this document. So is there a clinical trial exclusion?

MS. McNIFF: It's in the numerator. So if you look at the numerator details if trastuzumab is administrated

according to a clinical trial--

MEMBER PFISTER: Okay. All right. Understand. Understood.

And then the other thing is that with regard to the performance gap, and I hear what you're saying about QOPI being sort of self-selecting and so forth, but you know what's the actual -- if you have a mean of 99 percent, the range is 80, you basically have one practice that was 80 percent. And so I would suspect that probably the distribution of practices is, I would guess, virtually a 100 percent all of them. You have one practice, too, that was an outlier. So I guess if you could give us some granularity on that in terms of like how many practices weren't already totally compliant with this measure?

MS. McNIFF: I mean, I'm not able to give that to you right now. Again, when we went back and looked at the numbers, again if you look at the scatter plot they're mostly

toward the top, there are a few practices that are down more towards the 88 percent. You know, we acknowledge that the concordance within the QOPI practices is high.

MEMBER PFISTER: If you compare and contrast like the appropriate nonuse versus the appropriate use? Like there's a bit more described for the appropriate use measure than there was for the appropriate nonuse measure. Is that accurate?

MS. McNIFF: I had to catch up with you, but yes, that's right.

MEMBER PFISTER: See, I think that the -- the comment made earlier about in a lot of ways the bundling of this in terms of like do we measure it the right way would have been a more systematic way to do this. And if I were to look at these three measures, I would say that that, firstly, that we measure it the right way. And actually the developer, while it's laid out very nicely in both 1857 and 1858, that if we measure it the right way --

you know if we're doing that well, I would think that the nonuse would follow --

CHAIRMAN LUTZ: I guess the question is -- I mean are we breaking protocol too much if we go in the order we keep suggesting and do -- we're only at 11:50 in the morning. And should we do 1855 and then go back to 1858 and 1857? Is the developer okay with that? Because it sounds like --

MS. McNIFF: Absolutely. It makes good sense.

CHAIRMAN LUTZ: It might make more sense. And we can go on with this one and have the discussion, it should be easier. If we do that, if we could, maybe we should just do -- 1855 is the one we're talking about, right? Yes. Can we just do 1855 and go from there? Maybe we can just do it now, because I think we're saying this would come third in order. Is that going to mess you up? Is that all right? Let's do that. Because then we'll go in the order. We're do 1855 and then 1858

and then 1857 and then we'll have everyone -- I'm sorry, 1878. So shall we start from 55 or 78?

MS. McNIFF: I'm not sure whether the CAP -- I mean the CAP -- we're not the stewards for or the owners of 1855, so that's a little bit more of a --

CHAIRMAN LUTZ: So should we do then 1878 then is the one? We can do that? Can you do that one?

MEMBER CHOTTINER: 1878 is the percentage of patients with invasive breast cancer who receive HER2 testing. The numerator is HER2 testing performed and the denominator is all adult women with invasive breast cancer. The only exclusions are history of metastatic disease or multiple primaries.

This is a process measure.

The level of analysis is clinician, group practice, clinician team.

The importance to measure, I know

this is a large group of women and the testing is both prognostic and predictive in that it helps to determine the prognoses and predicts the response to trastuzumab.

The evidence level is high. The scientific acceptability I think we thought was high during the work group. It's a new measure, so the only performance gap we have demonstrated again is from the QOPI data with the same caveats that these were high performing groups and that although the performance measures were high, there's a concern about generally.

The useability and feasibility we thought were moderate to high.

The questions that came up during the work group in addition to the performance gap had to do with the statement that we do this testing for all women with invasive breast cancers and the only exclusion pathologically is too little tissue to test. And I think the issue is that the clinical

trials that have looked at trastuzumab have been the adjuvant setting and have, for the most part, been for women who nod negative or women who have two nods that are more than one centimeter. But you can correct me if I'm wrong, but I think that MD Anderson did some retrospective studies and we do that for tumors between .6 and 1 centimeters there is some prognostic value to the testing and that these patients that can be considered for trastuzumab. But there are really no data for smaller tumors.

And I think that's the biggest issue we had with that: Should we really be doing HER2 testing in women with DCIS with microinvasion or local DCIS or very small tumors? And my personal experience in the community hospital where I worked before I went to U of M is that this was something that we took up our pathologist because it does add to the expense of reading these and it really doesn't make much sense to be doing the HER2

testing on these very small tumors if it's not going to impact treatment outside of a clinical trial.

I do know that there are trials now looking at HER2 testing in DCIS and we participated in those and just called our pathologist and had done on a reflex basis.

CHAIRMAN LUTZ: Were there any comments from the work group that discussed -- Robert?

MEMBER MILLER: So I generally agree with what Elaine said. But I think -- I'm not sure with how this relates to what we're voting on, but I'll just say that there are the same MD Anderson series and others I believe looking at even smaller tumors. The T1a subgroup did seem to show that there was important prognostic value to HER2, so particularly in the ER positive group. So the HER2 positive T1a tumors or less than 5 millimeters clearly did much less well. We don't have the predictive information in that

group because the randomized trials almost all used patients who were centimeter or larger or ne positive. But I think that, again, I'm not sure how this goes back to a measure and whether we should require this or not.

On the call, I was the one that brought it up saying that I was just questioning whether we wanted to be sure if we're holding people's feet to the fire to do this test, is it relevant? And maybe that was more rhetorical or not, so I can give both sides of the street. But I would say that I'm not even sure the that T1a tumors are necessarily excluded from the discussion.

MEMBER HAMMOND: Based on the information that's coming out in the guideline panel that's now redoing the HER2 guideline again, it appears that there's a lot of heterogeneity in breast cancer. That metastatic disease has to be retested.

So from a perspective and also the data that Robert just brought up, I think from

a perspective of looking at it for the benefit of patients in the long run it's better for patients to have this data available to them and for their physicians to have that data available to them when they recur, if they do, for the purposes of prognosis and so on.

And to make exclusions into this measure will make it more difficult to -- or it will encourage people not to do it maybe in situations where they should. So, I would argue against having that exclusion in the measure.

CHAIRMAN LUTZ: Elaine?

MEMBER CHOTTINER: I think the issue I have coming from 20 years in a community hospital originally was that for one thing, you have to take costs into consideration. And I think that this particular Committee can't really be proactive. I mean, I think that we need to look at the data. And if you look at the NCCN guidelines, they're very specific about the

indications for treatment. And although I agree that I have treated patients with two millimeter tumors with Herceptin, but on a case-by-case basis. And I think to incorporate it into a generalized priority measure at this point in time is premature.

MS. FRANKLIN: I just wanted to say that if the evidence changes for this measure after we've endorsed, we can also do a review of the measure at that time.

CHAIRMAN LUTZ: Does anyone else have a statement on that topic or other? Bryan?

MEMBER LOY: A couple of things. I guess I'm a little bit perplexed about the lack of having sort of a time element to this measure. I think I heard you say heterogeneity issue and the proximity --

MEMBER HAMMOND: Well, yes.

MEMBER LOY: -- to treatment issue. I mean if you've got three year old data, you meet the measure you know because

you got it routinely in an early stage and then it recurred. And that's a little more troublesome.

And then I'm also just wondering if you all spent any time talking about the work we've talked about, those folks that perhaps wouldn't be candidates for trastuzumab because of cardiac function, for example?

MS. KHAN: We do talk about that in the actual --

MEMBER LOY: I'm sorry?

MS. KHAN: -- treatment.

MEMBER LOY: Okay. But I'm just saying to myself, you know if it's -- again from the payer perspective prognostic, okay so I'm getting news but if it's actionable news, what's the clinical utility would be the next set of questions. And if there's an answer, would love to hear it. But if it's predicted but predicted only for one regime that would excluded, that would be important.

MEMBER HAMMOND: I don't know what

the data's going to show in the long run, but I think it has differential significance in ER positive versus ER negative patients.

We also do this test retrospectively on patients. So putting a time exclusion on it would not be a good idea because sometimes you go back and measure their tumor from a long time ago so we don't want to put a time exclusion on it.

MEMBER LOY: Then I would ask how reliable is that information --

MEMBER HAMMOND: Very reliable.

MEMBER LOY: -- from a tumor that was three years old that has gone through chemotherapy, do we have good data that says that a recurrent disease that was even HER2 negative three years ago is now the same and vice versa in the face of chemotherapy. But we already got a heterogeneity issue, and now we're going to introduce a chemotherapy issue. That's --

MEMBER HAMMOND: Well people are

using that information. I don't if that -- that doesn't help you, I know. But in fact the testing does get done, mostly in people who never had it done in the first place is the problem.

So doing it, say if we're recommending in the new guidelines that metastatic disease be tested specifically, and that would argue that it should be proximate to the treatment. So I guess I agree with you.

MEMBER MALIN: I would just say, I think that is an evolving area. I mean, that actually may be pushing the envelop. I mean, and there's been some recent studies, you know smaller studies that have suggested that there's maybe more tumor heterogeneity than we thought previously. But until now the standard of care has been that when someone recurs, you use their original pathologic information and you don't go in and rebiopsy. And that's what would be required in a

situation is to rebiopsy someone to get newer tissue information.

MEMBER HAMMOND: Well, I think that's under active consideration in the redo the HER2 guideline, but we don't have the data yet and it could be changed when the measure changed.

CHAIRMAN LUTZ: David, did you have anything?

MEMBER MALIN: And I guess the other thing is I don't know if there was some concern that we would be over testing HER2, but I mean one would have to think about it in terms of a cost standpoint given that probably most people need the test, it's probably less expensive that it's just a routine then to have to request it on a case-by-case basis. And so I think the system has moved to it being routine like ER and PR positive.

MEMBER PFISTER: Okay. I would be cautious regarding the -- unless it's very clear that it should be tested, and that's

going to guide therapy. But clearly there's a harm to boxes that are done and bad things happen. So if it's something that's clearly part of the state-of-the-art, that's one thing. But some of it is going to leverage behaviors to do biopsies that aren't necessarily that established, I think would be I think a potential downside of leveraging behavior that way.

CHAIRMAN LUTZ: Well, this may be a little bit down the rabbit hole, but actually I think is evolved as the new standard of care in breast cancer that biopsies should be done for metastatic sites. I mean, I know it's not published in the CAP guidelines yet, but I think practically speaking that's what everyone is saying ought to be done now. And, you know there are certain sites that I've found don't lend themselves well to biopsies. But I think we've just seen practically I think the discordance rate is something like 10 or 12

percent with HER2, and I forgot what it is for ER. So increasingly at all of our tumor boards at my institution that's what it is.

So, again, maybe not relevant to this, but just clarify.

CHAIRMAN LUTZ: Bryan?

MEMBER LOY: I just want to go back and you're on the end of the spectrum that I think I appreciate what you're saying, but I also want to go back to the comment to the comment that was made earlier about the smaller lesions. I'm wondering if perhaps we might be promoting overtesting and still in that arena that we talked about. But I'm hearing the argument of don't exclude that because you might need it later and I'm thinking feels like we're asking for it to be both ways.

So, get the information now or skip later in a world where we don't quite yet know.

MEMBER MALIN: So I think that, I

mean at least in my experience this is usually obtained even on the core biopsy. It just done routinely up front before you know what the size of the tumor is. So, I mean I think the cost savings for not doing it for those few people where maybe you don't need it would be more than offset by the administrative burden of having to say "Oh, well what size is this tumor? Do we need to get it or not?"

And then secondarily, I think you know, I mean obviously this isn't the forum but I don't think the decision about whether or not to rebiopsy someone should be based on whether or not just their markers have changed, right, a ten percent change in marker? Because in metastatic setting you basically assess response within two months of treatment. So, you know you'd have to show that having to wait to assess that response results in a worse outcome than treating -- you know treating empirically and assessing outcome with potentially inaccurate marker

data on ten percent of the population sometimes results in a different outcome than re-biopsying and narrowing your chance of having a response a little bit better.

CHAIRMAN LUTZ: Karen?

MEMBER FIELDS: Just to comment on treatment of metastatic disease, though. If you're usually giving combination chemotherapy and not Herceptin alone so if you didn't understand your HER2 status, you might be giving a drug that didn't need to be given and not being able to understand which that drug the patient was responding to.

So, the tendency tends to be HER2 ne positive patients stay on Herceptin for life adding a variety of different synergistic drugs, and that may not be even most rational use of our health care dollars.

But I would just echo that I think trying to interpret what the next set of recommendations today is make it very difficult for us to proceed with any quality

guidelines.

CHAIRMAN LUTZ: Is there anybody on the line with anything to offer? Are you still there, Rocco? I didn't forget about you. Heidi? Larry?

DR. HASSETT: Can you hear me?

CHAIRMAN LUTZ: Yes.

DR. HASSETT: My name is Michael Hassett I'm with ASCO and I'm a medical oncologist, and I just make a couple of comments about this measure.

I think it's an important discussion that's been going. And I would say that regard to the DCIS and the metastatic occurrence setting at least the way I read the measure I don't view this as part of this particular measure because it was what was done in invasive breast cancer. And I would agree there's debate about whether to test DCIS cases or microinvasion cases for HER2 positivity, but this measure is really focusing on the invasive breast cancer cases

and the denominator describes that.

So the small T1a cancers, the invasive cancers, I feel that the information is potentially in type of forms of treatment while I'm not commonly giving trastuzumab-based adjuvant therapy to patients with 2 or 3 millimeter cancers, it does have some prognostic import for those patients. And I do consider that information when I figured out their risk of occurrence and a potential magnitude of benefit from anti-estrogen therapy as well.

I also think just from a generalizability perspective, interpretability perspective I think it might be more confusing to have a measure that is excluding a small focus of cancer cells and there are a number of nonrandomized trials that are suggesting the potential for benefit for HER2 directed therapy in the T1a/T1b subset of patients.

So, I would argue strongly in favor of having the measure apply to all

invasive cancers and not excluding the small cancers.

CHAIRMAN LUTZ: Elizabeth?

MEMBER HAMMOND: The current guideline doesn't exclude anybody from treatment. It says it should be a routine test just like just ER/NPR. And that's the current guidelines. That's not future. That's not going to change in the next iteration either.

MEMBER ALVARNAS: This is Joe Alvarnas. I would like to add to that sentiment as well. I think we have to be careful about exclusions and we can always base upon data and we re-evaluate this at the time of its renewal later.

CHAIRMAN LUTZ: Okay. Thank you.

So, is there anything that --

MEMBER DONOVAN: That's my agreement as well.

This is Heidi.

CHAIRMAN LUTZ: Oh, thanks, Heidi.

Does anybody have anything else

they want to discuss or go on to further discussion before we vote, or are we good to vote on this one? All right. We'll vote.

So just to be clear as we're making sure of the voting for the phones. Heidi, you're there.

I didn't not hear Rocco, did you answer?

MEMBER RICCIARDI: I am still here.

CHAIRMAN LUTZ: Okay. So we got Rocco and Heidi are left for voting.

Larry Marks I think is not on anymore. And Dr. Laver is gone. Dr. Alvarnas has joined us. Good.

MEMBER ALVARNAS: Are we sending in votes via the gmail thing to Lindsey?

MS. KHAN: Okay. All right. So we're going to --

MEMBER ALVARNAS: I'm sorry, I apologize.

CHAIRMAN LUTZ: She said yes. She

said you can channel your votes straight through her.

MEMBER ALVARNAS: Okay. Thank you.

MS. KHAN: So voting on 1a impact. High, moderate and low or insufficient evidence.

So you have 13 for high, three moderate and zero for low and zero for insufficient.

Voting on 1b performance gap. High, moderate, low or insufficient evidence.

You have four high, seven moderate, four low and one insufficient evidence.

Looking at the evidence, yes, no or insufficient.

So you have 15 yes and one no.

And going on to reliability 2a. High, moderate, low or insufficient evidence.

I think we're missing one person.

MEMBER DONOVAN: I'm going to put

my phone on mute when we're not talking.

MS. KHAN: So that's 10 high and six moderate, zero low, zero insufficient.

Looking at 2b validity. High, moderate, low or insufficient evidence.

So nine high, six moderate, one low and zero insufficient.

Looking at usability, high, moderate, low, insufficient.

Seven high, eight moderate and one low, zero insufficient.

Feasibility, high, moderate, low or insufficient.

Can we do it one more time?

Ten high, five moderate, one low and zero insufficient.0

And overall suitability for the endorsement, does the measure meet NQF criteria for endorsement, yes or no.

So 15 yes and one no, the measure will pass.

CHAIRMAN LUTZ: All right. So

just to be clear, since we're going out of order and some of the folks on the phone might not have heard all that, so we started with 1878, which was measure HER2/ne. We're going next to 1858 which is appropriately treat positive, and then we'll go to 1857 which is appropriately not treat negative.

So next will be 1858 and we'll let the developer tell us what we need to know and then I think David is going to be the one to describe the Subcommittee's thoughts.

MS. McNIFF: Yes, I would be happy to.

All of what I said before applies to this one, too. There is a change that was made that was an error that was identified in the work group call. And that is in the finding of the trastuzumab administration within one year. That change has been reflected. It's within one year, 12 months of diagnosis.

CHAIRMAN LUTZ: David?

MEMBER PFISTER: I was not actually on the subgroup call, so those that were certainly feel free to chime in.

I think that, again, the discussion of this overall as a measure I think is probably so as to not to sort of repeat a lot of what has already been said, I think is perhaps best done in the context of its relationship to the prior measure. So I think that one of the issues that came up on the importance, the available data is dissimilar, the performance gap issue is similarly -- at least basic data provided it's smallish, but not as small as it is for 1857

Kristen clarified the issue that came up about the timing of the Herceptin.

It also did come up in the call that, you know given the potential cardiac morbidity of the Herceptin that the exclusions are not super explicit about that. You know, my sense is it's probably purposely made that way because to overly explicit is probably

going to be ultimately overly explicit. And, you know it gets into the realm of judgment.

I think, again comparing the votes for the suitability of the measure for 1857 versus 1858, at least on the all there seemed to be that the preliminary assessment for the suitability for the most part seemed to be uniformly yes as opposed to the prior it was uniformly or seemed to be weighted the exact opposite direction.

CHAIRMAN LUTZ: Thank you.

Is there anyone on this call that wants to assure the facts of the Subcommittee?

Steve?

MEMBER EDGE: I note that the exclusions include the contradiction or other clinical exclusions. A consideration the NQF might want to have a consideration of making these analogous to the American College of Surgeons measures where those patients were not excluded from the denominator, but rather were considered concordant with the measure if

there was appropriate documentation that they should not receive appropriate treatment.

I think it would be confusing to users to be having to figure out who to exclude from their denominator rather than taking all people who have HER2 positive cancer who meet these criteria and then providing a reasonable either they got treatment or didn't get treatment rather than allowing the provider to choose who to report as a member of the measured group of patients. I think it'll be easier for the user. I think it will be more open and transparent. And I think it will allow granularity of the collection of data as to why that person was excluded. And it will allow them to have a uniform set of way of applying these measures.

CHAIRMAN LUTZ: Yes?

MS. McNIFF: Can I respond to that?

CHAIRMAN LUTZ: Yes.

MS. McNIFF: So that is actually

the way; it's analytical exclusion. The data are collected on every patient so that exclusion happens in the analytic of the measure. You know, we will collect this on every patient and the provider has to actually submit to us if there's a contraindication and it's pulled out analytically. And you can actually look in the -- you know, by that methodology you're actually able to look and see how often you're reporting the exclusion and have that date as well. But we absolutely do not -- I mean, I agree with you, Dr. Edge, that is not the approach that we take.

MEMBER EDGE: I think the NQF ought to look at this carefully and make this a homogeneous way of doing this rather than having us to go back and forth between those two different mechanisms for reporting. And I would argue for the American College of Surgeons' mechanism rather than the other, but I would recommend the NQF look carefully at that question when these are actually

operational.

MS. McNIFF: Just in response, I think that's actually a pretty significant change. And a lot of the changes I think we can bring back fairly confidently saying that the ASCO Committee would be happy make reporting a contra -- rereport clinical trial as a yes and for the numerator if the treatment was not done, by reporting a contraindication as a yes that the treatment was given is a conceptually major change. And so that one we would definitely need to do some real thought and work. ASCO does not specify that way.

CHAIRMAN LUTZ: Jennifer?

MEMBER MALIN: I mean, I think is value to harmonizing the approach so that exclusions are either handled in the numerator or the denominator. I think, you know personally as someone who has spent most of my career working on these kinds of things, I think it's much cleaner to do it through the

denominator because in the numerator it's open to a lot more interpretation. Essentially you end up having to count any notation that treatment was considered or recommended as passing the indicator, whereas excluding it from the denominator usually the criteria are much stricter.

MEMBER EDGE: If somebody is excluded because the doctor says they have a low ejection fraction and I'm not going to give them trastuzumab, how is that different whether they're excluded from the denominator or the numerator? Why is it more strict if they're excluded from the denominator? I'm sorry, I don't understand that one.

MEMBER MALIN: Because generally speaking, I mean it may not be operationalized this way in the American College of Surgeons data platform, but usually when the numerator statement says "Treatment was considered" or "Treatment was recommended", any notation in the charts that treatment was discussed,

recommended without any indication provided as to why it wasn't given is usually considered sufficient to pass the indicator.

