NATIONAL QUALITY FORUM

Moderator: Kim Patterson
February 26, 2020
08:00 am ET

Jodi Maranchie: Can anybody hear me?

Man 1: I can hear you.

Jodi Maranchie: This is Jodi Maranchie.

Man 1: Hi. Very good…

((Crosstalk))

Jodi Maranchie: I'm going to see if I can transfer to the computer audio if I can, then I will come in that way. Actually, if I do that you won't be able to hear me at all, will you? I guess I'll use both. Do we know what's going on?

Nicole Williams: Good morning.

Jodi Maranchie: Hi.

Nicole Williams: This is Nicole Williams with the National Quality Forum. Hello, thank you so much for joining us, everyone on the phone and in the room. This is the Measure Evaluation Meeting for the Cancer Project - Cancer Consensus Development Project for the Fall 2019 Review Cycle.

And I'd first like to just welcome the Committee joining us here in person and certainly on the phone. It's very early Wednesday morning. We have a packed agenda and we're excited to get started.

We have nine measures that we'll be reviewing today and we'll start with introductions and go through Disclosure of Interest. We then will give a brief overview of our evaluation process. We'll walk through each of the candidate's standards for consideration and review them against our Measure Evaluation Criteria. At the end of our meeting, we will open up the floor for member and public comment and then discuss next steps.

I'm joined by a great team here at NQF, as well as our Co-Chairs and I will open it up to the Co-Chairs for comments.

Karen Fields: Good morning, I'm Karen Fields. And I'd like to welcome all of - all of you to the Fall session for 2020 - or 2019, we're already in the next year. I'm from Moffit Cancer Center. I'