MEMBER EDGE: But wouldn't that be just as equally sufficient to pass the exclusion from the denominator? I mean, the College of Surgeons could switch around and analyze it the other way as well. But if NQF thinks that that's a better way to do. But, I'm sorry, but I don't understand why the doctor is saying that it's excluded because the patient is too sick to get the therapy is any different whether the doctor excludes it and we choose to put it in the numerator or the doctor excludes it and we choose to do it from the denominator.

MEMBER MALIN: I guess it wouldn't be different -- well, the ratios can appear different.

MEMBER EDGE: That's true, reportedly different.

MEMBER MALIN: But the numerator

statement I think is different if you say receive treatment unless the following, or have documentation that there was a contraindication, any of the specific exclusions. But if the numerator statement says "Consider treatment" or "Recommended treatment," that's much broader than received unless, which is the way essentially this --

MEMBER EDGE: The only value with putting this is in the numerator is that it allows you to see for an individual provider, institution, however you tend to attribute this whether that organization has a problem in that a high fraction of their patients are refusing therapy or they're choosing not to give therapy. So if an institution has 30 or 40 percent of their patients -- and Mr. Stewart alluded to this in his presentation. If that institution has a very high proportion of patients who are choosing not to get therapy, then that institution has got a quality problem in how they're presenting --

or a potential quality problem in how they're presenting that information to patients.

And an exclusion from the denominator we lose the potential to identify that quality problem. And that's one of the reasons why I think this is -- I actually don't agree with you that there's any different where you exclude them in terms of the indications on how it's documented. And I think there's added granularity and added quality evaluation and added opportunity for quality improvement by including in the numerator and separately reporting those patients who are not treated and considered excluded based on medical indication or patient choice.

CHAIRMAN LUTZ: So you are saying that this one is, as per the ASC --

MEMBER EDGE: I would recommend that the NQF look at this carefully, and it probably goes beyond our ability to make the answer today. But I would suggest that when

operationalizing this through NQF and through CMS that this course should be more carefully reviewed. I think it's a really important question. I don't think we came prepared to address the question today. And I don't think we're fully prepared to answer the question today. I think you've got some concept from Dr. Malin and myself and others. But I think this is a really important one that the NQF may want to address.

MS. McNIFF: I was just going to say, so I just wanted to clarify that this particular piece of the conversation is regarding recommendations as to what you would like to see in the future. And we're looking at the measure in front of us. Is that a recommendation for changing --

MEMBER EDGE: I personally would recommend my recommendation --

MS. McNIFF: Right.

MEMBER EDGE: -- and I suspect it will be taken today for this approval. But my

recommendation would be that they be switched and I would recommend that the NQF and with this measure have those cases excluded from the numerator and not from the denominator.

I would recommend that we turn this back to the developer with that recommendation.

MS. McNIFF: Okay.

MEMBER EDGE: But after the fact, I think this is something the NQF should look at very carefully before these kind of measures are implemented.

MS. McNIFF: Karen, did you have anything?

MS. PACE: So, yes. Exclusions is a big topic of interest and it is something that our Consensus Standards Approval Committee is going to be looking at a little more closely.

Currently our guidance specifically about the issue of patient preference or patient declining is that the

measure if that's include in a measure, it should be transparent. So the ways that that could be transparent is exactly as you've talked about: Is a numerator category. The other way is that you have to report both rates -- both with and without those exclusions because of the very reason you're talking about. If one provider has a higher rate of patients declining in treatment, you know what's going on there?

So, it is certainly a broader issue than this project or these particular measures.

In terms of the harmonization, I think that's something that you'll be talking about later if individual measures on their own merits meets the criteria, then you know if these are big issues in terms of related measures, you know how they would define the denominator and exclusion populations. That's something that the Steering Committee can certainly weigh in terms of when they're

addressing related and competing measures.

CHAIRMAN LUTZ: Yes, Karen?

MEMBER FIELDS: So to the developer. At the beginning you summarized what changes you made in response to our previous discussion. And the only one that I heard was you changed it from a four month window to a year window. You still didn't go through and do our recommendations about more clarity in cardiac exclusions, correct?

MS. McNIFF: So I would ask Dr. Hassett to comment on that.

DR. HASSETT: I think one of the challenges with -- and you guys have been having this conversation, is how to rank corporate exclusions into the mix for these folks.

The vast majority of -- this measure is targeting folks who receive chemotherapy for breast cancer, and the vast majority of these folks will have already had preexisting cardiac evaluation. So, at least

from my perspective, the probability of that cardiac evaluation in addition to including the characteristic of chemotherapy receive cardiac evaluation would be very unlikely to leave somebody out of this because they wouldn't have gotten in the measure in the first place, because they probably would have gotten chemotherapy.

MS. McNIFF: And to add to that, there is a clinical exclusion option, right? So that goes --

DR. HASSETT: Oh, yes. Yes. And, of course, yes, if there is a clinical comorbid condition option.

MS. McNIFF: Right. It's already there.

DR. HASSETT: So we felt that with those elements that the concern about getting folks into this measure who shouldn't be there for cardiac issues were addressed.

MEMBER EDGE: One quick comment, Kristen, is that I would suggest you also

change the title on this measure just like we did for the other one to reflect that this is trastuzumab administered with adjuvant chemotherapy for a patient with AJCC staging and for clarity for the user. That this measure isn't intended to be addressing people with metastatic disease. The fact that they have AJCC stage I to III cancer, the stage doesn't change when they have metastatic disease, so that does not clarify that. I would add the same thing for consistency and clarity.

MS. McNIFF: Yes. And I meant my opening comments to reflect both of the measures. We will absolutely do that, make that change.

And the page, Dr. Fields, is 9 -- oh, but I'm looking at a different document. It's 2a1.8.

MEMBER FIELDS: For those of us in the room that have prescribed it, are the label indications do they say cardiac

exclusions? I don't remember. Just I mean, I think there's some very clear cut ones where we don't worry about necessarily remeasuring the ejection fraction. If somebody had congestive heart failure or some -- you know, a history of those things, those are contraindications that are pretty well standard. And so I still am disturbed that we don't enumerate that a little bit in the exclusion criteria rather than the general statement. But maybe just changing the title and making sure everybody understands that the quality measure isn't punitive, it's more meant to just be a quality measure will help that problem.

MEMBER PFISTER: How do the measures here handle when, you know sometimes I see these things come through where, you know have it tested at one place, it's registered HER2 negative, it's tested in another place it's another place it's HER2 positive. And how does one trump the other or

is basically that, you know any positive will count as a decision to justify giving it and any negative will be justification to the other measure?

CHAIRMAN LUTZ: I think, isn't that in the directions we printed out when I asked a similar question for the other one? That's in the directions for use the following definitions to determine status.

MS. McNIFF: Actually in the instructions, Bob, information from the most recent report.

MEMBER PFISTER: So it's going to be whatever the most recent report is?

CHAIRMAN LUTZ: Robert?

MEMBER MILLER: I don't know if it's relevant to the discussion, but the answer to Karen's question, the label does not list any contraindications but cardiac is a boxed warning, it's listed under warnings. So it's technically not contraindication.

MEMBER PFISTER: Is there data

that suggests the most recent report is the most accurate report, or is that just you did it for a feasibility measurement?

MS. McNIFF: This is a feasibility issue. I mean others in the room many want to comment data about which report. But it was done for feasibility instruction.

MEMBER HAMMOND: I don't think there's any data about that.

CHAIRMAN LUTZ: So just to be clear, so Stephen made a plea that we submit this back without a change in the exclusion criteria. Are we comfortable to go ahead? Do we discuss further whether to give that back to the medical? I guess that's the unanswered question in my mind. Are we moving it to a vote or are we agreeing and saying we should move back and have those definitions more clear?

MEMBER EDGE: I would say this is a feasibility issue and I wouldn't actually necessarily insist or ask that you take a vote

on delaying the other votes. I think this is a broader question when you look at these clinical contraindications that I think the NQF ought to very carefully make these the same. And I think there's arguments on both sides.

But I'm not sure, for the purpose of practicality, that I would suggest that you insist on turning this back to the developer while you have that discussion, because I don't think the developer is going to recommend that they change it at this point.

MEMBER MALIN: I mean, I would certainly recommend that we defer on the issue of addressing harmonization because I think it goes beyond just the numerator/denominator issue. It goes to the issue of the specific categories themselves.

And then also, you know, what we haven't explicitly here is are these measures for a defined data set or not? So, for example, the College of Surgeons measures have

been implemented using their data, but I don't know that there's anything about NQF endorsement of the measure that says that they only think it's valid with their data set.

And so the exclusion criteria are going to get operationalized potentially differently in different data sets. And so, I mean, I think it's a broader topic that probably should be gone into in more detail.

CHAIRMAN LUTZ: I have been told we are allowed to vote as to whether we're going to vote, so if you want to -- but I certainly agree. I mean, I don't know that I made attention to the exclusion criteria that closely in all the other ones we've done, so it's sort of stopping procedure for this one measure for this one developer, whereas I don't recall whether we've gone that far in depth in any of the others. So I'm not sure if it's fair to put them under the criteria.

But, yes, we can vote as to whether we'd like to vote.

MEMBER HAMMOND: I would like to make comment that based on what Bob said about the labeling requirements that we can't really. I would like to see more specificity about the cardiac exclusions, but since the labeling don't have it, I don't think it's fair to do that to providers.

MS. McNIFF: And would you feel more comfortable if there was a specific notation along with the clinical exclusion contraindications that, for instance, cardiac?

MEMBER HAMMOND: Yes, heart failure for example.

MS. McNIFF: Yes. Right.

MEMBER HAMMOND: I mean, you can measure that with an ICD-9 code, it's not difficult to get that data. I would feel more -- but I'm not sure that it's fair to require it because the labeling requirement doesn't say that. So --

MS. McNIFF: If it's more of an instructional -- but clearly there as an

instruction instead of a data element?

MEMBER HAMMOND: An instruction would be great.

MS. McNIFF: Okay. I mean that we can certainly do.

MEMBER ALVARNAS: This is Joe Alvarnas. I was away, so I wasn't sure if the developer is in the room.

And I know last time when we met we wanted the developers to walk away, come back an hour later and push back a respond. Are they available for us to put this on hold for a little while, let them rethink and either push back or suggest modifications?

MS. McNIFF: Hi. This is Kristen McNiff talking representing ASCO as the measure developer. And I think we're fine right now.

MEMBER MALIN: Are we just looking for a motion to vote on whether we should vote on this? I move to vote.

MEMBER EDGE: Second.

CHAIRMAN LUTZ: We're voting, folks.

MS. KHAN: So we are voting on 1a impact high, moderate, low or insufficient?

MEMBER EDGE: Is it true that the NQF can make these kind of adjustments if they felt they were to put those clinical indication exclusions into the numerator or the denominator, they could modify this after the fact to do so?

Oh, that's a different matter then, because then I would retract my second to this motion because if you can't then take these and harmonize them so that they can be operationalized to the public in a consistently uniform fashion, I think that's a serious matter, actually. I'm then in disagree with it.

MS. PACE: And I'm sorry. I didn't introduce myself. I'm Karen Pace on NQF staff and work with the measure evaluation criteria on different methodology issues.

So the measure stewards own these measures. And so you're reviewing the measures as they were submitted. And basically we can't change measures. The Steering Committee cannot change measures.

If there is something that you think is a fatal flaw in terms of measure meeting the NQF criteria, then your voting should reflect that. So if you feel that the exclusions make this really an invalid performance measure in terms of being able to identify differences in quality, then that should be reflected in your vote for validity or ultimately whether the measure is recommended.

Now, you know you can if a measure goes down, you know you can then talk about conditions for your recommendation for endorsement. And so the Steering Committee could say that, you know we think this measure should be recommended on the condition of X,Y,Z and then the measure steward needs to

respond to that. And, you know it may be that they agree and we'd change it. It may be that they disagree and they give their rationale for that. It may be that, you know it's such a major change that it would require additional testing to really implement that kind of change.

So, there's no kind of one black and white thing, but NQF does not change measures after they're endorsed. The Steering Committee has some ability to recommend measures on certain conditions that the measure stewards reply to you about, and then you make a decision on that.

You know, your suggestion about NQF and having some standardized approach to exclusions, you know that's a much broader issue and it goes to making changes in our criteria, and that's a much longer process in getting that implemented across all topics and all measure developers, it's going to be a much longer process.

So, you know what you have at hand is the measure that's before you and voting on whether the measure before you meets the criteria based on what they've submitted in terms of the reliability and validity testing and how it's specified, and you know whether there's evidence that backs how it's specified, et cetera. And if fails, then, you know, you could recommend it on a condition and see what the measure developer's response is tot hat.

CHAIRMAN LUTZ: Elizabeth and then Bryan. Just don't want to skip you. Bryan?

MEMBER LOY: Just a comment.

(1) It feels like some of the discussion that we're having now is largely around the harmonization. I think I heard the developer say that didn't own all of these. So, it would be kind of hard on a measure-by-measure basis to really execute upon what you just described.

(2) I'd just comment to the group

it feels like this isn't the first time that this has come up. I mean, we've kind of all throughout our deliberations here have asked ourselves the question: So how good is good enough in terms of adhering to these measures? And it feels like to me at some level we've kind of acknowledged all along the way that there's some imperfections and some exclusions and maybe some things that we haven't completely contemplated.

And I don't know what it is about this measure that kind of brings that -- escalates it to a higher level, but it seems that at some level we ought to be acknowledging as a group that a 100 percent compliance is maybe not the --

MS. PACE: So let just clarify other thing. As I mentioned, what you're to be doing now is reviewing each individual measure against the NQF criteria. If after you go through this and you have related measures with the same target population, then that

becomes a harmonization issue if you know the denominator is specified differently, if the exclusions are specified differently. And that can be brought back at that time to go back to the developers.

Your vote today is really not a final recommendation. It's preliminary pending addressing any harmonization and competing measures issues. So I don't know if this measure has related measures that are targeted to the same population or you're just talking in general about --

MEMBER HAMMOND: No. No, just about the broader issue.

MS. PACE: -- the method of doing exclusions? Okay.

So you're right, harmonization and competing measures need to be addressed later, but this not about specific measures that are related or competing, but just the broad concept of how to do exclusions, I believe.

CHAIRMAN LUTZ: Karen?

MEMBER FIELDS: So this is a new measure and explain to us how the new measures get adopted. Because before we were talking about new measures have a year of review or --

MS. FRANKLIN: No. This measure has been tested, so it would be fully endorsed--

MEMBER FIELDS: Okay. So some of the other ones where there's --

MS. FRANKLIN: -- if that's the Committee's decision.

MEMBER FIELDS: -- no testing date--

MS. FRANKLIN: Those are time limits.

MEMBER FIELDS: Okay.

CHAIRMAN LUTZ: Yes?

MS. FRANKLIN: And we have a comment from --

MS. McNIFF: A point of clarification. We do in fact, these three measures ASCO does own. I think maybe there's

one from CAP that's related.

And I just want to make sure, a point of clarification. NQF does not dictate how exclusions are handled and in fact has endorsed many measures that handled exclusions by pulling them from the denominator analytically, is that correct?

MS. PACE: Yes. We don't dictate measure specifications. We do have criteria about exclusions that say patient preference should be transparent.

So to what extent that has been a key issue for any one measure, it has varied. In some cases it has been. So there is a criterion about that that would apparently apply to your measure. But in general we have measures that -- I would say that most of them are, you know excluded from the denominator. But we have examples of measures where there are numerator categories and it really depends on, you know the particular measure and measure developer. But right now our criteria

do not require one way or the other.

The only criteria that the exclusion should be necessary, they should be identified in the evidence or they should be of sufficient frequency that it's really worth the data collection effort, or if patient preference is one of the reasons for an exclusion, that it should be transparent.

MEMBER EDGE: Well in my mind, first of all, I'm not sure that patient choice is an exclusion. It's a concordance with the measure. You appropriately consider that the patient should consider trastuzumab in this situation and it's been decided actively not to do so for a specific reason.

Based on what you just said, I'm feeling even stronger that the way that this is handled, this specific "exclusion" is handled in this measure reduces the value to the public, the value to the providers for quality assurance and reduces the transparency. So if the goal is transparency

to the public and transparency to users for the purpose of quality improvement, the way that these exclusions, the way that this is included as an exclusion reduces the transparency because you can't see how many people were considered, how many people were eligible for the treatment, how many people received it and now many did not receive it because of valid medical reasons.

CHAIRMAN LUTZ: We'll see how many folks you swayed in voting. Time to vote.

MS. McNIFF: I mean I don't want to draw out this conversation, I think it needs to go to vote. But that seems to me to be a reporting issue and that by reporting out either the numerator categories or the exclusions that go with the denominator, each way you're able to demonstrate the impact of patient preference and the impact of contraindications.

MEMBER EDGE: I would agree with you that that could be dealt with in a

reporting way as long as the data are collected. Are you currently reporting that to your providers in that fashion?

MS. McNIFF: Yes, we actually report this measure and we report recommended and received and you're able to drill down to look at the information exclusions. Now that's within QOPI. This, you know it's recommended to be --

MEMBER EDGE: Is that recommended in this document for how this should be reported? The developer be willing to put in a reporting recommendation that the number of patient are excluded because of those kind of clinical issues be reported?

MS. McNIFF: I don't think that's an option, is it?

MS. PACE: The question again?

MEMBER EDGE: Can the measure have rules for reporting that say that you report the people who are eligible based on including the exclusion that if the patient says no,

they won't be in the denominator? And that they will also be reported how many people are excluded from the denominator because of that, which is the way the developer specifically is reporting to that providers now in their data reporting system.

MEMBER MALIN: I need a clarification. I don't recall in any of the other measures that we've reviewed where the exclusions were in the numerator that the reporting was going to stratify how the people passed the measure. So it's not like that's providing people at a reporting -- you know, if you're talking about quality reporting that people are going to use, nobody's talking about stratified results. So it's not like, you know if 50 percent pass a recommended measure because the doctor discussed it with them and they refused, you would have no way of knowing that.

MEMBER EDGE: Well, we actually did discuss that, not quite so in detail when

we were discussing the American College of Surgeons measures. And specifically we discussed how those data were collected and whether the specific data element included, whether it was because the patient refused or the doctor said no, or there are other reasons. And that's why I was suggesting that after -- I think we're going too far down this road right here, but I think that this is something that would be valuable to harmonize across these measures so there's a consistent method of reporting so that the public get a consistent report. So the public when they see these data don't have to dive into the methodology about how one measure was defined and how another measure was defined. Our goal here is for transparency to the public.

MEMBER MALIN: Are you able to identify those who refuse?

MEMBER EDGE: Yes. We specifically discussed that with Mr. Stewart, and you might want to invite Mr. Stewart to

come back to the table to discuss how that is collected if you want. But, yes, the answer is yes.

And, again, our goal is transparency to the public, and I think we're losing that transparency with this measure.

MS. PACE: So let me just clarify a couple of things. First of all, NQF endorses the measures, how they're implemented which includes reporting currently is not part of the endorsement. So we don't attach guidelines on how the measure is reported.

If the measure was actually specified that there's numerator component, that would be pat of the measure and the expectation would be that's how it would be implemented.

So, you know we don't attach reporting guidance to say that information goes back to the provider or that that information could be available. It's not part of the measure.

So what you're voting on is the measure as specified. And if for some reason the measure does not receive a preliminary recommendation for endorsement and someone wants to bring up a condition on which you might want to push it forward, that could be done at that time. But again, you know I don't know if this has been an exclusion in other measures that you've taken a look at, but you need to think about some of that balance in terms of how you've been looking at measures.

And again, whether this one element is in your mind a fatal flaw or not, then you vote that accordingly.

I think your general recommendation about harmonization of methods, not just of the actual specifications, is something that you could discuss as a Steering Committee about whether you want to make that recommendation, and certainly we can have some discussions about the developers about that.

But that's really a separate issue and it's something that we ask all Steering Committees to come up with recommendations regarding performance measurement, whether it's identifying areas where we need additional performance measures or if it's specifically about methods that apply across measures. You're certainly encouraged to do that.

MEMBER ALVARNAS: I know that you had scheduled a discussion of streamlining the process for how measures are evaluated. Would it be worthwhile including this as part of that much broader discussion?

MS. PACE: Yes, we'll certainly be looking at this for sure. Thanks.

CHAIRMAN LUTZ: All right. Shall we vote, see if this sinks or swims? We're getting to it. Yes, we did measure and then appropriately treated and then appropriately not treated. Right, but it's a different developer.

MS. KHAN: So voting on 1a impact.

High, moderate, low, insufficient evidence.

So you have 14 high, two moderate, zero low and zero insufficient evidence.

And measuring performance gap. High, moderate, low, insufficient evidence.

Can everyone put their vote in one more time, please?

So you have three high, nine moderate, two low, two insufficient evidence.

Looking at the evidence, yes, no or insufficient.

So you have 15 yes and one insufficient evidence and zero for no.

Looking at reliability. High, moderate, low, insufficient.

You have six high, eight moderate and two low, zero insufficient evidence.

Looking at validity. High, moderate, low, insufficient evidence.

Five high, seven moderate, four low, zero insufficient.

We're moving on to usability,

high, moderate, low, insufficient evidence.

Four high, eight moderate, four low, zero insufficient information.

And Feasibility high, moderate, low, insufficient information.

Five high, nine moderate, one low, one insufficient.

And overall suitability for the endorsement, does the measure meet NQF criteria for endorsement, yes or no.

So 13 yes and three no, So the measure will pass.

CHAIRMAN LUTZ: Shall we do member and public comment and then hit lunch.

MEMBER HAMMOND: Yes.

MS. FRANKLIN: 1857.

CHAIRMAN LUTZ: Do you think we discussed 1857 enough to go ahead and vote? Okay.

MS. KHAN: So. Okay. So 1a impact. High, moderate, low, insufficient evidence.

MEMBER LOY: Just so I understand, are we voting on 1857 with the revised language or is it as is? There was use of the word appropriate. So we do we handle it like--

MEMBER DONOVAN: Yes, they already said what they'd do.

MEMBER LOY: Okay. So could you put it back up on the screen one more time?

MEMBER DONOVAN: Maybe we can hear it again out loud?

CHAIRMAN LUTZ: So if I'm understanding correctly, so it now says: "Trastuzumab appropriately not administered to breast cancer patients when human epidermal growth factor receptor is negative or undocumented." So the additional is medicine appropriately now administered versus simply saying not administered? What was that change? And then was there an adjuvant therapy in addition to that as well?

MS. McNIFF: Yes. The use of the word -- I hate to wordsmith at the moment.

The use of the word "appropriate" has it's own specific meaning, and we can put that in there but I think that'll probably be fine.

We can certainly change the title for clarity, absolutely.

MS. BOSSLEY: Right. What we can do is I think we can let ASCO go back and kind of wordsmith and recirculate it, but let's if everyone's comfortable have you vote, assuming that there will be some language in there that's appropriately or whatever terminology. If you have concerns with what they circulate again, we can redo the vote or send it back to them.

But I mean I think they've heard it, they're going to make the change. So if you're comfortable, we can just vote that way.

MEMBER LOY: Given the discussion, I feel I'd be remiss not to at least look at the exclusions on this measure. Can we take a look at those?

MS. BOSSLEY: So for those on the

phone the exclusion is just patient transfer to practice after initiation of chemotherapy.

MEMBER ALVARNAS: Thank you.

MS. KHAN: All right. 1a impact. High, moderate, low, insufficient evidence.

So you have nine high, three moderate, four low and zero insufficient evidence.

1b performance gap. High, moderate, low, insufficient evidence.

Can we have everyone press it one more time, please.

You have two high, six moderate, seven low and one insufficient evidence. That's eight and eight, so it doesn't pass.

MS. BOSSLEY: So it is actually a split.

MS. KHAN: Yes.

MS. BOSSLEY: So in the instance of this typically we have you go on and continue voting and let's see how the rest of this plays out. Because what staff will do is

make sure reflects at the moment that you all really didn't come to consensus on this one subcriteria at the moment.

Are we on reliability or -- 1b, I'm sorry. I lost track. Walking in after being on a webinar makes me lose track. Sorry.

So in this instance all three subcriteria must be met to pass importance. So the impact, the opportunity for improvement and then also the evidence.

Here we've actually got a split. I don't think we can say whether this subcriteria was or was not passed because it's 50/50. So we should move on to the evidence piece and see if it passes that component. And then I think we should have a discussion again to make sure that are all in agreement. And usually what we typically do is you have a split vote on one of the subcriteria, it in essence doesn't quite pass but it's one of those that it's hard to tell, you'll move on to scientific acceptability if it passes

evidence. I think that's the next thing that we need to do.

This is one where it's always fun when we have a split vote on a subcriteria, and it's really let's move it through the rest of the process and see how it plays out against the remaining subcriteria.

Does that make sense? All right.

MS. KHAN: So looking at evidence, yes, no or insufficient.

So you have 13 yes, two no and one insufficient evidence.

So we're going to go forward, right?

MS. BOSSLEY: So again because you did have a split vote there's no real way to know. I think we just need to follow the stream and let's do scientific acceptability and move it through the rest of the process.

MS. KHAN: So voting on reliability. Again, high, moderate, low or insufficient evidence.

It's six high, seven moderate, three low, zero insufficient.

Looking at validity. Again, high, moderate, low or insufficient.

So four high, eight moderate, four low and zero insufficient evidence.

Moving on to usability. High, moderate, low or insufficient.

So five high, eight moderate, three low and zero insufficient information.

Feasibility.

So you have six high, six moderate, four low and zero insufficient.

And overall suitability for the endorsement, does the measure meet NQF criteria for endorsement, yes or no.

So you have nine yes and seven no, so the measure will pass.

MS. BOSSLEY: So I think -- I wasn't here for most of the discussion, so I apologize. But I want to make sure that staff have enough of a kind of a rationale to

understand why people voted and we had a split vote on the opportunity for improvement. So if -- again, more it's more to Angela and Lindsey if they have enough information. Because we want to explain kind of where we landed on this.

Again, it was a close vote, but it did pass and we have the split vote in the opportunity for improvements.

Feel like you do? Okay. ASCO feel comfortable? Okay.

I just want to make sure because I wasn't in the room.

CHAIRMAN LUTZ: All right. So any public comments or any NQF comments from the group or on the phone?

MS. TIGHE: Of you could open all lines, please?

OPERATOR: At this time there are no questions.

CHAIRMAN LUTZ: Shall we vote on whether to eat lunch?

MEMBER ALVARNAS: I Vote yes.

MS. BOSSLEY: Any comments in the room? Okay.

(Whereupon at 1:15 p.m. the above-entitled matter went off the record and resumed at 1:47 p.m.)

A-F-T-E-R-N-O-O-N  S-E-S-S-I-O-N

1:47 p.m.

CHAIRMAN LUTZ: All right. So we're going to get started again with 1855, which is another HER2 discussion. We have our submitting group here. And I think Heidi was going to give us thoughts about how we should have the work group sort of present as we vote.

MS. BOSSLEY: So I have a request. You all might not like it, but it is a request. To standardize across our different committees across the different topic areas, it's most helpful if we have you discuss importance. So all three set criteria first and then vote on importance. Then move onto to scientific acceptability. Discuss that. Then vote. That's what we did the last time. And again, for consistency's sake, we kind of got away from it this morning. I'd like to bring us back and have us do that.

I don't think it will take more

time, but it really helps people -- I think the developers follow the discussion. It helps staff to be able to capture the rationales. And when we go back to try to capture and make sure we got it all, it's much easier to track that way and it is better in mind a thought process. So if you all are willing, my request is that we go back and do it that way. No, not repeat. Not at all. Starting from 1855. I would never ask you to do that. I promise.

CHAIRMAN LUTZ: She means 1855, the submission, not the year.

MS. BOSSLEY: Right.

MS. FRANKLIN: So if we could have the developers for 1855 give us an overview. And I would just like to note that this is also a time limited measure, or it's eligible for a time-limited recommendation for endorsement.

MS. BOSSLEY: Everyone remember when it's time-limited what that means? No,

everything -- I wanted to make sure. So for time-limited it means they've provided all the information with the exception of reliability and validity data. So under reliability and validity for site specific acceptability, you will specifically just look at whether they've provided precise specifications. That's it. Because you won't have anything else. So on that one I think we have provided specifically for that so you're sure you know what you're voting for. Make sense?

(No response.)

MS. BOSSLEY: Okay.

DR. SPEIGHTS: Are we ready? Okay. 1855 is a quantitative HER2 evaluation by immunohistochemistry. Uses a system recommended by the ASCO/CAP guidance.

MS. FRANKLIN: Sorry. Sorry to interrupt.

DR. SPEIGHTS: That's okay.

MS. FRANKLIN: Could the participants on the phone please mute your

lines if you're not speaking? Thank you.

DR. SPEIGHTS: Ready? Okay. In discussion of the last three measures we saw that HER2/neu testing is essential in determining whether patients do or do not receive trastuzumab. Our measure does not focus on which patients should receive HER2 testing as much as if we're going to do it we need to do it right and report it in a reproducible and clinically relevant manner.

Several years ago it was noted that when people -- when patient samples which were tested for HER2 at one facility were subsequently retested at a reference facility, then there was discrepancy in a set to 25 percent of the cases. This led to the ASCO/CAP guidelines for all phases of HER2 testing being published in 2007.

In 2010; actually two years ago this month, there was a survey of about 700 labs which showed about 84 percent of them were using the CAP/ASCO recommended

guidelines. So we see that there is a performance gap. We feel this is a very important measure. Obviously, we've talked about the large numbers of people with breast cancer and the high impact of appropriate therapy for these patients and the need for selecting the appropriate patients to be administered trastuzumab.

We see then that it basically is a very important measure in the sense that it has very important implications for patient care, there is a documented performance gap, and that we are focused on assuring that the key information from the pathology testing for HER2/neu is done in a standard manner and reported in a standard manner. You've already seen some of the criteria for HER2/neu reporting in discussion of other measures. So basically, we feel that IHC evaluation of HER2/neu should be reported in a consistent manner as indicated by the ASCO/CAP guidelines.

CHAIRMAN LUTZ: Okay. And I think -- who is our discussant for this one?

MEMBER FIELDS: I am.

CHAIRMAN LUTZ: Karen.

MEMBER FIELDS: So I think that was an excellent summary. And I just wanted for the group to add a couple of other issues.

So the measure itself measures the percentage of patients with quantitative breast HER2/neu IHC evaluation who either use the ASCO/CAP recommended either manual system or computer-assisted system with an algorithm that includes when to --

(Whereupon, there was interference from participants on the phone line.)

MEMBER FIELDS: You want to try again?

MS. FRANKLIN: To those participants on the phone, if you're not speaking, please mute your lines. And, Arnika, could you let us know if you can mute that line?

OPERATOR: Yes, one moment.

MS. FRANKLIN: Arnika?

OPERATOR: Yes, one moment.

MS. FRANKLIN: Okay.

MEMBER FIELDS: Okay. So the numerator is all patients receiving quantitative HER2 IHC testing according to the guidelines, and the denominator is all patients who got HER2/neu IHC testing. So there were no exclusions. And as we noted, it's a new measure.

I think for the group to understand the reason for the performance gap also is the FDA indications and the manufacturing recommendations for the measurements differ from the ASCO/CAP guidelines. So ASCO recommends to call a positive IHC test. It's 30 percent of the cells completely take up the dye, and then it's positive. Less than 30 percent, then we recommend FISH testing or we recommend HER2 CEP17 testing just to verify whether or not

HER2 is over or under-expressed in those tumors. And then less than 10 percent is negative. The manufacturers recommend more than 10 percent is positive. So that's the difference between the disparity and why some labs may not adequately be reporting.

Also, a comment from a clinical standpoint. Usually it falls on the clinician to go back and request the testing if you get the equivocal results rather than it's an automatic. The pathology department automatically follows those guidelines. At least that's been the way over the years it's evolved for trying to get those equivocal tests redone so that the provider could use the information about whether or not to treat a patient with trastuzumab or not.

So we'll discuss section 1, impact. Obviously, breast cancer, there's a very high number of diseases. It's costly to treat and trastuzumab is one of our most costly drugs and contributes to the overall

cost. So I thought that the impact was high.

The opportunity for improvement I think was well described by the developers, that only 84 percent of the labs surveyed used the ASCO/CAP guidelines.

And the evidence. I'll make a comment on evidence. I think that there's no direct evidence about comparing a tumor marker, in different ways use a tumor marker. It's all direct evidence. The clinical trials where we're describing whether a patient was more or less likely to respond, the measure is an indirect measure because there's central review of the tumors and going back and reanalyzing who was going to respond. So there's a huge body of indirect evidence related to using trastuzumab in these patients, that the ones that truly respond are the patients that have the true positives or have evidence of over-expression of the gene.

So this is a guidelines-based recommendation and the guidelines are very

well written and understandable. So I think that I would have rated the literature as the quantity of the literature was high. The quality was moderate because it's indirect, not direct. And the consistency is high. And so, I felt that it was reasonable -- that's a -- importance to measure was yes, but I open it up for discussion from my other group members and any other comments from the investigators, or the sponsors.

MEMBER HAMMOND: I agree with Karen has said. She has documented in her remarks another source of this performance gap, and that is that in the guideline it specifically says what you're supposed to do if the test is equivocal. It specifies that clearly that you have to do certain specific things, and clearly that's not happening. So the goal of this performance measure is for us to document and try to improve the problem we have with this testing and not following the guideline recommendations, which would, we

hope, make a big difference in what happens to these patients and the accuracy of the testing.

CHAIRMAN LUTZ: Jennifer?

MEMBER MALIN: I had a couple of questions. So under numerator details it says that you report one of the following CPT Category II codes. The first one, 3394F, is quantitative HER2 IHC evaluation, but the second one is quantitative non-HER2 IHC evaluation; e.g., testing for ER, for estrogen and progesterone receptors. I don't understand how that would be a passing criteria for the HER2 testing.

DR. SHAMANSKI: It's because with the codes you cannot differentiate the two types of testing. So we had to have a separate reporting code for testing that was not for HER2.

MEMBER MALIN: But why would quantitative testing not for HER2 meet the criteria for the --

DR. SHAMANSKI: Because if you're coding with breast cancer and with IHC codes and pathology aren't -- they're not specific to HER2.

MEMBER MALIN: But here it says specifically 339 -- am I just --

DR. SHAMANSKI: Those are the reporting codes.

MEMBER MALIN: Right?

DR. SHAMANSKI: Those are not the -- the denominator codes --

MEMBER MALIN: Right. No, I'm saying but the numerator codes. So those are the measure that's specific to HER2, correct?

DR. SHAMANSKI: Correct, but you have to have some way of picking up those cases that are not HER2. They're going to get picked up in the denominator, so you have to have some way of reporting them.

MS. BOSSLEY: But for performance it's only the 3394 that counts?

DR. SHAMANSKI: Right.

MS. BOSSLEY: Correct?

DR. SHAMANSKI: Correct.

MS. BOSSLEY: So actually --

DR. SHAMANSKI: For reporting, it's for both of them so that you can account for those cases, which are approximately 50 percent of the cases.

MEMBER MALIN: Okay. So maybe this just needs to be clarified.

DR. SHAMANSKI: Yes.

MEMBER MALIN: Because the way this is worded, it looks like if you --

MS. BOSSLEY: Right.

MEMBER MALIN: Yes.

MS. BOSSLEY: Right. It looks like -- right now if you read this, I would interpret that both of these would count for the numerator.

MEMBER MALIN: Right.

MS. BOSSLEY: But that's actually not the case.

MEMBER MALIN: It's basically --

MS. BOSSLEY: So I think we need to --

MEMBER MALIN: -- having either one of those --

MS. BOSSLEY: Yes.

MR. MALIN: -- puts you in the denominator. And then the only thing that counts for the numerator is -- so we can work with the developer to make sure that's clear.

MS. BOSSLEY: So we can work with the developer to make sure that's clear.

MEMBER MALIN: Okay.

MS. BOSSLEY: Yes.

MEMBER MALIN: Okay. I think I may have just missed this. Is this a time-limited one?

MS. BOSSLEY: Yes.

MEMBER MALIN: Okay.

CHAIRMAN LUTZ: Okay. I think Elizabeth and then David. David?

MEMBER PFISTER: It was a little unclear to me. Is the denominator here any

pathology reading? So for example, let's say that someone has their slides evaluated locally, then kind of goes to another place, has their slides reviewed. The second place probably sort of sees what was done the first time and may dispense with certain things because they sort of view it already been done. And how is that captured as not being non-compliant?

DR. SHAMANSKI: So the measure is physician-specific. So it's just saying as a physician if you're doing this sort of evaluation you are using the ASCO/CAP guidelines regardless of whether there's been previous studies or not. I don't understand why you would not want to do that.

MEMBER PFISTER: No, I was just saying if it is physician-specific. So I'm good with that.

DR. SHAMANSKI: Okay.

MEMBER PFISTER: So but then let's say you've got two different pathologists that

cross paths on this case. And so, you have pathologist 1 that maybe was the first intake and follows the guidelines and gets it done. Then the second pathologist might confirm a diagnosis of breast cancer, might kind of be mindful of what had been done already with the other pathologist. And how is that eventuality sort of captured in a way that doesn't penalize the second pathologist?

DR. SHAMANSKI: If the second pathologist is actually not doing a HER2 evaluation, it won't get picked up in the denominator.

MEMBER PFISTER: Yes.

MEMBER HAMMOND: It wouldn't be able to charge for that.

MEMBER PFISTER: Yes.

MEMBER HAMMOND: Those are charging codes.

MEMBER PFISTER: Yes.

MEMBER HAMMOND: So they would not be able to charge for HER2 and therefore they

would not be measured about it. That code would never be in the system. That clear?

MEMBER FIELDS: So I guess what you're saying is the trigger is always when you order HER2 IHC and then it needs to be done correctly?

MEMBER HAMMOND: It's not when you order. It's when you do it.

MEMBER FIELDS: When you do it?

MEMBER HAMMOND: Yes, you do it.

MEMBER FIELDS: When you do it?

MEMBER HAMMOND: Yes. Right.

MEMBER FIELDS: And so, then any other ordering of FISH or variations on amplification isn't related to this measure?

MEMBER HAMMOND: Correct.

MEMBER FIELDS: Okay. Is that --

DR. SHAMANSKI: Yes.

MEMBER PFISTER: So then, I mean, I'm just thinking in real time like how these things kind of come through. Maybe Steve can comment on this. But like, let's say one of

the breast pathologists might submit some slides. They kind of put in like the order. It gets kind of processed. And arguably they may end up doing a HER2 that's redundant on what's been done previously. And then they don't do any for the work of knowing what's been doing previously. But having done that HER2, then even though they're not following up on it further because it would be redundant, they're going to get penalized for having done in the first place.

DR. SPEIGHTS: I mean, our measure really just focuses on whether the pathologist uses the ASCO/CAP guidelines for interpretation. Other problems such as not knowing a previous result, repeating the test, difference in interpretability and interpretation between pathologists are not really the focus of this.

MEMBER EDGE: On this test when you did the HER2 test you used the guidelines for testing as recommended by ASCO/CAP, NCCN,

whatever? And then that should be documented in the path report?

MEMBER HAMMOND: Right, and the guideline states that anybody who looks at a HER2 test should be using the guideline recommendations. So anybody who does that first or second time, it doesn't matter. They should be using the same criteria.

MEMBER EDGE: So is this something that should be measured on a case-by-case basis, or is this --

MEMBER HAMMOND: Yes.

MEMBER EDGE: -- something that is better measured on a laboratory-by-laboratory basis? Like, you know, if I have my blood sugar measured, I'm supposed to be in a laboratory that has documented that they measure blood sugars accurately. Shouldn't the same thing be true for this? Isn't this a CLIA issue?

MEMBER HAMMOND: Well, there are two parts to the test. In the guideline, this

is made clear. So there's laboratory component and there's pathologist component. This measures only the pathologist component. We need to have a measure -- and hopefully the measure developers are hearing me say this. We need a measure for the laboratory component as well. That's whether or not the test was accurately done and the specimen is handled correctly. So by institution. We should have a measure by institution as well as a measure by physician, just like we've talked about with these other measures that we've discussed previously.

DR. VOLK: Dr. Hammond, this is Emily Volk. I'm part of the Measure Development Team here. I think we certainly appreciate the content of that comment. I'm a little unclear on how we would operationalize that with the parameters set by the PQRS program.

MEMBER HAMMOND: Well, I don't know. The answer is, Emily, I really don't

know, but I know there have been measures that we've discussed where they were institution-specific. Maybe CAP is not the one to make this measure, but it would be nice if we had measures that were measuring whether or not laboratories were compliant with this guideline. That means that they're watching the fixation of the sample, the way in which the test was done, the quality indicators for that laboratory's performance. That's not what this measure is about. This measure is completely about the other part of the test, which is just pathologist-specific.

DR. VOLK: Agreed. Agreed. I'd love to talk to you about that more off line.

CHAIRMAN LUTZ: Bryan, did you have something?

MEMBER LOY: I just want to make sure I understand. You showed us a part of the screen that showed some alphanumeric codes that really made the distinction between HER2 and non-HER2.

(Off mic comments.)

MEMBER LOY: Well, I thought I saw them up here on the numerator statement. There. They're alphanumeric. As a payer, that gives me a little bit of pause because not all systems process those codes.

And then the second question that I had was that there's a CPT code that I'm kind of worried about because it's not necessarily specific for HER2/neu that folks use probably even more frequently than they would the alphanumeric codes that are much broader. They do HER2 and ER/PR and others. How are we dealing with that in terms of --

DR. SHAMANSKI: So just to be clear, the CPT billing codes and ICD-9 codes are the codes used to determine the denominator. These are reporting codes. And so, the reason you have the second code for non-HER2 IHC is to exactly address the problem you're talking about, is that those CPT codes are not specific. So we have to account for

those other cases in some way.

MEMBER LOY: Got it.

DR. SHAMANSKI: And this is the best way.

MEMBER LOY: Okay. So in order to even be measurable, you have to submit these reporting codes, is that correct?

DR. SHAMANSKI: Correct.

MEMBER LOY: Okay. So one other question. If I report 3394F in my numerator, does that mean that clinically I've met the ASCO/CAP recommendation?

DR. SHAMANSKI: Correct.

MEMBER LOY: Or is there a further review of the actual pathology report that's required to meet that criteria?

DR. SHAMANSKI: Well, by reporting that code, it indicates that that was done, that the report meets the criteria.

MEMBER LOY: Okay. Thank you.

CHAIRMAN LUTZ: Is there anybody online that has a question? I don't know, if

Rocco, Heidi, Joe -- if any of you are there, but we don't want to forget you. Anybody?

(No response.)

MEMBER DONOVAN: We're here. I don't have anything to add.

CHAIRMAN LUTZ: Okay.

MEMBER LOY: One other question. FISH. Is there any --

MEMBER FIELDS: What about FISH?

MEMBER LOY: Pardon?

MEMBER FIELDS: FISH is not --

MEMBER ALVARNAS: No comments on my end.

MR. LOY: So if somebody chose to do FISH instead of IHC -

MEMBER FIELDS: It wouldn't qualify for --.

MEMBER LOY: So we're just going to exclude that out of the universe for this purpose?

MEMBER FIELDS: Yes.

MEMBER LOY: Okay.

MEMBER HAMMOND: It's just IHC.

MEMBER FIELDS: Then I don't understand the measure at all, because I thought that it was when to use FISH appropriately to quantify your IHC.

DR. SHAMANSKI: No, we require that laboratories -- well, we don't require it, but we like to have them provide to us a score, which is sort of semi-quantitative, and a quantitative number for the immunohistochemistry as well as the FISH. Both of those could be quantitative tests.

This particular measure only measures the immunohistochemistry part. It doesn't measure the FISH part. So another measure would have to be created to measure whether or not the pathologist is compliant with the FISH codes.

MEMBER FIELDS: But the guideline itself tells you when to use FISH?

DR. SHAMANSKI: Yes. Yes, the guideline --

MEMBER FIELDS: So how can we have a measure that measures if you're doing the guideline if you don't --

DR. SHAMANSKI: Well, because this is --

MEMBER FIELDS: -- do the whole test?

DR. SHAMANSKI: -- one element of the guideline. It's not the entire guideline. As we talked about a moment ago, you know, there are laboratory components, there are FISH components, there's immunohistochemistry components. There are many components of the guideline. One could look at this as a surrogate for all ASCO/CAP guideline compliance. We need measures that tell us whether people are complying with this and get rid of that gap. And so, this is our first effort to try to start to get there.

DR. SHAMANSKI: I think there's a word missing here in the measure title. It's the scoring system, not the system, which was

in our original measure. Just that word seems to have gotten dropped. But, so we're measuring that aspect of the guidelines.

MEMBER FIELDS: Okay. Well, so to get to real quality improvement then, we need the labs to start appropriately interpreting the pathology and ordering the appropriate rest of the work-up, because otherwise we're leaving it to the clinicians to interpret that for treatment decisions. I mean, that's not the point of today. I understand now. You're eliminating it just to saying 1+, 2+, and 3+. That's all you're doing.

MEMBER HAMMOND: According to the guideline, which means that there are requirements in there for how they have to do the test. So they're not supposed to use that reporting code unless they are compliant with the guidelines. So we would assume that this is a surrogate for them doing all the other things you talked about, but we aren't measuring those other things. We're measuring

one element and hoping that it's a surrogate for all the other elements.

MEMBER FIELDS: Just as a clinician, the assumption that we can make is that once we get a 3+, we're done. We don't think about it again. And 2+, somebody's gone and is going to give us another report that tells us exactly what we needed to know.

MEMBER HAMMOND: And this report, if it's equivocal, should have a statement in it that says the IHC is 2+, the IHC HER2 test is 2+ positive. By the ASCO/CAP guideline, that requires that the test be confirmed by doing a FISH test on the same sample, and that report will be subsequently provided. And if those words are not in there, then they haven't complied with the guideline. That's part of the guideline.

MEMBER FIELDS: Okay. Then I guess you need to really change the title to say scoring, because that's a huge difference. Yes, okay. That's fine.

MEMBER HAMMOND: So this is just a surrogate. It's measuring one part of this whole guideline. And we're hoping that it will address the performance gap and make it better in the future.

CHAIRMAN LUTZ: All right. So I'll defer to my NQF brethren here. We've led a little bit further ahead. I don't know, in terms of voting whether you want to --

MS. BOSSLEY: So, I do. I think we should have a vote on the importance because there's clearly a discussion around the evidence and as the measure that's before you. So let's do that and then let's see how it goes against that. And then we'll move onto scientific acceptability, because you moved right into that already. Then we'll move onto the rest, usability and feasibility -- Well, we talked about the -- you did it. You weren't the one who moved into scientific acceptability. Others did. But that's okay.

MS. KHAN: So voting on 1a,

impact.

MEMBER FIELDS: Dr. Ricciardi, are you still on the line?

MEMBER RICCIARDI: Yes, I am on the line. Sorry.

MEMBER FIELDS: Okay. We're just waiting for your vote.

MS. KHAN: Okay. So, we have 11 high, 4 moderate, 1 low and 0 insufficient information.

Moving onto performance gap, 1b.

We're missing two people.

Five high, eleven moderate, zero low, and zero insufficient.

And going onto evidence. Yes, no, or insufficient.

Can we press them one more time, please?

So, we have 14 yes, and 2 no.

MEMBER FIELDS: So moving onto reliability and validity. So the question No. 1 is is the measure precise? And now that we

understand all of the differences in CPT codes and reporting codes, I think that the measure is precise and you would be able to measure it.

Reliability. There's no reliability testing available. But because this is adopted for a one-year period to test the reliability, I think that makes it acceptable for approval.

Validity has to be determined once we determine whether or not it's a Reliable measure. It seems like a valid tool for me as a clinician and as somebody that uses this information to make treatment decisions. So I would assume that it meets validity criteria, or it's worth discussing that.

And the disparities in healthcare don't apply in this measure.

CHAIRMAN LUTZ: Any further discussion on those things? I know we already covered a lot.

(No response.)

CHAIRMAN LUTZ: Okay. Can we vote on those?

MS. KHAN: So we are going to be voting on reliability and validity for untested measures. The measure specifications, numerator, denominator and exclusions are unambiguous and likely to consistently identify who is included or excluded from the target population, identify the process, condition or event being measured and compute the score. And they should also reflect the quality of care problem in 1a and 1b and the evidence cited in support of the measure focus, 1c.

So we're going to vote 1, yes or 2, no.

I think we're missing two people.

So, we have 15 yes and zero no.

And moving onto usability. High, moderate, low or insufficient. You want to discuss it first?

CHAIRMAN LUTZ: Karen, we

definitely need to hear what you have to say about that.

MEMBER FIELDS: Usability? I don't know that I understand what the public reporting implications would be at this point in time. I think it's a useful measure for quality improvement, however. So I would say that it seems to meet the usability criteria.

And not feasibility yet, so --

CHAIRMAN LUTZ: Anything else about usability?

MS. KHAN: So usability. High, moderate, low or insufficient.

I think we're missing one person.

All right. Six high, five moderate, two low and two insufficient.

MEMBER FIELDS: And feasibility. Yes, it's definitely data that's generated as a byproduct of the process. It should be available on electronic formats. And I would assume that it has a moderate susceptibility to inaccuracies, but it should be fairly

reliable. And I think that the strategy that they outlined to collect the data is feasible.

CHAIRMAN LUTZ: Anyone else?

(No response.)

CHAIRMAN LUTZ: Okay.

MS. KHAN: Feasibility. High, moderate, low or insufficient.

So, we have 4 high, 11 moderate and 1 low.

And overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? Yes or no.

Fifteen yes and one one. So the measure will pass.

CHAIRMAN LUTZ: All right. Not to confuse anyone, but the next one I think by virtue of who's available to be moved up in line is 0391. I know, Elizabeth, you have to leave in four minutes, correct? Do you have the capability in four minutes and four seconds to tell us what we need to know?

MEMBER HAMMOND: I think I can.

MS. FRANKLIN: Sorry. If there's a developer on the line from AMA-PCPI, could we please open their lines, or from College of American Pathologists? Or if they're in the room? Okay. There they are. Okay. There they are. Sorry. Okay.

MEMBER HAMMOND: All right. This is a maintenance measure that was originally endorsed in 2008. It is a measure that seeks to show that the staging information is being collected on all patients with breast cancer resection specimens. It has been shown over and over again in the literature that staging information is very critical to patients. We've talked about this in other cancers at our last meeting. And the data and the way in which this is presented is very analogous to those other sites. So staging information is used to treat patients and this is an attempt to collect that staging information and to demonstrate whether or not it's present.

The impact of breast cancer is

high. There is a performance gap related to proposing this or recording this staging information. As we said last time when we were talking about other places, we know that the outcome of a patient is directly related to stage, but whether or not the recording of stage relates to outcome is not necessarily known. This is a process measure and it is supported only by indirect data, but there's a lot of indirect data that supports it.

Because I'm going to be leaving, I would like to just go on and mention my thoughts about acceptability. I think the reliability of this measure is very high. The data is collected in a meaningful way and the measure is a valid measure, although I would rate its validity as being moderate. The information would be meaningful to the public because staging information hopefully is something understood by the public. So I think it has a high usability criteria. It is feasible to collect since the data is

generated during clinical care. So I believe that this measure should be accepted for endorsement.

And there were no specific issues that I felt needed to be addressed. Let's see. Oh, the only thing that was brought up that I think is really a serious problem that can't be addressed by this particular performance measure is that often there's staging information embedded in several pathology reports, and one of the difficulties is how do you decide which pathology report you would use.

Typically that's the latest pathology report is usually the one that is usually used, but in some cases it's the initial report. And because we don't have valid codes to measure a summary report or we don't even have a form of a summary report yet, that issue cannot really be adequately addressed. But it occurs across all of pathology reporting. It's not specific to

this breast cancer measure.

So basically, the situation is very similar to the measures we passed at our last meeting related to other cancers and staging measures. Does anybody have any questions for me before I run out the door?

MEMBER MALIN: Maybe I missed this. What happens if it's just an excision and the lymph node biopsy hasn't happened yet?

MEMBER HAMMOND: Well, if it's only an excision, there won't be any lymph node status. But typically in that situation what should be said is that the lymph node status would be designated as an X, which means that the person writing the report has no understanding about the status of the lymph nodes at that time. So if you look at all the pathology staging reports, you should find one where there's the most information, and that most information should be the one that's used.

So if you're only looking at an

excision specimen, it will be pT with a number, pN with an X, and pM for metastasis with an X. But if there are lymph nodes, it will be both.

(Off mic comments.)

MEMBER HAMMOND: Oh, there isn't?

(Off mic comments.)

MEMBER HAMMOND: All right. Well, I should have left before I -- Well, then I was not --

MEMBER EDGE: There is no such thing as MX. So they will not be listed as MX. I'm sorry.

MEMBER HAMMOND: So what do you do in the situation where you have no knowns? Do you record it as being --

MEMBER EDGE: No, M. M. Metastases. There is no MX.

MEMBER HAMMOND: Oh, there's no M? Okay. But there is an NX?

MEMBER EDGE: A patient is either clinically M0 pathologically --

MEMBER HAMMOND: Oh, good.

MEMBER EDGE: -- M1 or clinically M1.

MEMBER HAMMOND: All right.

MEMBER EDGE: There is no such thing as MX.

MEMBER HAMMOND: Okay. So there is a way to tell.

CHAIRMAN LUTZ: Thank you, Elizabeth, and safe travels.

MEMBER HAMMOND: Thank you.

MEMBER LOY: If you caught them in a slice time where they had gotten an excisional biopsy and they wrote down p and NX, would that be counted as compliant before they'd gotten the full specimen?

MEMBER HAMMOND: Yes, because they might never get another specimen.

MEMBER LOY: Right. Right. Okay.

MEMBER HAMMOND: It would.

MEMBER LOY: So as long as they have used the appropriate notation --

MEMBER HAMMOND: Codes.

MEMBER LOY: -- no matter where you've gotten them --

MEMBER HAMMOND: Right. Right.

MEMBER LOY: -- they could --

MEMBER HAMMOND: Right. There is a strong meet though; and we talked about this on the conference call, for something called an integrated report, which would be at the end where all the information was recorded in one place. The College of American Pathologists is actually working on this through their electronics interfacing groups trying to come up with something like that. And at that time, when we ever get it, that will be something we can bring back for a measure.

CHAIRMAN LUTZ: I know we're a little out of order, but do our AMA or CAP folks have anything to say?

DR. SPEIGHTS: I don't think we have anything to add. Emily?

DR. VOLK: Nothing to add.

CHAIRMAN LUTZ: Okay. Then I guess we need to go first to importance.

MS. KHAN: So, voting on 1a, impact.

CHAIRMAN LUTZ: Is there any further discussion on importance?

(No response.)

MS. KHAN: Oh, we are voting on 1a, impact. So if you could send your votes in to Lindsey.

Twelve high, three moderate and one low.

CHAIRMAN LUTZ: Anybody have comments about opportunity for improvement?

(No response.)

MS. KHAN: So voting on 1b, performance gap.

So it's nine high, five moderate, one low and one insufficient.

And voting on the evidence. Yes, no, or insufficient.

CHAIRMAN LUTZ: Any further comment on evidence?

(No response.)

CHAIRMAN LUTZ: Okay.

MS. TIGHE: Dr. Marks, can you send your vote, please?

MEMBER MARKS: Sorry.

MS. KHAN: We have 14 yes and two no.

CHAIRMAN LUTZ: Any discussion about reliability?

(No response.)

MS. KHAN: Voting on reliability.

Can everyone just press it one more time, please?

So that's 10 high, 4 moderate, 1 low and 1 insufficient.

CHAIRMAN LUTZ: All right. Anything additional about validity testing?

(No response.)

MS. KHAN: Voting on 2b, validity.

So we're missing two votes. If

you could press it one more time.

So four high, eight moderate, two low and one insufficient.

CHAIRMAN LUTZ: Anything about usability? Bryan?

MEMBER LOY: One shortcoming we might have identified here in the process is that we may not have complete information. It seems to me in order for this to really be linked to a health outcome, even indirectly, you would want what Dr. Hammond had advanced before, and that is, you really want the complete integrated report. So if we give somebody credit for something, meaning they did the appropriate pathologic staging on an excisional biopsy but that didn't get it accomplished when they actually did the node dissection and all the accompanying pieces of it, we might find that our results might not reflect what we're really trying to measure.

Have you all given any thought to that, measure developers? I mean, I

understand it doesn't need to be perfect. I'm not trying to say that it's still not useful. It just seems as though it kind of clouds the issue, if that makes any sense.

DR. SHAMANSKI: Can I just add one point? This is on resection. Biopsies are not included in this measure.

MEMBER LOY: Okay. Well, I misunderstood her, then. I thought that if it was an excisional biopsy and it was staged properly, that you got credit in the numerator, is what I thought I heard. Is that not true?

PARTICIPANT: No.

MEMBER LOY: So only when you have a complete --

DR. SPEIGHTS: If an excisional biopsy or lumpectomy, tylectomy, whatever names it goes under, can completely remove a tumor, it may or may not be accompanied by lymph nodes.

MEMBER LOY: Correct. Okay.

DR. SPEIGHTS: With this, as with any measure, all we can report on is what we have. And we really need to have the complete tumor resected and the margins free to really say the T category (telephonic interference) big it is.

MEMBER ALVARNAS: I'm sorry, but how is that related if the path report doesn't have an N stage result? For example, how do we differentiate just a T stage, but not an N stage? How do we differentiate that? How are we differentiating not meeting the criteria versus not having nodes submitted, for example?

DR. VOLK: Again, you would use the NX designation if nodes were not submitted. This is Emily Volk from the Baptist Health System in San Antonio, and I'm a practicing pathologist here. And I think what this measure does is encourages the most accurate up-to-date staging at every point along the way in the patient's journey.

DR. SHAMANSKI: And I would just add; this is Fay Shamanski from CAP, that it's breast cancer resection pathology reporting. The CPT codes that are included in the denominator are 88307 and 88309, if that means anything to you. Those are not biopsy codes.

MEMBER ALVARNAS: Are those breast surgeries or axillary surgeries, or both?

DR. VOLK: Both. Any time there's a margin that needs to be evaluated, it changes the code from a biopsy code to a resection code.

MEMBER ALVARNAS: So for example, a patient goes in for axillary surgery, but they don't enter the breast again, is that going to be captured? Because that would be a pX if they don't have the old report, for example.

DR. SPEIGHTS: Again, the most we can report on is what we have. It is possible that the tumor may be resected at one facility. Patient goes elsewhere and then has

an axillary dissection in which the lymph nodes are removed.

MEMBER MARKS: But (telephonic interference) so the pathologist is given appropriate credit, if you would, reporting what they have based on the information available to them.

DR. VOLK: This measure would capture that.

MEMBER FIELDS: So we had this same question. I brought it up on the group call. And in breast, it's very common that they have multiple re-resections. So I think unless we get to the point of really trying to have summary reports, it still won't give us the level of quality we need in this particular disease. It's true that other diseases have multiple resections for margins, but in breast it's pretty traditional that you have the lumpectomy. And returns to the ORs are not that uncommon. And there's multiple stage procedures. So that was our --

MEMBER MARKS: But I think the majority of patients have read the synoptic report, the synoptic path report. Right?

MEMBER FIELDS: So I don't think the resection code issue answers or solves the problem about getting to the quality end point that we need, which is we need to know what the TNM stage is before we make a treatment decision.

MEMBER MARKS: Do you have any clinicians who actually make the decision based on the path stage given the path report as opposed to the later staging based on assimilation of the two or three path reports that we have in the clinic?

MEMBER FIELDS: I would say I know a whole bunch of clinicians, because we're relying on the pathologists to tell us what the stage was. I'll let the surgeon answer that.

MEMBER EDGE: I think the question was does he know a clinician who rely on the

single path report rather than both the aggregate of all the path reports, plus the imaging studies, plus the clinical examination that goes into it? And I don't ever make a recommendation --

MEMBER MARKS: Right.

MEMBER EDGE: -- based on a single path report. To me unfortunately it makes me concerned that this measure -- this is why this measure really isn't linked to outcome.

MEMBER MARKS: Right.

MEMBER EDGE: And it makes me really struggle with whether we should be approving the measure. Did the person write down on a piece of paper as opposed to did the doctor provide a treatment that was appropriate for the true stage of the patient? But that's another question.

MEMBER MARKS: No, I agree.

MEMBER MALIN: I mean, I think, you know, some of the times the pathologist-specified T stage can be misleading. So let's

say it's the third excision, or whatever. And I've seen this happen before, you know, either it was a different pathologist at the same institution or a different institution that they didn't aggregate across. And then you see on the third path report, you know, a specific, you know, T stage that's just reflecting the tumor that they got out of that specimen, not the two other things that happened before, and it can be wrong. And so if you as the clinician aren't making sure that you've checked it -- so at least personally I don't ever rely just on the pathologist-specified stage. I always calculate myself.

MEMBER FIELDS: Well, I mean, I think -- but you have to have all of the information. And --

MEMBER MALIN: But have you looked at what they gave you?

MEMBER FIELDS: Except for the most common scenario where the pathology stage

is what you needed or the timbers that had positive margins, but you go back in the -- there were close margins and the margins were clear. And so, the bottom line is somehow we need to get to the point where somebody does a summary of the data that we have so that it's not in multiple stages. And there are some pathologists that are very compulsive about that and do that. And then there are some that just don't do that.

And so I guess opportunities for the future would be getting to that level of reporting so it's helpful. And it's true, you have to use everything. You have to use physical exam, radiographic images, and everything else. But there's lots of times where you had a close margin. You go back because the margins were close. There's nothing there The path stage is the stage. So the answer is yes lots of time.

MEMBER MALIN: But if it's that one report -- I mean, maybe this is being

harsh, but how much added value is it for them to go ahead and put it in a category versus just seeing the tumor size there on that report?

MEMBER FIELDS: If it's in a summary document, it would be lots of value.

MEMBER MARKS: But there is no construct beyond that summary document.

MEMBER FIELDS: We're just saying that that was our request.

MEMBER MARKS: I agree it would be nice if they read the summary document, but if there is a mechanism to generate that, I'm not sure of the validity of this metric.

MEMBER LOY: I can't disagree with what you just said, but I have to say given where we are -- and we live in a world where we don't have that synthesis of all path reports -- this certainly has to be more desirable to have this document than to not be documented. So it certainly seems like a valuable step given where we are today, but it

certainly seems like we should be making recommendations for the future world.

CHAIRMAN LUTZ: So necessary but not yet sufficient, or not yet comprehensive?

MEMBER LOY: I would agree with that.

CHAIRMAN LUTZ: Okay.

MEMBER MARKS: I'm not sure if this would come into your discussions or not, but on the opportunity costs. And yes, this might be a good first step, but this is -- you know, perhaps other measures that may be -- one could spend one's energy on that might be more useful in the pathology realm. I don't know what those are, but I'm just saying, it seems like a good step forward doesn't mean necessarily we should do it because there are opportunity costs.

MEMBER FIELDS: Just one final comment though. I think they showed us a huge performance gap which was --

MEMBER MARKS: That's true, yes.

MEMBER FIELDS: -- 32 percent of the reports don't have all the elements. And then there's another report that's similar to that. So I'd have to say mom and apple pie comes first and then we get to better levels of reporting and quality.

MEMBER MARKS: That's fair.

CHAIRMAN LUTZ: All right.

DR. SPEIGHTS: Obviously, you know, we have to report on what we have. We can't say the tumor size unless we have a completely resected tumor. And sometimes we just have to say that according to an outside report there was a one-centimeter tumor seen that involved the margins elsewhere. We have another centimeter tumor here and we have to give our best assessment of the final T stage based on what we see.

Now our path reports, at least in my institution, say something to the effect that this staging information is based on the pathology specimen. There can always be a

lung CT or something that shows a metastasis that we aren't privy to that could upstage. But I think what we're trying to do is to close the gap so that we give appropriate T and N categories whenever we can.

CHAIRMAN LUTZ: All right. Anything else? We're actually -- Elaine? Sorry.

MEMBER CHOTTINER: Is there any attempt to incorporate the clinical staging in any way because of the larger number of patients who are receiving neoadjuvant therapy where the clinical stage is actually going to be more accurate? Is that reflected in the reports?

DR. SPEIGHTS: If the patient has received previous treatment, it should be a Y, and there should be a Y in front of the T, receiving neoadjuvant or what have you, which implies a caveat that we report again what we see pathologically, but that hopefully the neoadjuvant treatment has downstaged the

disease.

CHAIRMAN LUTZ: All right. I think we left off on voting on usefulness. Is there anything else to add for that?

MEMBER MARKS: Did we actually vote on that yet?

CHAIRMAN LUTZ: We're just about to.

MEMBER MARKS: Okay.

MS. KHAN: So voting on usability. You can go ahead and send your votes in now.

So we have four high, eight moderate, three low and zero insufficient.

CHAIRMAN LUTZ: Is there any additional discussion about feasibility?

(No response.)

MS. KHAN: And voting on feasibility.

So we have five high, eight moderate, two low and zero insufficient.

And overall suitability for endorsement. Does the measure meet NQF

criteria for endorsement? Yes or no.

I think we're missing one vote.

So we have 12 yes and 2 no. So the measure will move forward.

CHAIRMAN LUTZ: All right. Then the same logic applies in terms of who's available to present for us.

Next one is 0392, which would have been after the break, but we'll do it now. Anything from our AMA or CAP folks to give us the framework?

DR. WITTE: This also is a maintenance measurement. It was developed by a broad multi-disciplinary group convened by the AMA and supported by the College of American Pathologists Use Guidelines. It's been in use in multiple places. It's been in the PQRS program. Obviously colon cancer is frequent. The gap in the most recent data was about 25 percent. It focuses on guidelines and it focuses on those elements of the guideline (telephonic interference) useful in

guiding therapy. And it has been useful and, we believe, reliable.

CHAIRMAN LUTZ: All right. We have John and Bryan listed as double-teaming this one.

MEMBER GORE: So basically when we look at importance, we discussed the prevalence of colon cancer, being a very common cancer among men and women and the need for accurate pathology reporting. As for example, distinguishing between stage 2 and stage 3 colon cancer, it is very important to delivery of adjuvant therapies. And in terms of performance gap, the surprising identification of inaccurate complete pathologic staging in up to 25 percent of the pathology reports missing elements such as grade or nodal status. And so, in terms of importance, our work group universally declared this to be an important measure to report.

Bryan, did you have anything to

add?

MEMBER LOY: No.

MEMBER GORES: In terms of disparities, there's not really much on there, but in terms of importance to measure and performance gap.

CHAIRMAN LUTZ: Anybody have anything to add on importance?

(No response.)

CHAIRMAN LUTZ: All right. Should we vote on 1a?

(No response.)

CHAIRMAN LUTZ: So for those of you on the phone, we are voting on 1a.

MEMBER ALVARNAS: Thanks for clarifying.

CHAIRMAN LUTZ: So we're going to go back to square one on voting on 1a. One moment.

MS. KAHN: Importance to measure in our report. Impact. One high, two moderate, three low, four insufficient.

We need three more votes. If you could try voting again.

DR. TIGHE: Dr. Marks, if you could send me your vote.

CHAIRMAN LUTZ: I'm sorry, David. Did you have something to --

MEMBER PFISTER: Yes, in my colorectal -- in a trip, we make sense, because you know obviously grade and reports, but what management decision is association with grade of a cancer?

DR. WITTE: I'm sorry, could you repeat that?

MEMBER PFISTER: Yes, like it certainly makes sense that they use the T and N data to make a management decision, but under what circumstances does grade affect that management decision? You know, once you've got a diagnosis of invasive cancer?

MEMBER RICCIARDI: This is Rocco Ricciardi from Lahey. The only thing I could think that would be of any value that we use

would be with a T1 tumor based on the depth of invasion in the submucosa and the grade of the tumor.

MS. FRANKLIN: You faded out. Hello? You faded you just a bit.

MEMBER RICCIARDI: Oh.

MS. FRANKLIN: Could you repeat that?

MEMBER RICCIARDI: Yes, what I was saying is that based on the grade and sometimes the level of invasion into the submucosa we'll base a decision about a T1 rectal tumor as to whether or not treat it locally versus a more extensive resection.

DR. SPEIGHTS: Following up, sometimes when colon polyps are locally resected whether it's poorly differentiated or not can be determinative of whether to do a more extensive resection or not.

MEMBER GORE: One thing, we also were curious on the call was about why it's just T, N and grade and not margin status?

DR. WITTE: We are trying to remember the discussion on that. I apologize for not being able to bring back five-years-ago discussion. I think part of the reason was there wasn't -- when we reviewed the data, if I remember correctly, that what was missing was not that, so the gap -- we tried to pick the stuff that was higher gap, is what I recall, but I'd have to go back and review that.

CHAIRMAN LUTZ: Jennifer?

MEMBER MALIN: I think, you know, similarly to some of the discussions we've had about some other measures, you know, it might be worth considering updating this measure. You know, grade is not so important, but number of lymph nodes evaluated is. And I think certainly margin status is arguably more important. And then other things like, you know, if you want to look at things that actually impact outcomes. Things like lymphovascular invasion, you know, evidence of

rupture, things like that would be more relevant.

MEMBER LOY: Yes, I think we're going to get there, but further on we're going to hear some other things that might be more contemporary like KRAS testing, etcetera. I think all signs; at least in my view, are pointing towards a more synthesized report that encompasses all the clinically and important and molecular diagnostic predicted biomarkers, et cetera, into one report. And I just don't think we're there yet.

MEMBER MALIN: But I think this is just kind of a global issue. You know, with a lot of these measures that we're seeing for re-review, they were sort of barely reaching a threshold the first time they were, you know, endorsed for being kind of relevant and driving improvement. And for, you know, what, is it five years later, to not have something that's trying to move the bar, I personally find disappointing. So I think it would just

be, you know, good to re-look at the evidence, not just to bring back the same measure, but to have a little more responsibility on the part of the measure developers to see really what needs to be done to move the bar.

CHAIRMAN LUTZ: John?

DR. WITTE: We certainly are sensitive to the synthesization of the report as we were on the previous measure, and that certainly is in our docket, as Dr. Hammond indicated. There still remains a performance gap for this measure. I think the criticisms are registered and taken to heart.

MEMBER ROSS: So I have a question on the performance gap. So is that the data that was originally presented in 2008, or is that current data? I'm confused about that, the 21 percent.

MEMBER GORE: Looking at 1b, the data does give for the demonstration of performance gap is 2008.

MS. CHRISTENSEN: Yes, so the data

there, the 10th percentile, the 25th, 50th, 75th and 90th percentile is 2008 data, which is unfortunately the most recent that CMS has been able to make available for us to report publicly.

MEMBER ROSS: So I agree with Jennifer about raising the bar. And we've now -- at last meeting and this one have sat through a number of those in which we're validating staging, which is the essence of oncology care. And I still remain surprised that in 2012, we're revalidating staging.

But this doesn't seem to make sense, because in the last four years there have been so many presentations at all of the oncology meetings, NCCN, addressing the points that you're talking about, Jennifer; number of nodes, how the resections are done. And to just go ahead and validate another staging that is at least four years old in terms of the data that documents a gap that may no longer exist doesn't make sense to me.

DR. ANTMAN: Mark Antman speaking for the PCPI. So unfortunately, as Keri was saying, we can only report the most recent CMS data that we have for the PQRS system. And obviously if we had more recent data, we would provide that.

If I may jump off of what Dr. Witte said a moment ago, this is very valuable feedback. This is a measure set. As we've been saying, it is five years old. And so that means that it is one of the measure sets, one of the PCPI measure sets that is certainly due for a review and for an update. And by all means, the recommendation of this steering committee will be paramount in the discussions of the work group in considering how to update the measures. So I think we have to defer to this committee as to whether or not you feel that the measure as it stands is still beneficial and is still better to retain endorsement rather than have no measure in the meantime until we're able to update it. But

by all means, these recommendations I think will be very useful for knowing exactly how to do that update.

CHAIRMAN LUTZ: Do Pat and then John.

MEMBER ROSS: Kind of a follow-up to that. So I guess I may not understand the process; and I should because we've now sat through three days of it and a few conference calls, but I'm a surgeon, so indulge me.

So in 2008 a committee like this validated this and said move this forward. Am I interpreting that correctly?

DR. WITTE: Yes.

MEMBER ROSS: No?

DR. WITTE: Well, partially I'd say.

MEMBER ROSS: So let me finish the question and then perhaps you can educate me.

DR. WITTE: Well, not much.

MEMBER ROSS: But how can we now say -- if we validated this, someone collected

the data, right? We've been collecting this data for four years and you can't give us any new news?

DR. WITTE: Well, let me just say that the data that was presented when this was originally approved was data that came from literature studies, not from performance measurement formal program studies.

MEMBER ROSS: But where's the data that's been collected for the last four years?

DR. WITTE: That's what Dr. Antman spoke to.

MS. CHRISTENSEN: So CMS is the organization that runs the PQRS program. And they provide information back to doctors, but they do not make that information publicly available. So we are unable to give you data because they don't make it available to us either. So they are collecting the data. They are looking at the data and they do make a determination every year about what measures they're going to keep and what measures they

will retire from their program, but we do not have the data that we are able to give to you for anything more recent than 2008 just because it's their data.

MEMBER ROSS: Does that make sense?

MS. CHRISTENSEN: We would very much like the data that is more recent, but unfortunately the government gets to make that decision.

CHAIRMAN LUTZ: The way it's been described to me is we give these measures, they sit up on a shelf, and based upon what Medicare sees from real-time data, they decide which ones to take off the shelf and use and which ones to put out in the trash. Is that too simplistic? That was the way it was described to me.

MEMBER ROSS: It seems somewhat not reasonable to just revalidate it. I don't know. Doesn't make sense.

MEMBER GORE: So, I mean, this is

a comment I was going to make: the question then becomes is this something analogous to what we did with the melanoma measures last time? I mean, there are clearly elements of this that are important and we all believe in accurate pathologic reporting, but there are more elements to the path report that we know are important that maybe were less useful five years ago. Do we just recalibrate the measure? You know, I don't know.

MS. FRANKLIN: We have to --

MEMBER GORE: I know we have to evaluate it as is.

MS. FRANKLIN: Yes, as is and go through the criteria and vote. If you find that you don't think the evidence is sufficient to support the measure, you can still make a determination as a steering committee, if you want to vote to move the measure forward if the benefits outweigh the harms.

MEMBER GORE: Yes, and I think the

hard part is I think it's important and I think it meets a lot of the criteria, but it could be better.

MEMBER ROSS: Right. So I guess to the sponsor, I mean, why isn't the burden on the six of you to have brought us an updated version instead of just bringing us one that is in a maintenance mode?

DR. SHAMANSKI: You know, the measure was developed in 2007.

MEMBER ROSS: Right.

DR. SHAMANSKI: And it was endorsed by NQF in 2008.

MEMBER ROSS: And none of us practice the way we practiced in 2007.

DR. SHAMANSKI: Okay. We then spent the next year-and-a-half testing these measures. So now we're already up into the end of 2010. There's only so much time, first of all, to get this stuff done. So I think there is a lag in --

MEMBER ROSS: A half a decade lag?

DR. SHAMANSKI: It takes a long time.

DR. WITTE: If you think this frustrates you, you should have been on our end of the testing.

MEMBER ROSS: Well, it does frustrate us, yes.

DR. WITTE: Believe we are taking your comments to heart, because I think they are very important. When this was developed, we had the data for about 10 elements, as I recall, in the colorectal cancer report and we selected the three that we thought were most important and had the biggest gaps.

Some of those other seven elements had very small gaps. And we thought, not to add to the burden of all of the record keeping, we'd just pick the three that we felt were most important either for guiding therapy or for being absent from the report.

Now we have not gone back to get another (telephonic interference) as far as

I'm aware. But that's how we got to the three that we have.

Going forward I think the suggestion that we have a summarizing path report is an excellent suggestion. And in fact, as Dr. Hammond said, the body of pathology agrees with that and it has groups of people working on how would we get to that? There are currently no mechanisms to have a code so we could keep track of it, and we're working on that.

But as far as being able to either tell you that there's more data after what the CMS has given us, we're kind of stuck. And I guess we don't have any other data, because our data would not be anywhere near as broad as what CMS could give us as far as --

MEMBER ROSS: No, I understand everyone's well motivated; and I apologize for being stuck on this, but I've been listening all day and perhaps I just needed to get it off my chest. My psychiatrist will be happy

that I'm doing this.

So I'm disappointed that we bring experts and interested parties to the table to validate something based on no data. We're trying to make a decision on whether something is worth collecting, and a stakeholder has that information but doesn't share it with us. The only information that can validate whether to reaffirm this is what's been collected in the last three years.

DR. WITTE: I'm not in your group, but it strikes me that you have another place to communicate that. We would certainly, I think, be in favor of you doing that.

DR. ANTMAN: And if I may add, so, Dr. Ross, we certainly share the frustration that you're expressing. I think as my colleagues have said, we are not the collectors of the data and we can only work with the data that we do have.

I will note; and I apologize if it's already been noted by this committee in

this discussion, but I think it's noteworthy that the performance gap that's cited in our documentation is from information that was collected in 2010. So that is somewhat more recent data that does note that the gap, the performance gap in the -- or the percentage of reports that are missing at least one of the required elements was still 21 percent at that point. And so at least we have that as more recent information that we can provide. Although, as noted, we do not have more recent actual testing information.

CHAIRMAN LUTZ: All right. We'll let David get something off his chest, and then John's turn.

MEMBER PFISTER: No, I think that the -- you know, I think this is a very worthwhile discussion though. I think that in a lot of ways I think as Patrick implied, that, you know, while it may seem we spend disproportionate amount of time on this, on this particular measure, it's not unique to

this measure at all.

And I think that when it came up earlier about there being a venue where you look at certain process-related issues, like the exclusion of the numerator versus denominator, it's a more fundamental methodologic sort of approach, which, you know, applies across the board to multiple metrics.

Yes, I would certainly share my impression from the last meeting and this meeting that when things come up for reassessment that there's often very little that -- you could change the date on the submission form and it's basically the same submission form that was looked at the prior time. And that, you know, I think it may be worth, you know, being more explicit with the subsequent forms, not just for this, but for other measures as well that are coming up for sort of renewal. This is what's new. And it would at least leverage a little behavior to

say, well, if we don't have much out there, that's probably not a good thing, although it may be beyond your control in terms of providing what's new.

But I think that there is a certain kind of -- when things get past the first time, it's often on the presumption, well, more is coming. But by and large, I find that for a lot of the measures that come back that really more isn't coming. And it's sort of like we don't really raise the bar in our assessment of the measures proportionately. And I think that that's sort of something which I think is a general part of the process which I think is worth revisiting. It was sort of a touchstone for this particular measure, but I'm not sure this is in any way unique to this measure.

MEMBER GORE: So toward that, and this is maybe a better discussion for the kind of future directions of NQF, as they evaluate new processes for how the performance measure

process works, is there a possible process for essentially like amendments or updated modifications to measures?

MS. FRANKLIN: Yes, there's an annual update for the measures.

MEMBER GORE: Oh.

MS. FRANKLIN: As new information becomes available --

MEMBER GORE: Okay.

MS. FRANKLIN: -- the developers are able to amend their submissions.

MEMBER GORE: Okay.

MEMBER LOY: One more comment. I heard Angela say that, you know, you need to vote on this. You might want to consider voting that there's insufficient evidence to support. I think where I find myself is there's really insufficient data to even have an opinion at this point one way or the other.

So I just recommend that if there's a way to weight some of these questions in the renewal mode or the

maintenance mode -- because I for one would be very hesitant to say, no, take this away. Because at least in my view, and I think the point's already been expressed, this is table stakes at this, you know, time in 2012. If you don't document your pathology well, that's a very different expectation now than it was even five years ago.

And to the synthesis comment, you know, I heard you use the word criticism and I just wanted to pull back from that just a slight bit, because I don't think we're there yet and I think the measure developers have been contemporary in that we've kind of chosen the important things. We have measures yet to look at today that do address lymph nodes and do address KRAS. So I think those are very important.

Still, I think there's opportunity, but I also think there's opportunity for these maintenance pieces to have a stiffer requirement on some sort of

data. But again, I would be very hesitant to say, no, take it away, take the measure away because of insufficient data, because I don't know if the problems been solved yet or not.

MS. FRANKLIN: And, yes, just to answer your point, the steering committee can look at whether there's an impact, there's a high impact for this measure and whether or not there's still an opportunity for improvement, and whether there's a strong link to outcomes, or desired outcomes when making the decision as to whether you want to move the measure forward.

CHAIRMAN LUTZ: I think we had Karen and then David.

MEMBER FIELDS: So my question was process. So can we approve a measure and with a -- you had talked about we're allowed to make some suggestions or recommendations about it, or a caveat?

MS. FRANKLIN: Yes. Yes, you may.

MEMBER FIELDS: So we can do all

of that with this one vote? We can say we approve the measure, but we expect -- in one year we want the rest of these data elements in there and we need data. Is that how we do it?

MS. FRANKLIN: Yes, you would walk through the votes as usual. And if it looks like it's going to fail at the evidence level or at the importance level, you could make a vote or a decision as a committee to invoke the exception, which is that the potential benefits outweigh the potential harms. And at that time continue through the voting. And at the end we would talk about recommendations for future development. And if you had caveats as well for the developer, if the developer is able to address them in this measure, you could base your decision on those changes that could be made.

MEMBER FIELDS: So I think just for the summary issue, summary report issue in colon cancer, breast cancer is 31 years of

adjuvant therapy, or 30 years of adjuvant therapy where we've come up with what are the really key items. Colon cancer wasn't quite as far along in 2008 when they were developing that as far adjuvant therapies. In colon cancer it sounds like we need to get to a better standard of just reporting before we get to the more sophisticated summary reports because we've changed the therapy of breast cancer over the years.

So I think the caveat should be we need much better reporting across the board on preliminary data in colon cancer than even we're getting. That's all.

CHAIRMAN LUTZ: I think we're David and then back to Patrick.

MEMBER PFISTER: No, certainly I hear what you're saying about the difference between colon and breast. But, you know, in substance if you look at 0391 versus 0392, I'm not sure that that additional data explains the magnitude of the difference in comments

about one measure that was basically already passed here versus something which would generate 25 minutes of discussion. And I think a lot of the issues are basically identical for both measures. And that's why I say, I think it's a more fundamental issue. So some of the issues about, well, recommendations kind of go back to developers and assess its own merit. Those are, I think, equally applicable to the breast measure which we just passed. It's just that this discussion is occurring 30 minutes later.

MEMBER ROSS: Angela, I have a question to understand. So let's say that we vote down. I mean, so first of all, we'd have to question the judgment. Who's going to say that not doing appropriate staging on colon cancer is a good thing? It would not make good press for the Cancer Steering Committee to vote against it, right? But let's say we did it. When would this group of sponsors then have the chance to bring the new,

improved version forward? Is it a year from now?

MS. FRANKLIN: It would be during the next time that we have a project related to cancer.

MEMBER ROSS: So it's almost an impossibility to correct any of these --

MS. FRANKLIN: Therapy time, right.

MEMBER ROSS: -- in real time, right?

MS. FRANKLIN: Yes.

MEMBER ROSS: That's disappointing.

MS. FRANKLIN: Mark?

DR. ANTMAN: Thanks, Angela. Just a comment on the PCPI process, Dr. Ross, just to clarify our timing in working with our colleagues at CAP in updating these measures. Typically we convene our own work groups, our own panels of experts to consider the measure, and in this case, to update a set of measures.

Angela referred earlier to the annual updates that are available for all currently endorsed measures. We're able to use those annual updates only for situations where there has been a coding change to an element that's in a numerator or denominator of a measure where we can make a somewhat insignificant non-substantive change to a measure.

But if there's a substantive change, such as what's been discussed here, retiring, if you will, one element of a measure and replacing it with others, or perhaps adding new ones: that would require a very substantive discussion of our panel of experts. And so that's by way of saying that unfortunately that's not something that we could do in a very short time frame. It would require our reconvening the group. But it might be possible by the next time that NQF convenes the cancer group, depending on how much time passes at that point.

Typically our measure development

processes take in the neighborhood of a year or a little more or less, partly depending on the testing process involved, but it's not something we can turn around quickly because we need a panel of experts to approve it.

MS. CHRISTENSEN: If I can just, sorry, piggyback on that, these measures are a bit caught in the gap of our colleagues at NQF revising their process, which we have enjoyed the new process. But the timing on these hasn't come out quite the way that we maybe would have liked, and that's because when they were first endorsed we then went and did a testing project with our colleagues. And as Mark said, those can take somewhere between six months and a year. You guys I'm sure have all gone through the IRB process, and it's just harder to get done faster than that.

So once we had that information -- that does go back to the work group when we do testing projects, things that we find that

need to be clarified and things that perhaps could be updated. We do take those back to our measure work groups. Unfortunately, in this case, if we had made changes to the measures, we would not have been able to submit them because they would not have been tested again in time for this policy. So it just is kind of a timing issue to figure out what are the best measures you can submit at any time. With the new process, we should be able to adjust our timing to fit that.

CHAIRMAN LUTZ: Do Jennifer then David.

MEMBER MALIN: Maybe it is just because I'm steeped in oncology that I feel like oncology changes more rapidly than other fields, but it does seem like, you know, things change pretty -- I mean, guidelines get updated several times a year. And so I wonder if maybe there's something with the NQF process where instead of the committee reviewing just the final set of measures,

saying up or down on this measure, if there couldn't be six months ahead of time someone who reviews how well do these measures fit the context of what's happening in breast cancer today.

MS. FRANKLIN: That's part of what our new CDP two-stage process that we're piloting will do. But that hasn't come on line now. And unfortunately, I think this project kind of falls in the gap. We don't anticipate seeing another cancer project for at least a year, or more than a year.

MEMBER MALIN: Okay. Well, that's good. Maybe that will help some of --

MS. FRANKLIN: But that is contemplated in the new process.

CHAIRMAN LUTZ: All right. Anybody else? Anybody on the phone have any additions?

MEMBER RICCIARDI: Yes, this is Rocco Ricciardi. Just a couple things. One I'd say that I still get path reports today

that don't include this information, so I think it's still valuable today. I know that's anecdotal, but I think, you know, I can comment that my colleagues and I do still see this. And two, it looks like we have a metric or a measure that looks at measuring the number of lymph nodes, which I believe is very important. Thank you.

MEMBER ALVARNAS: This is Joe Alvarnas. I agree. I think that we do want more perfect measures, but I think given that there's a performance gap, sadly, with even this level of measure, I think we should just vote upon that gap rather than become paralyzed because the measures may in fact not be perfect.

At the same time, I think we have to be able to plan for measures that are brought forward in a more timely fashion to maintain the currency. Because you're right, in oncology and hematology the state of the art evolves so rapidly that five-year cycles

may be way too long for these things to maintain their complete relevance.

CHAIRMAN LUTZ: Very good. Thank you. If someone could remind me where we are on the voting.

PARTICIPANT: 1b.

CHAIRMAN LUTZ: 1b?

MS. KHAN: So voting on performance gap, 1b.

So we have seven for high, five for moderate, zero for low and one insufficient.

And 1c, the evidence. Yes, no or insufficient.

So 12 yes, and 2 insufficient evidence.

Voting on reliability.

CHAIRMAN LUTZ: Does anyone have anything else that they want to say about reliability in this? John?

MEMBER GORE: So there was, I think, fairly robust evidence presented of the

reliability of ascertainment of this measure. The working group had no concerns about reliability. And we go to validity now, too, correct?

MS. KHAN: Yes.

MEMBER GORE: In terms of validity, this was one of the measures where there was an expert panel that kind of decreed the importance to report. And there was pretty uniform consensus about the importance of the measure and the validity of the measure.

MS. KHAN: So we're going to go ahead and vote on 2a, reliability.

So we have five high, and nine moderate, zero low and zero insufficient.

Voting on 2b, validity.

So we have five high, seven moderate, two low and zero insufficient.

And did you want to have a discussion on usability?

CHAIRMAN LUTZ: Say anything else

on usability, John?

MEMBER GORE: I don't think there's much to say about usability. And it's a little bit pursuant to some of our previous conversation, but the working group didn't have any concerns about the usability of the measure. The accurate pathology report definitely can be used to evaluate pathology labs, institutions, whatever.

MS. KHAN: So voting on usability.

Can we have everyone press their button one more time?

Still missing one vote. If you could push your votes again.

So you have four for high, nine moderate, zero low, and one insufficient.

And feasibility?

CHAIRMAN LUTZ: Say anything on feasibility?

MEMBER GORE: The elements are all easily abstracted on. For example, electronic health record and our standard parts of a

synoptic path report.

MS. KHAN: And voting on feasibility.

So you have nine high, five moderate, zero low and insufficient information.

And overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? Yes or no.

Thirteen yes and one no, so the measure will pass.

CHAIRMAN LUTZ: So is there anything else, any recommendations beyond what we've said for our developers who want to give thoughts, suggestions on this one? Karen?

MEMBER FIELDS: So I guess we wanted to make a recommendation that they try to add some other pathologic elements to the list of elements that they're measuring, if that's feasible, although it sounded like for next year's measure that's not feasible because they could only replace new elements.

But perhaps histologic grade needs to be replaced with something a little more contemporary, like margins.

And then number two, we also would like for the next year's data -- the caveat would be we need to see the data from the period up to that time.

MS. TIGHE: And just for my notes, is this recommendation only for 0392, or for 0391 also?

MEMBER FIELDS: I think 0391 had some different issues. I think that we thought that the requested pathologic elements were broader. These were just three elements that we didn't think were sufficient, unless someone disagrees with me.

MEMBER PFISTER: I mean, I think it was equally applicable to both measures. I mean, I think the pathologic elements, looking at both measures, are identical, right? I mean, it's T, M, grade, you know?

I think you're absolutely right

that, you know, there's a longer line of -- to the extent you're hoping to have some correlates without them here just because there's more adjuvant data. There's going to be more correlation with outcome, but I think that, as I think has been implied by other discussions, there's certainly other factors, albeit different for the diseases which are kind of raising the bar in terms of what oncologists look at when they're trying to make these management decisions now. I think making some of the old paradigms, probably just that, old paradigms. And some of the new paradigms which are now being actively used.

And so I think, you know, some of the updating issues, some of the relevance to, of let's say the applicability of something like grade, which is a historic-sort of cultural thing that often drives decisions. And I'm not saying there aren't particular circumstances where it sort of does factor into what you do, but is that on the same

footing as some of the other markers which are now like very heavily vetted? You know, that's a larger discussion. But I would say some of the caveats, I think, are very similar for both 0931 and 0392.

MEMBER FIELDS: So I'll respond to that. So in breast, though, we still know that the major prognostic indicators are TN and then ER/PR status and HER2 status, and I think some of the previous measures addressed some of those issues. Whereas in colon, I think we're just now getting to understanding KRAS and a little bit more information about nodal status.

So I do think they're a little bit different. But I agree, all of them need to be up to date, all of them need to reflect modern therapies, and the fact that we use all these data now for treatment decisions where we used them perhaps less at the time in colon cancer from that era.

I have a question, though. Are we

allowed to go back and make a recommendation without the group that voted on the breast thing? Or we can just make a caveat on both of those right now and we vote that and make that as a recommendation?

MS. FRANKLIN: We can make it a recommendation for 0391. Add that as a recommendation for 0391. We don't have to vote again. I don't think --

MEMBER FIELDS: Okay. So we must make --

MS. FRANKLIN: Right.

MEMBER FIELDS: These aren't voting things? These are recommendations?

MS. FRANKLIN: Right. They're not voting elements.

MEMBER FIELDS: Okay. That's all I needed to understand.

DR. ANTMAN: If I may just add, I do want to clarify that -- and I'll stand corrected if my colleagues at CAP disagree, but when we updated the measures, I don't

think any of us said that we would have to simply replace one element of the measure with another. Now we heard the recommendation to replace histologic grade with the margins.

MEMBER FIELDS: No, there was a statement; and maybe I took it out of context, that said we just try to replace measures rather than we add new measures.

DR. ANTMAN: Ah.

MEMBER FIELDS: Because it sounded like a huge process to add new elements to the measures.

DR. ANTMAN: I see.

MEMBER FIELDS: If you can add new elements to the measures, I think you're hearing our group calling for that.

DR. ANTMAN: Right. Okay. So all I wanted to clarify was that we're happy to take whatever recommendations you have on additional elements that you think should be in this measure, and that can all be part of the work groups in their deliberations.

DR. SPEIGHTS: And speaking from this side of the table, I think we can say that we have heard your concerns and we'll certainly work with the ongoing CAP efforts for integrated and comprehensive summary reports, and we'll certainly work on these. Thank you.

MEMBER MALIN: I know we've kind of beat this to death, but I wanted to sort of just in the spirit of thinking about the measurement process -- the point of these kinds of measures, especially -- these are really about communication between the team members. And really what you're trying to encourage is that pathologists and members of each institution have a process in place whereby they're making sure that they're documenting and communicating what's important to the person who's receiving the information.

So, and especially as we see new genomic tests, it would, you know, make these measures kind of useless if they're not

reflecting is this a pathologist who's keeping up and making sure he's providing the information that the clinicians need to make treatment decisions. It's not so much, I think, you know, about what the elements that are included in the measure are just one way to capture that.

CHAIRMAN LUTZ: Good. Our competing issues I think are folks from ActiveHealth with submission 0623 said they only have until 4:00. But then again --

PARTICIPANT: Do we have members from the ActiveHealth Team on the line?

(No response.)

PARTICIPANT: Arnika, could you please check to see if there's anyone from Active Health whose line may need to be opened?

OPERATOR: If so, could you please press star one?

The line is opened.

DR. CHIN: Hi, this is

ActiveHealth. Yes?

CHAIRMAN LUTZ: We're curious. Did someone from you guys say that we need to go over 0623 with some time frame in mind, like before 4:00 p.m., or how did that I

DR. CHIN: Yes, that's okay. No, we're fine.

CHAIRMAN LUTZ: Okay. If you don't mind then, we'll take a short break so everyone can kind of walk around their chair once.

DR. CHIN: Okay.

CHAIRMAN LUTZ: Thanks.

(Whereupon, at 3:45 p.m. off the record until 4:02 p.m.)

CHAIRMAN LUTZ: Okay. I think the request was made that the folks from ActiveHealth identify themselves by name. That was one of the first requests. If we could, please?

DR. CHIN: Sure, this is Dr. Lindee Chen from ActiveHealth Management, and

we have --

DR. PALACKDHARRY: This is Dr. Palackdharry, ActiveHealth.

DR. MENTHA: This is Laneesh Mentha. I'm the pharmacist. ActiveHealth.

CHAIRMAN LUTZ: Good. We appreciate that. And I think that if we're okay, I think we can go ahead and -- if you don't mind, if you can just go ahead and give us some background and framework for the submission. And then we'll go from there.

DR. CHIN: Sure.

MS. TIGHE: Sorry, just one quick point. For those in the room I had mentioned it. You have a copy of 0623 on the table in front of you. They have made some updates to it as a result of the work group call, and so we'd just ask them to point out those changes to you. Okay. Go ahead, Lindee.

DR. CHIN: Okay. Sure. So our measure is titled "The History of Breast Cancer - Cancer Surveillance." And we're

looking at the percentage of women with a history of breast cancer treated with curative intent who had breast cancer surveillance for a local regional recurrence annually. We updated the description of the measures to be more clear. I think there was some confusion about what types of cancers we were looking for exactly last time. So we updated the measure description, the numerator description and the denominator description. And we also had changed the numerator time window based on the preliminary work group suggestion as well. The other pieces that we had updated are the reliability and validity testing areas. We went back and looked at our data and did the statistical analysis that I think the group was asking for. I think we misunderstood the wording of the question, so we went back to our data and tried to give the statistics I think that the committee was looking for.

DR. PALACKDHARRY: This is Carol

Palackdharry. Did you want us to summarize how we updated it, or just that we updated it?

MS. BYRON: I would appreciate you describing how you updated it.

DR. CHIN: Sure. That would be better. Great. Okay. So in terms of the numerator statement, we're looking for women with a history of breast cancer treated with curative intent who had surveillance for breast local or regional recurrence annually. We updated the time window just a few months. It was 12 months before, but after the previous discussion with the preliminary work group it went back to 15 months.

We had 15 months in the past on our previous endorsement. We had moved it to 15 months to align with sort of the annual recommendation, but then we went back to 15 months because of the discussion around that women aren't going to get it within the 12-month window because of insurance reasons, that we need to give them a little bit more

time to get testing completed. So that was one of the changes that we had made.

The other piece that we wanted to emphasize in our descriptions is that we're looking for non-metastatic invasive breast cancer. So we just put that clarification in the description.

MS. BYRON: I just want to bring in, there was some question during the preliminary work group meeting about what the rules were pertaining to DCIS. And we wanted to make it clear that DCIS is -- all in situ breast cancers are excluded from this measure. It's invasive cancer only.

DR. CHIN: Okay. So we updated most of the descriptions to reflect that. And the other piece that we did in terms of the validity and reliability testing, like I said earlier, we added sort of the numbers in our test sample and their our statistics around it. So we had added our signal to noise ratio. We also added the other sort of

discussion around our sample size for our validity testing. And that's sort of more details around it that I think the committee was looking for.

CHAIRMAN LUTZ: All right. Thank you. I think if Heidi Donovan's on the line, I think she was going to give us our first overview of this. Are you there, Heidi?

MEMBER DONOVAN: I am here, yes.

CHAIRMAN LUTZ: Great.

MEMBER DONOVAN: Okay. So just everyone knows, I was not on the phone call of the small group discussion, so I hope others on the phone call will weigh in.

I guess we'll just start with the importance to measure. I think there were two discussions. One of them has been addressed, the question of whether 12 months was an appropriate timeline given some insurance restrictions. I think it's great that they've extended it to 15 months.

I think in terms of the importance

to measure, the other concern was that while everybody is very much in favor of annual screening and that meets -- that is appropriate and consistent with NCCN guidelines, there is not adequate evidence out there that screening does improve survival outcomes. And so I think that was one of the issues that came up. As I said, that's countered by the reality that we do find the early cancers in a group of patients who at high risk for recurrence.

I think that I'll stop there. Let's see, they've addressed the issue around DCIS isn't excluded. I think there was some question about whether there needed to be an age limit for annual surveillance, and they have provided some rationale to not put in an age limit other than if women have short left expectancies. That's somewhat unclear that we can talk about that as well. So I'll stop there and let other people weigh in on this.

CHAIRMAN LUTZ: Okay. So I guess

we're looking for any other comments about importance. So I will say -- I hate to speak for him, but Dr. Marks wanted to strongly state that he doesn't think that this changes survival and was therefore not for it. Larry I think is not on the line, but asked us to at least mention that it was still his strong belief even after the changes.

But starting there, does anyone want to argue differently?

MEMBER FIELDS: So I had some questions for the developer. I guess your clarification about "invasive" probably needs to apply then to the description of the measure, the numerator statement and the denominator statement, because only the denominator statement says that it's invasive. And so the question is following the DCIS patients.

And then I had another question. I understand the intent, but I got confused about how you were trying to describe

reconstruction and whether or not those patients were candidates for follow up. It sometimes implied that they might have needed surveillance and they wouldn't -- I mean, I would assume they would need bilateral -- I mean, they met the criteria for bilateral mastectomies and therefore they weren't eligible. But there was some implication that you might follow them with MRIs or something like that.

And then my last question was there was sort of an interchange between screening, or follow-up mammograms and breast MRIs, and I didn't think that there was much literature or data to support MRIs as a follow-up or surveillance study in the patients.

So that was three big questions, but I'll stop and let you comment.

DR. PALACKDHARRY: Sure. This is Carol Palackdharry. So let me take those one at a time.

In terms of the updated terminology, I just wanted to make it clear that in situ carcinomas were never included in this measure. It was always only invasive breast cancers. And so the coding that we use in our elements, we only use codes for invasive breast cancers. So does that answer that the first question?

MEMBER FIELDS: It does. It's just you probably want to go through the document and make it consistent, because sometimes you say invasive and sometimes you say breast cancer.

DR. PALACKDHARRY: Okay.

MEMBER FIELDS: And that's the difference.

DR. PALACKDHARRY: That's right. If you just look at the very first page, you have women with a history of breast cancer treated with curative intent. And the numerator statement says you have breast cancer treated with curative intent. And then

in the denominator it's a history of non-metastatic invasive breast cancer.

DR. CHIN: Yes, you know, that's a good point. We'll make that more clear.

DR. PALACKDHARRY: Yes.

DR. CHIN: So the second one, yes. And the second one, let me just clarify it by saying that in the revision that we submitted to you guys, we removed everything about reconstruction. And previously when we first -- meaning that we removed all women who have had bilateral mastectomy regardless of any kind of reconstruction from the denominator.

MEMBER DONOVAN: I mean, how do you measure previous local recurrences? Are they included in this or excluded?

DR. PALACKDHARRY: With previous local recurrence?

MEMBER DONOVAN: Right, so not metastatic.

DR. PALACKDHARRY: We do not exclude them if they've had a previous local

recurrence unless that recurrence led to a completion mastectomy which then gave them bilateral mastectomies. If you add, you know, two unilateral mastectomies together. But if they still have breast tissue left, if it wasn't coded as a full mastectomy, either bilateral at one time or two unilaterals, then they would still be included. Did that make sense to you?

MEMBER DONOVAN: Yes, just wanted to clarify.

DR. PALACKDHARRY: Sure. Thank you. And the last thing was we are not suggesting that women receive MRI, but we are counting MRI as a completion since some women are clearly recommended to get MRIs on the basis of dense breast tissue or radiation changes, or whatever. You know, and there are organizations that do recommend MRI for high-risk women in combination with mammography. But we're not recommending that. We're just taking that as a completion.

CHAIRMAN LUTZ: Okay. We're going to Bryan, Jennifer, Robert and John.

MEMBER LOY: Yes, and I think it's just important for me to disclose that I do have a working relationship with ActiveHealth Management, that I disclosed last time.

But I wanted to ask the question. I heard Dr. Lutz say that Dr. Marks said that there was no impact on survival. Is that correct?

CHAIRMAN LUTZ: He emailed us a request to make that statement as a point that he wanted to bring forth.

MEMBER LOY: And I'd just like to hear the point of view of the other committee members on that particular point, and if ActiveHealth had a response to Dr. Marks' concern.

DR. PALACKDHARRY: This is Carol Palackdharry. I actually do have a response to that, because at least -- well, I'm an oncologist, and so I would just say that when

I look at the data; not me, when we all look at the data which is looking at the survival of women with invasive breast cancers who had breast-conserving surgery with radiation therapy versus mastectomy, I think it's long been realized that although the survival 10 and 20 years out is the same, the incidence of relapse-free survival is not the same. And the reason the overall survival becomes the same is because if you've detected a recurrence in the conserved breast, you can salvage that breast by mastectomy. That's the reason the overall survival is the same, is because they get salvaged with early detection. So I guess I would disagree with Dr. Marks' statement.

CHAIRMAN LUTZ: Please do.

MEMBER MALIN: There is no randomized trial data to support mammography improving survival in women who've already had breast cancer. And there have been several randomized trials that have looked at the

impact of intensive monitoring using imaging and laboratory data to see if there's an improved outcome in terms of survival. And those several studies are now probably 10-plus years old, but both of them were negative and showed no improvement in survival.

And so, you know, most of the time, you know, I think the rationale for doing it is that, you know, presumably people are at risk for contralateral disease and, you know, it's a low-risk procedure, so why not do it? But there's no evidence that it improves outcomes.

And at this point, you know, with modern radiation therapy techniques and hormonal therapy, local recurrence risks are in the low single digits. So something like two to three percent of women will have a local recurrence.

DR. PALACKDHARRY: But I think it's also important to point out that false-positive rate is higher --

MEMBER MALIN: Right.

DR. PALACKDHARRY: -- in the patient population overall.

MEMBER MALIN: So one of my concerns though is just how broad the denominator population is for this measure. So, I mean, it seems like you've excluded people who are at death's door, but you know, breast cancer is a very, very common disease and there are a lot of, you know, 80-year-olds and 90-year-olds who are unlikely to benefit at that point from having any breast cancer identified early. We certainly can identify it early, but whether it will, you know, decrease their morbidity or mortality at that point, given other things going on, you know, is questionable. And certainly most screening guidelines tend to put an upper limit on the age at which you would actively screen. And I guess I'd be interested in your thoughts as to why this population should be any different.

DR. PALACKDHARRY: This is Carol Palackdharry. Now none of the guidelines that I'm aware of actually have an upper age on the surveillance guidelines. It would be -

MEMBER MALIN: Well, NCCN guidelines specifically say that all of their recommendations should not be applied to anyone over age 70 because there's no data on that. Or there could be, but there's no absolute recommendation. So they have a general caveat across the whole guideline.

DR. PALACKDHARRY: Well, you know, we can take a look at that again. We'd be happy to put in an upper age limit of 70, if that's what the data supports.

MEMBER MILLER: So I also have some of the same concerns, but I guess maybe just to approach it from a different way and say that I think there are multiple data sets that show that if you're just talking about the breast conservation population, which is at least theoretically the population of

patients that you might expect to have the greatest likelihood of salvaging; however you want to define that, even that population of patients -- so patients that have had lumpectomy and radiation who have an ipsilateral breast tumor recurrence, they have a poor prognosis irrespective of what happens after that occurs. The NASBP has shown that in both node-positive patients -- and this is with modern systemic therapy. NASBP showed it even in node-negative patients that the chance of distance recurrence and death is very high.

So I guess I'm concerned that in the section -- it's 1c.1, which is the relationship between process and outcome -- I mean, again, we understand this is a process measure, but it has to speak to an outcome that's reasonable. The last sentence is simply factually incorrect. "Women who have had breast conservation have a higher chance of recurring within the remaining ipsilateral breast, but early detection allows for salvage

mastectomy and thus an equivalent overall survival."

I'm sorry, that statement is simply not true. There are no data showing that. And I guess my whole concern is this whole measure is built on that assumption that you can identify something early and fix it. So I'm really troubled by the scientific assumptions based on this.

DR. CHIN: I could perhaps point to a couple of publications, if you wouldn't mind. I'm looking at one right now from Breast Cancer Research Treatment in 2010. I'm just going to read some from the abstract. They followed 17,286 women for five years. Between 1996 and 2006 these women had a combination, some were DCIS, or they could have had early-stage 1 and 2 breast cancer. And what they found was that four percent had a second breast cancer event. There were 314 recurrences in that and 344 second breast primaries; I am assuming in the other breast,

there.

They state here when they went and identified that about a third of the recurrences, 37.6 percent, and the second primaries were not screen-detected, so two-thirds were screen-detected in there.

MEMBER MILLER: I'm sorry, and your point regarding survival is? What are you getting at?

DR. CHIN: Actually, yes, I thought that there was a survival statement in that one.

MEMBER MILLER: Okay.

MEMBER FIELDS: So the NCCN guidelines suggest mammography every 12 months.

DR. CHIN: They do.

MEMBER FIELDS: And at 6 to 12 months post-irradiation for the treated breast. And it's true that in an academic center where they pay attention to margins of local recurrence rates in the two to three-

percent range. In the general patient population the local regional recurrence rate is still in the range of 10 to 15 percent over a lifetime. And completion mastectomies then make long-term survival the same, whether you had a mastectomy or a lumpectomy, if you look at the long-term data from some of the early studies. So NCCN's still recommending annual mammograms in this patient population.

MEMBER MILLER: So certainly no dispute about that. I think I'm not sure you can prove cause and effect by those two statements. So I mean, you know, everything we've known since, whenever, the '70s, that lumpectomy in -- or '80s maybe, lumpectomy and radiation associated with, you know, equivalent survival to mastectomy.

I guess my concern is a justification for a quality measure in 2012, I just don't think you can use those data to justify that the surveillance act is what is going to make that difference. And I think

that's what I'm concerned, that the authors, the developers have put this in this abstract form without a reference as a matter of fact. And I'm just saying I don't think that's correct information, and I think this underpins their whole reason to put this measure forth.

So I'm not disputing at all that these women should have surveillance. Yes, it's just what's the outcome you're expecting from that? Are they going to live longer? Have the same outcomes as someone that never had an ipsilateral breast tumor recurrence? I don't think you can say that.

MEMBER FIELDS: Right. And then we still know that the risk of a new breast cancer in the opposite breast remains the same as it was in the first breast, unless you have a -- well, you become a higher-risk patient then. So we still recommend annual surveillance in that patient population.

So what you're disputing mainly is

the rationale and the literature support for the follow up rather then the need for the follow up?

MEMBER MILLER: I'm not disputing the need for the follow up -

MEMBER FIELDS: Right.

MEMBER MILLER: -- because it's consistent with guidelines.

MEMBER FIELDS: Yes, okay.

MEMBER MILLER: I'm saying I understand this is a process measure, but every process measure implies some type of outcome. I'm not sure I understand what --

MEMBER FIELDS: Okay.

MEMBER MILLER: -- outcome -- the outcome that is purported in this abstract document is -- I question its scientific validity.

MEMBER FIELDS: Okay. Okay. That was my question.

MEMBER MALIN: I think the other thing, too, is that there's already a quality

measure out there that all women of a certain age should get screening mammography, right? So the question is what's the added value of having an additional one specifically for breast cancer survivors that maybe focuses on a slightly different interval and has a more conservative interval. Is that, you know, really meaningful and add to kind of the quality of reporting that's out there?

MEMBER DONOVAN: I guess then the other measure that's just screening there, there is no measure for breast cancer survivors then. That excludes people who've had a diagnosis.

CHAIRMAN LUTZ: I'd say we'd have to look, but I'm not sure it does, because you're screening for a new cancer in the same breast and for a new cancer in the contralateral breast. So it's still a screening situation, I believe. I mean, we can double-check the wording, but I don't think someone's excluded from screening just

because they've already had active treatment.

DR. PALACKDHARRY: Yes, they are excluded.

MEMBER MALIN: Maybe someone from NQF staff could pull up the measure for us?

MEMBER FIELDS: In the denominator exclusion on the other measure we're going to evaluate next --

PARTICIPANT: Right.

MEMBER FIELDS: -- its says who had a bilateral mastectomy or for whom there is evidence of two unilateral mastectomies. It doesn't say that they had a diagnosis of breast cancer.

PARTICIPANT: Mastectomy--

MEMBER FIELDS: Yes, it's like you can't do it, right. Right.

CHAIRMAN LUTZ: David, I'm sorry, you've been waiting patiently. Did you have something?

MEMBER PFISTER: A couple of things just to reiterate some of the points

that have been made. I mean, on the first they talk about that the goal here has to do with local regional recurrence detection. But there doesn't seem to be a real, like, specification of a time frame. If the emphasis is on local regional failure detection, one would expect that most of those are going to be early events and that after the first five years it's probably going to mainly second primaries that you're going to pick up on surveillance.

I think that similarly the issue about what the impact of the imaging of a post-mastectomy breasts is on survival as an end point, I think, is not established. So, you know, I think that that's an assumption. I think as far as how this population fits into the screen recommendations, you know, I don't know that off the top of my head. I think it's worth asking though.

CHAIRMAN LUTZ: John, did you have something?

MEMBER GORE: This is just more of a question for the developer. And this isn't pertinent to the question of the scientific importance of therapy for local regional recurrence, but it's a question about the unit of measurement. And so who is expected to be measured with this metric? Because, you know, some of these women may be in the survivorship phase of their breast cancer and it may be unclear who is being assigned this quality metric? Who is being evaluated with this metric?

DR. CHIN: So it's those women that we find with a history of breast cancer, invasive breast cancer with surgical or radiation treatment in the year prior to the measurement year. Because if there were --

PARTICIPANT: I need you to answer the --

MEMBER GORE: What provider is being assessed with this metric?

DR. CHIN: It's the provider who

is coding for the breast cancer for the patient or whoever is caring for the patient. So we have an algorithm for which we try to identify providers who are coding that we seen claims for the breast cancer diagnoses. And then by default then it would probably go to the primary care provider if we don't need those codes. Those are sort of the algorithm that we go through.

MEMBER LOY: So outside of your algorithm would other entities be able to reliably attribute back to a provider in a similar manner, or is the measurability of this measure dependent upon your proprietary --?

DR. CHIN: No. No, I mean, it's typically whoever you're finding that is coding for or treating the patient, or has treated the patient for the breast cancer, or actively caring for the patient. We don't say that you have to attribute this measure to anyone in particular.

MEMBER LOY: Okay. And --

DR. PALACKDHARRY: So that way if the person, if the woman were say transferred back to her primary care physician or her gynecologist after the acute phase of treatment, and if the oncologist or the radiation oncologist isn't coding at that point, so their follow up, that's the primary care-

MEMBER LOY: So what if they're seeing both? What if they're seeing a primary care doctor and a medical oncologist in follow up? Who gets the attribution?

MS. FRANKLIN: But so I just wanted to clarify that the level of analysis is current specified as the population level, the national population level?

MS. TIGHE: Yes, so correct me if I'm wrong, but ActiveHealth actually uses this measure for their clients and they use it at all different levels. The measure in front of you today is only specified for the population

level. So they're talking about their uses for it, but the NQF endorsed measure would only be used at a population level.

MEMBER LOY: Okay. And then the other question I might have being a non-medical oncologist, I heard the developer say that there was some value in relapse-free survival. Could you help us to understand if that's clinically meaningful, or in what way?

DR. CHIN: I guess relapse-free survival per se, I don't personally think is clinically meaningful. I think we could probably have another, you know, 10-hour discussion on what the literature says about that. But for this reason it's that women who have breast conservation do have a higher risk of relapsing within the breast tissue that's remaining. And if that is detected, then that breast can be removed by salvage mastectomy and that woman then is expected to have the same survival, overall survival, as a woman who was treated with mastectomy.

MEMBER MALIN: I think the problem with that sort of logic though is so we know that people who have breast-conserving surgery and radiation have the same survival as mastectomy. There is no data. Presumably I think most of it is because it's the distant metastases that kill you, not the local recurrences. And so even though we certainly don't let local recurrences just lie there; we treat them, and there's some retrospective data that suggests giving those people additional chemotherapy may improve their outcomes, it's really very speculative. And there's certainly no evidence of a process outcome link there.

Can we ask about or are we at the point where we can ask questions about the kind of reliability and validity of the measure, or are we still on importance?

CHAIRMAN LUTZ: I think we're still on importance.

MEMBER MALIN: Still on

importance?

CHAIRMAN LUTZ: John, did you have something?

MEMBER GORE: Just speaking to the structure process outcome link. And my question is just if this measure is relevant to a population, then how is it used for quality improvement?

CHAIRMAN LUTZ: Bryan and Jennifer, either one of you have anything? You're fine.

MEMBER GORE: Do the developers have a response?

DR. CHIN: Sorry. Can you repeat the question again?

MEMBER GORE: So if the unit is the population, if it's used to evaluate a population of patients, how is that used for quality improvement?

DR. CHIN: Well, you know, our clients or any sort of person using this measure would monitor their population of

people and how many of them are doing the surveillance on an annual basis. So I guess if they're performing well, they would either look at sort of what they're doing in terms of recommendations to patients to improve surveillance. Is that what you're asking?

MEMBER GORE: Well, with many of the metrics that we look at the unit analysis is such that you can discriminate quality among or between providers. And so if the unit of analysis is the entire population, then all you know about is whether your whole population is doing well or doing badly. And I wonder about the opportunity for quality improvement when it's looking at the whole without trying to drill down any deeper.

DR. CHIN: Well, since you clarify, I think we didn't say that this measure couldn't be used at those different levels. We do not do the level of analysis, the statistical analysis at those different levels for our measure. I think that we were

trying to answer the question on the form that says that if you're going to use this measure at those different levels you need to do the different types of analysis at those levels and the statistical analysis. And we do not have the time to do that level of analysis per provider and such. So that's why we said we went for the population endorsement. But it's not that this measure isn't being used by some of our clients to look at their providers and how they're doing on these measures.

MS. TIGHE: So ActiveHealth is using it at all the different levels, but they've only provided reliability and validity information for the population level. So that's the only level that we can evaluate it at.

CHAIRMAN LUTZ: All right. Is there anything else before we vote on importance?

(No response.)

MS. KHAN: So voting on 1a,

impact.

MS. TIGHE: Heidi or Rocco, if you're still on the line, you want me to send your votes?

MS. KHAN: We have zero for high, four for moderate, five for low and three insufficient evidence. So we are done.

MS. BOSSLEY: Let's do all of importance. Let's do the gap and the evidence, too. I think that would be helpful.

MS. KHAN: Okay. Moving onto performance gap.

We have one high, four moderate, one low and seven insufficient evidence.

And moving onto 1c, evidence.

We have zero for yes, seven for no and six insufficient evidence. So the measure will not pass.

MS. BOSSLEY: So this didn't pass importance. No need to move forward because this must pass.

CHAIRMAN LUTZ: All right. Let's

see. The next one is No. 0031, breast cancer screening. I think NCQA is going to give us the frame work and then Nicole is going to give us the details of the discussion.

MS. BOSSLEY: They are making their way to the table.

MS. BYRON: Hi, I'm Sepheen from NCQA. Mary Barton. The breast cancer screening measure is a HEDIS health plan-level measure. It's a longstanding measure in the HEDIS health plan measure set. And it looks at biennial, so that's once every two years, mammograms in women ages 40 to 69. And it's applicable to commercial, Medicaid and Medicare health plans.

CHAIRMAN LUTZ: Okay. Nicole?

MEMBER TAPAY: Yes, I would just add in terms of impact, there's potential high impact because of the benefits of early detection in terms of survival. There's actually significant room for improvement, even in the white population, and the African-

American population. We're only at 68 percent, and it's even lower for other ethnic and racial minorities. The group found it to be a reliable measure with a high degree of usability. As was stated, it's being used right now for HEDIS.

I think a lot of the controversy was around the validity. As many of you know, there was a U.S. Preventative Services Task Force recommendation to only begin it at age 50. And so the actual recommendation of the group was only three to two to recommend it to go forward. And I think largely because of that it wasn't clear from the explanations why there was a divergence. While they cited a number of other groups, ACOG, ACS, that concurred with this recommendation there wasn't really a clear rationale for keeping it at age 40.

CHAIRMAN LUTZ: All right. So is there -- please.

MS. BYRON: So NCQA is aware of

the differences between some of the national guidelines that are out there. And because of that, we are starting a reevaluation of the measure. And the difficult position that we find ourselves in is that there are national guidelines that are recommending different things. And for the task force, the recommendation for ages 40 to 49, that screening should be an individual decision based on shared decision making and other factors like that.

And so we did not feel that we could immediately change the measure. And what we anticipate is that we will be working with an advisory panel to discuss how we might address these issues. One possibility is that we might stratify the measure by different age groups so that we would be looking at 40 to 49 and, you know, 50 and up, something like that, so that you could say what the rate for a health plan would be in these different age stratifications. And that might be one way

that we would have a measure that doesn't come into conflict with guidelines but is still able to produce some meaningful information for quality improvement.

MEMBER MILLER: So I have more of a procedural question, I guess. For example, if, say, the major objection that one of us were to have was strictly about the age issue; I'm just trying to think through, does that apply more to the importance to measure part of it, or is that really a reliability/validity question? I mean, I could argue both sides, but so is that --

CHAIRMAN LUTZ: I think it's hard to separate it.

MEMBER MILLER: -- overthinking it?

CHAIRMAN LUTZ: I had the same question because I keep arguing with myself there. I don't know what you guys think, but it seems like it's a little bit of both those things.

MS. BOSSLEY: I think the big thing will be, as we walk through this, to be clear on each criteria what the concerns are, because you could really raise them in both places. It's based on specifications as well as concerns with the measure as specified doesn't quite match the evidence that is important, if that's what you were thinking. Just have to be clear, yes.

CHAIRMAN LUTZ: Well, and then if I could, if we can say it in importance; and maybe this is the part I'm not supposed to say, but it concerns me that we might have something that's still being actively discussed and intelligently argued between respected bodies. And if we solidify this in a quality measure which can then be used as a payment issue, what we say can pull us onto one side or the other. And I'm not saying it's wrong. I'm not saying it's not a bad measure. It's just a real rough time given all of those disparities. I mean, you know,

we could probably intelligently argue about the 40 to 50 age group from either side for a long time.

Nicole?

MEMBER TAPAY: This would just be another clarifying question for NCQA. How long does your reevaluation process take?

MS. BYRON: For this measure we'd be reevaluating this summer. We plan to convene our advisory panel in July. The issue is we offer all of our measures for public comment. So because this is HEDIS plan measure, we align it with our HEDIS health plan publication and set. So that means that public comment would occur this coming spring, so it would be like spring 2013. And then any changes would be published in the HEDIS volume that summer.

MS. BOSSLEY: And just as a reminder, we have an ad hoc review process. So if this measure should go forward as it is now, if you all voted to maintain endorsement,

when NCQA brings back a revised specification, most likely that would go through what I would call an ad hoc review where a small group of experts review the evidence, the changes that they may or may not make, and determine whether the measure endorsement should continue. So there's a process to accommodate the change in the future. Again, just want to make sure you understand what the options are, maybe.

MEMBER LOY: I just want to make sure I understand the stewards' position on this. This is a maintenance endorsement, maintenance review. So did you all take a look at the same evidence that USPSTF took a look at and have a similar conclusion that 40 to 49 is still appropriate, or was that part of the process?

MS. BYRON: What NCQA does when we develop measures is try to actually look at guidelines and trusting that the guidelines are following the process of basing their

recommendation on systematic reviews. You know, that said, we don't just take any old guideline. We do consider the USPSTF to be a highly-regarded and a very well-researched guideline that we usually trust.

We usually follow the recommendations that were put forth by the Institute of Medicine's "Guidelines We Can Trust." Guidelines for guidelines. And so we do try to look across the guidelines, see what they're saying, trust that they're basing it on systematic evidence reviews. We don't tend to do primary evidence reviews ourselves, because we are trusting the guideline developers to do that.

And so, you know, when a situation like this comes about, we find ourselves in the middle and we have to make those difficult decisions about what to do for the measure.

CHAIRMAN LUTZ: Karen, please. Help me.

MEMBER FIELDS: So I guess your

statement before makes it even harder to have this discussion, because most of the oncology societies actually still are going around and endorsing 40 and above on an annual basis. And so even though there's some other data out there that's been confusing ACS, NCCN, ASCO, everyone is still endorsing that. So I guess until some of the big oncology societies start to think about changing those endorsements, I don't know that NQF endorsing a measure that's a maintenance measure makes us choose sides. I think the medical societies that represent us already chose sides. That's just an editorial comment. I don't know. And I value the other members' comments about that.

MEMBER PFISTER: You know, I think that, as you pointed out, there's clearly not consensus with the guidelines, and certainly the oncology societies have aligned with not changing the age range. And oftentimes not changing the age range is kind of a path of resistance in political hotbed situations like

this. I think what's a challenge here is that certainly the organization that's recommending a different age range is certainly not by any metric viewed as a non-credible source.

And, you know, I guess what I'm wondering is if you were -- and this is a quality measure where you are going to evaluate activities based on what would be widely appreciated, something should definitely be happening. And if in a certain decade of life there's a disparity among people saying what should happen, then I think that questions of the robustness of the metric is applied to that decade, whether the focus should be on those age groups, which there is no disagreement that they should definitely be getting it done.

It's such that the measurement is not a distraction from sort of like, oh, and by the way, when you looked at those performance gap statistics, they weren't looking very good in the area where they

should definitely be getting it done. And everyone agrees there, you know? And so, you know, sometimes situations like this can lead to the sort of distraction from stuff that there's broad consensus should be happening.

MEMBER GORE: I just wanted to clarify what provider is being measured here as well, because there's a part where it says it's a physician-level measurement. And so how is that determined which physician is being measured or evaluated by the screen of their patient population?

MS. BYRON: Well, the measure has been re-specified for electronic health records, and so that means that it is for what they call eligible providers. So I believe it's any providers, because this is a primary screening, or it's a secondary screening measure for a general population.

MEMBER GORE: So if this is for a population, how is it anticipated that this is used for quality improvement? Sorry to ask a

redundant question from my previous one.

DR. BARTON: Let me make sure I understand that question, if I might. So like for a population, for a health plan?

MEMBER GORE: So who are we evaluating with this measure? So if you're using this to understand rates of screening in your entire health plan, then my question is how is that then used for quality improvement? Because there's a part where it says it's a physician-level measure. How is that physician determined so that it's not, for example, punitive to someone who saw that patient once? That patient, like me, doesn't see a primary care doctor ever. And so how is that determined?

DR. BARTON: Well, I think NCQA's greatest experience is with the HEDIS set being applied in health plans. And I think that this is a conversation that Heidi and we go back and forth on a lot, is how we talk about measures that have been specified at

other levels where we may or may not have data to show their use in those areas, but where we want to make it available to people to use because we think there's good justification for using a measure on other levels than the one we use it for.

So I'll just say that for health plans, health plans are required to submit data on measures with the agreement that we're going to publicly report their results. They're compared on websites that NCQA maintains. They're published in Consumer Reports. So, and if I were a health plan and I saw that I had a poor overall score, I would certainly go to my component care groups and suggest that they be willing to compare publicly their rates so that there could be all of the boats, you know, working together to increase the rate for the health plan.

MS. BYRON: And also, NCQA does publish benchmarks for the measures, just so that health plans can compare themselves

against these benchmarks. So that said, we have actually seen a lot of work from health plans at our different conferences where they do best practices, and many of them have used this measure. They've looked at their rates. They've seen that they might not be as high as they would like, or they've stratified their rates according to different race/ethnicities or other, you know, socioeconomic status and they've seen maybe that their rates are good for some populations and not others. And so then they've been able to do quality improvement around that, like provider education or reminders sent out to patients, their members. So we've seen it go at different levels.

MEMBER GORE: So how is the benchmark determined and is there a benchmark for this measure?

MS. BYRON: There is; and I would have to look it up, and it's based on the data from all of the other plans.

MEMBER GORE: So it's an average or a quartile? Okay.

MEMBER MILLER: Yes, so not to be redundant, but I just want to point out I think the discussion we're having here is reflective in part of the national discussion we had between the Preventative Services Task Force being comprised generally of people that don't treat cancer patients. And I think, not everyone, but a number of us are oncologists or predominantly deal with patients who already have the established diagnosis.

And so I guess, you know, I'm just putting out there as an oncologist I have a totally different perspective on this, that, yes, I mean, I think 40-year-olds should be screened because I see the bad end of it. And I understand the data and I understand this is a controversy, but I think, who was it, I think David said it, was, you know, maybe focusing our attention on the people, the 52-year-olds that aren't getting screened is

really where the money is.

And I just worry about this kind of distraction. We're coming to try to identify this as a quality measure that the individual physician at, you know, name the medical group is going to get dinged on because they didn't do their 42-year-old. I just worry that's going to be reflective of this whole national angst over this. And we're not smart enough to figure this out anyway. Certainly I'm not smart enough to figure this out.

CHAIRMAN LUTZ: Well, and I know we're not supposed to change, but just in terms of thinking about the processes, a hypothetical, if this has been brought simply with 50 to 69, would we have -- I mean, obviously we discussed the 40-year-olds and 40 to 49 would be left out, but would we have any problem saying, well, that's a quality measure, or would that still be a confusing issue because we're not including the people

in their 40s? I mean, it seems a fair corollary to ask. I don't know.

Bryan?

MEMBER LOY: And to that point I would just say was there any consideration either by the developer; I guess to the work group as well, in terms of the shared decision making component of this, to say we had the discussion about shared decision making and we excluded that somewhere? I won't get into where it comes out of, numerator or denominator.

MS. BYRON: So this measure is actually an administrative measure only, so it only pulls from claims. So our data source is claims and there is not way that we would be able to capture that.

MEMBER LOY: Yes. Understood.

MS. BYRON: So we are balancing, you know, being able to capture all of that information but keeping it feasible.

MEMBER LOY: Got it.

CHAIRMAN LUTZ: We've got David and then John.

MEMBER PFISTER: Yes, I mean, I think that, you know, this is a very, I think, passionate issue for oncologists, as you pointed out. But, you know, I think that we are a little bit at the mercy of what we see and, you know, it's sort of whether it's, you know, the last case we've seen and also the morbidity we see. While there's certain insights that oncologists have on this particular issue, you know, I think in all fairness there are certain insights that people that aren't oncologists have in this issue that look at, you know, health in a different way, see the downside of some of the false positives and never even make it to an oncologist, you know? So, and I think it kind of goes either way.

And it's not that, as a provider, following a given set of guidelines isn't a very defensible thing to do here. You know,

certainly you have very credible organizations that, you know, say that this is the deal. It's just a matter of, if you're going to evaluate a health plan, an individual or so forth, when you have a pretty major player in this business saying, you know, that it's not so clear-cut in this group, and then we're having a pretty explicit quality measure which applies to that decade, it implies a certainty, or an evaluative certainty which I think it seems isn't so clearly there amidst the controversy. And I don't think we're going to resolve it in this room.

I think what the pressure on the situation is, is that it's a metric used to assess, you know, quality. And someone might say, one group will say, well, gee, not doing it in this decade is under-penetration. Another very reputable group will say it's actually over-penetration. So it's not even like a neutral thing. And so, you know, I think that's the quandary we're in here,

because this is something which -- you mention about evaluating the health plans, so how meaningful is that evaluation of a health plan in that decade of life in terms of being able to interpret that, except for the fact that two different guidelines panels disagree and you end up with a number that I'm not sure how actionable that number is?

CHAIRMAN LUTZ: I'm sorry, before we go to John, because directly to that, I mean, in practice if I see someone who's 40 and is being seen, a female being seen for another reason, I start discussing the merits and drawbacks of screening at the age of 40 and let them get screened if they want. And then I am much more dogmatic about starting screening at 50. I mean, I don't know. Again, I'm not trying to rewrite the submission, but I mean, in some ways there has to be some common sense to this.

MEMBER GORE: Yes, and I'm going to build upon that as well in that this

measure, it's not -- you know, the physicians who are going to be evaluated by this measure, either by their health plan or individually, are overwhelmingly going to be primary care physicians. And so when you look at survey studies of what informational materials primary care physicians use, they do predominantly use the USPSTF.

And so I think when you're feeding back to primary care clinicians about, you know, their breast cancer screening when their predominant guideline says to start at 50, I think they would be hard-pressed about being dinged for not doing it between 40 and 49.

MEMBER FIELDS: I guess the most important point is the median age of breast cancer in the United States is 51 or 52. That means that half the patients are below that age when they're diagnosed.

The median age of breast cancer is 65? Okay. Then I am incorrect.

The problem with screening

mammography, the state of the art is such that the specificity of the test goes up with age and it's not specific in the younger patient population. Unfortunately, it's the test that we have. It's the widely-available screening tool. So there's a patient population in the ages of 40 to 50 that we just haven't addressed.

So that's the reason we have guidelines problems right now. And my only answer is I think that the main issue that we have as a problem is that we don't have a better way to screen that patient population. And that's where the disparity comes. I mean, that's where the issues come in.

I also think that there's a disparity problem between Medicare and commercial payors and Medicaid payors. And so, I don't know that this measure addresses that. I mean, the most important thing is to understand why Medicaid and Medicare patients aren't being screened at the same rate as

commercial payors if these guidelines are out there and available.

MEMBER MALIN: I think, you know, there's complex issues related to disparities, but I know from some of the work that the Quality of Care Department at WellPoint has done, you know, part of the challenges that -- you know, screening requires an activated patient population as well. And, you know, often patients who are in under-served communities have more pressing needs in terms of survival than getting out for their screening. And so, you know, it's not necessarily having access and having physicians. I mean, there's a lot to get people in for their screening.

MEMBER LOY: Just to emphasize from the health plan perspective again, if HEDIS scores are valuable to employer groups and a health plan finds themselves with less than their competition, then there will be some pressure back to find out who those

segments are that aren't being screened. So I think it would be na‹ve to think that there would not be some pressure back on the system, and it wouldn't be pressure to get towards a shared decision making or a conversation. The measure, as, you've already pointed out, would be addressed or acknowledged through claims payment. So I don't know of a claims payment mechanism that could overcome that shortfall.

And then the other point that I might make would be many commercial plans in the industry use USPSTF as their basis for screening coverage. And I don't know how folks are coping with that today. I don't know how regional versus the large plans are coping with that discrepancy today amongst different guidelines that exist out there. It is hypothetical, but I think that there could potentially be a situation where you've got a quality measure where there might not be commercial coverage that's available, if your commercial coverage allows for those

screenings that are acknowledged by USPSTF.

CHAIRMAN LUTZ: John?

MEMBER GORE: And to build on that --

CHAIRMAN LUTZ: I'm sorry, they want to respond to that. Sorry.

MEMBER GORE: Oh, sorry. Sorry.

DR. BARTON: I think it's true that there are insurers who look to the U.S. Preventative Services Task Force. I think though that an act of Congress that came along in 2009 instructing HHS to disregard the U.S. Preventative Services Task Force recommendation on mammography screening was a powerful message.

MEMBER CHOTTINER: I think, without trying to oversimplify. I'm looking at this, and in 2012, with the evidence that we have, I think it's hard to answer that first question about evidence with anything other than it's insufficient right now to recommend for or against in that patient population. So

the question becomes: do you judge the measure on that basis or is there any opportunity to modify the age range in the measure?

So did the developer want to respond?

MS. BYRON: So because the task force did not come out with an insufficient evidence -- it's actually a C grade for the 40 to 49, and I think the issue is that across guidelines we don't necessarily have agreement. By stratifying the measure as 40 to 49 and then, you know, 50 to 59, or 50 and up, we may be able to get around some of these problems that I think the entire medical community in addition to measure developers are struggling with here.

It's possible that if we are able to get that change for the measure, which I can't promise anything because, you know, we do rely on our advisory groups to help us go through that process, and you know, we would post it for public comment, and I imagine we

would get lots and lots of comments -- but by stratifying in that way, it may give us that ability to either disregard that age group for that stratification -- if I were someone implementing measures for whatever reason; quality improvement or payment, I might be able to say we would only like to focus on the 50 and up age because that is where across the guidelines we have agreement.

For the 40 to 49, that could be something that you not use in a program or in a payment system. You know, that would be up to people implementing the measure, but would give people an opportunity to address the different age groups in different ways. So that is where we think we might be able to go with this measure.

CHAIRMAN LUTZ: So do we need to vote, or after the discussion do we have them go talk that over? I mean, what's the best way procedurally to deal with that?

MS. FRANKLIN: So we can go

through the votes for impact and opportunity for improvement. And if we feel strongly there is an opportunity for improvement and strong impact, but there is a weakness in the evidence base, we'll have to take that vote. And then if it fails at 1c, which is the evidence base, we can as a steering committee decide if we want to move the measure forward anyway and invoke the exception at that time.

MEMBER FIELDS: So the way the data was presented, it's hard to tell if there is an opportunity for improvement because the age range wasn't stratified in the way they gave us the data. So we can't necessarily say where the shortcomings in the data are, if it's in what age groups, unless you have a clarification on that.

MS. BYRON: That's true. For HEDIS, right now the measure is not stratified. So we anticipate that may be a possible change in the measure to allow for that stratification.

One way you can look at it is that you can compare across different product lines. So commercial plans have a rate that's right around 69-70 percent as a mean. Medicare I think is around the same, if not a little bit lower. And Medicare, it's down to about 50 percent. So that's one way to think about whether or not there's, you know, an opportunity for improvement.

MEMBER FIELDS: So the way you could interpret that might be that the Medicaid patients are the younger patient group?

MS. BYRON: Or disadvantaged.

MEMBER FIELDS: So then that's why they're falling outside of the guidelines? Because the Medicare patients have it covered as part of their coverage.

MS. BYRON: For 65 and up. So it probably is fair to say they're a little younger, but they are also, you know, disadvantaged in other ways. So I think it

would be hard to assign the reason to just one factor.

MEMBER FIELDS: It just makes it hard to follow that one recommendation, which would be since we can't interpret the data, I would think.

MS. KHAN: Voting on 1a, impact.

So we have eight for high, one moderate, zero for low and two for insufficient.

And voting on performance gap.

We have four high, four moderate, two low and on insufficient evidence.

And going onto 1c.

I'm missing one person.

So we have two yes, one no and eight insufficient evidence.

MS. BOSSLEY: So this is where again, as Angela said, there is the exception. It typically deals more with consensus-based guidelines that you're looking at.

I'm wondering, and again, just let

me throw out an idea, because you still have comments to come. And so one thing we could do is have you assess the measure against all the criteria. At the end of the day it could go out with you seeking additional input from the membership and the public before you make a final recommendation. We have done it in the past. It is an option before all of you.

Or if you feel like you have assessed this and you don't want to move forward, we won't move forward. But that is again another option you can run through. You can assess all the criteria and then see where we land, and then if you choose, we could actually put it forward for more input before you make a final final recommendation. Throwing it out as an option.

MEMBER MILLER: I actually like that idea, and I'm not just saying that because it's 5:15. Again, maybe I'm speaking as a cancer doctor who sees people who already have a diagnosis of cancer and I have my bias.

We said it's a health plan measure. It's a primary care measure predominantly. And I would love for those stakeholders to have an opportunity to influence my opinion.

MEMBER PFISTER: Yes. No, I think getting more input is definitely the way to go. You know, this measure's been around a very long time. Okay? And obviously this is a very controversial area where there's been a lot of back and forth about it and I think that, you know, getting comprehensive input on this I think is particularly critical here.

MS. BOSSLEY: So in order to do that, often, especially at the end of the day, it's helpful if you could assess the rest of the criteria because that may be helpful when it goes out for comments. So they can see how you assessed it against scientific acceptability, usability, feasibility, and then the overall. Again, we can just put it out if everyone agrees that it would just be seeking additional input, you're not yet sure

what recommendation you should make, if that makes sense to everyone. Because I think everybody's in this dilemma and we just need more comment, it sounds like, from the external stakeholders.

CHAIRMAN LUTZ: Okay. So that means we continue on with the voting? Okay.

MS. KHAN: So moving on to -- are we going to have discussion?

MS. FRANKLIN: We're looking at 2a, reliability, under scientific acceptability. And if there's any discussion about that?

(No response.)

MS. KHAN: So voting on reliability, 2a.

CHAIRMAN LUTZ: Nicole, do you have anything you want to say about reliability? You don't have to. We just didn't want to leave you out. Can someone nudge Nicole, wake her up?

MEMBER TAPAY: I mean, I think

that the group had felt that it was fairly clearly stated in terms of the reliability. The question was more around the validity and the age. So I don't have anything more to add.

MS. KHAN: So voting on 2a, reliability.

So we have six high, three moderate, zero low and two insufficient evidence.

And voting on validity.

I think we're missing one person.

So zero for high, six for moderate, two low, three insufficient evidence.

MS. FRANKLIN: All right. Moving on to a vote on usability.

But first, Nicole, did you have any comments around usability, and discussion from the group?

MEMBER TAPAY: We didn't really have any on that point.

MS. KHAN: So voting on usability.

We have two high, five moderate, two low and two insufficient information.

And feasibility. Was there anything?

(No response.)

MS. KHAN: Voting on feasibility.

That's nine high and two moderate, zero low, zero insufficient information.

And overall suitability for endorsement. Does the measure meet NQF criteria for endorsement?

So we have two for yes and nine for no.

CHAIRMAN LUTZ: All right. Just to prove how strong we are, there's one left and the developers have said that they're available to 6:00 and really requested if we could do it tonight, that would be good for them. Besides, Patrick said the dance bars downtown don't really get going until about 9:30 or 10:00. So we got a lot of time.

That's what he told me.

(Laughter.)

MS. FRANKLIN: So do we have someone from AMA/PCPI who will tee up the measure for us?

MS. TIGHE: Actually, we may just want to actually let Dr. Miller go first, because he has to run off.

MEMBER MILLER: I have to catch a 6:05 train.

MS. FRANKLIN: Okay. Dr. Miller, if you could start us off. Go ahead.

MEMBER MILLER: Yes. So very quickly, this is a measure similar to one we saw many hours ago. This is adjuvant therapy of hormone receptor positive breast cancer measure. This is the use of tamoxifen or aromatase inhibitor for appropriately selected patients, stage IC through IIIC, that are ER/PR positive. This is a process measure and the level of analysis is at the clinician, the individual physician group.

In terms of the impact, little doubt of the importance of this. The most common type of breast cancer. Evidence very high that the intervention is effective in improving disease for survival and overall survival.

There is a performance gap in terms of the QOPI measures. Performance was at 94 percent, but other patterns of care study, particularly in under-served populations have been considerably less good than that, 80 percent or so.

And I'll just summarize very quickly and say I didn't have any concern with the evidence. It was high-quality evidence, multiple studies. So I'll leave it at that.

MS. FRANKLIN: Any comments from the developer?

DR. ANTMAN: My colleague Sam Tierney is on the line, so I'll defer to her to see if she wants to add anything.

MS. TIERNEY: Thank you for your

comments. The only thing I would add; because it wasn't available at the time we submitted the measure, is that there was some data from PQRS in 2010 related to this measure that showed that the average performance rate was about 90 percent. So that information is not available in a range, so we're not sure of the range of variability within that, but I just wanted to also share that additional information from the recent use of the measure in PQRS.

MS. FRANKLIN: Okay. Thank you. So focusing on importance, do we have discussion around importance? Dr. Loy?

MEMBER LOY: Just would say that point's already been made today. It feels like we're missing the compliance/adherence piece of this, rather than just the prescription. I just would say, as we move forward, if that's a consideration the developers would take away. I think that's contemporary.

MS. FRANKLIN: Additional comments about the importance? Karen? Okay. Any other comments?

(No response.)

MS. FRANKLIN: Then we're ready for a vote.

DR. HASSETT: Can I make a comment?

MS. FRANKLIN: Oh, yes. Sorry, on the phone?

DR. HASSETT: I'm sorry. This is Michael Hassett. I'm a medical oncologist in breast cancer.

Two quick comments, one about the gap issue. There are a number of studies that I've looked at compliance relative to this measure in other patients that would suggest that there are some particularly disparity-focused populations where compliance is much lower, probably in the 60-percent range.

And with regard to the adherence issue, I would certainly support the concept

of an adherence-related measure as well. I think we actually probably need both on the market, an initiation measure and an adherence measure.

MS. FRANKLIN: Thank you.

DR. HASSETT: Thank you.

MS. FRANKLIN: Any other comments from those on the phones?

(No response.)

MS. FRANKLIN: No? I think we're ready to vote. All right. Then we're ready to move to a vote on 1a, impact.

MR. CUNNINGHAM: Okay. Now voting on 1a, impact.

We have 10 high, one moderate.

Moving onto 1b, performance gap.

Seven high, four moderate.

Moving onto 1c, evidence.

Ten yes, one insufficient.

CHAIRMAN LUTZ: So any discussion about reliability?

MEMBER MILLER: So none about any

of the other measures felt comfortable with our discussions. My own analysis, we met all the other criteria.

MEMBER FIELDS: I just had a question. When you looked at the expert panel for validity, 80-90 percent of them put it in category four or five, and we're sort of used to seeing higher validities there. And I assume they're saying it was because there was a high exception rate to who wouldn't get the drug, but I just wanted to ask how to interpret that, or any comments. Because it seems to me like a very valid measure.

MS. CHRISTENSEN: Yes, we didn't actually ask for comments on that. That's something that we've changed since then to find out more if it's not a four or a five, you know, what their particular thing was. There were two people on here that put a three, which was -- just make sure I get the word right, sorry -- neither disagree nor agree. So it's not disagreeing. It's just

not very high.

MEMBER FIELDS: That was like 30 years of data on disparity in healthcare. I just didn't understand if there was something we were missing about --

MS. CHRISTENSEN: Yes, we felt the same thing.

MEMBER FIELDS: -- the validity of the test. Okay.

MR. CUNNINGHAM: Onto 2a, reliability.

Ten high, one moderate.

Moving onto 2b, validity.

One more vote.

Eleven high.

Any more discussion?

MS. BOSSLEY: I get the feeling you all feel you've discussed this enough. You want to just vote? Okay. We'll vote.

MR. CUNNINGHAM: All right. Cast those votes.

Eleven high.

Moving onto feasibility.

We need one more. Please hit it again.

Nine high, two moderate.

Moving onto overall suitability for endorsement.

Eleven yes, zero no.

CHAIRMAN LUTZ: It is incumbent upon us to ask for public comment. Anyone for public comment?

(No response.)

CHAIRMAN LUTZ: All right. Hearing none, we will see -- oh, go ahead.

DR. CHIN: Hi.

CHAIRMAN LUTZ: Go ahead.

DR. CHIN: Hi. This is Lindee, Dr. Lindee Chin from ActiveHealth again. I just had a suggestion for the steering committee about our measure again.

CHAIRMAN LUTZ: Go ahead.

DR. CHIN: So I just wanted clarification that if we're basing our measure

on an NCCN guideline that's as recent as -- it's been updated as of January 2012, still recommending surveillance for this group of people -- I guess I'm confused as to what other data that you would have liked to have seen to qualify the importance of this measure.

CHAIRMAN LUTZ: I'm not sure if there was data, so much as there's so many different options as to things that can be considered in quality measures. And I think it's trying to put some gradation on the things that are most pressing and things that are important, but maybe not the most pressing in terms of measurement. I'm not sure there was anything missing as much as it's sort of put upon us to find the things that are emergent and important and topical right now. I don't know if that helps.

MEMBER MALIN: Also, I think the rationale for surveillance in the NCCN guidelines for breast cancer, for colorectal

cancer, for lung cancer is based on the data in the non-impacted population. And so it's harder to make a case, I think, that there's evidence to support a different indicator for the affected, the survivorship population rather than just using the same indicator that you would use for the general population.

DR. CHIN: And I guess our concern is, though, then if you just put these people under the bucket of screening, then the other measure, you're not going to capture those people that are under the certain age limit that you're capturing with screening.

MEMBER MALIN: So you're saying that maybe we need a measure for young breast cancer survivors?

DR. CHIN: Perhaps. I'm not sure. That's why I'm trying to figure out how do we capture that population, because the screening measure's going to miss those people who are younger than the screening guidelines.

MEMBER MALIN: Well, currently

they wouldn't, right? I mean, although we just voted down the measure. But the current measure starts at age 40 and the number of women with breast cancer who are younger than 40 is incredibly small. If the general screening measure gets revised to be 50 and above, then I think you may have a case to be made to come in with a targeted measure for breast cancer survivors who are under age 50 who wouldn't fall into the regular screening guideline.

DR. CHIN: And I guess my other question is, I'm just wondering why this was endorsed a couple of years ago, but now it's not. And we didn't really change the measure that much because we were applying the same guidelines. So I'm just confused as to why in the past it was believed to be more important than it is today.

MS. FRANKLIN: And this is Angela. Our criteria here at NQF has changed and become a little more stringent over the last

couple of years, and it would have been a different level of review then than there is now. So our criteria have changed, and that's what the committee is looking at in reviewing the measure today.

DR. CHIN: Okay. Thank you.

CHAIRMAN LUTZ: Any other public comment?

(No response.)

CHAIRMAN LUTZ: What are you thinking? I'm good with 8:30. Depends whether Pat will be over his dancing or not. No, I'm good.

MS. FRANKLIN: So for tomorrow we have a motion on the table to start with a working breakfast at 8:30 tomorrow morning and the review of the measures will begin during that area. So we'll start tomorrow at 8:30 with our discussions. Thanks, all.

(Whereupon, the meeting was adjourned at 5:37 p.m.)

## C

**N**

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Cancer Endorsement Maintenance Steering Committee

Before: NQF

Date: 05-23-12

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

---

Court Reporter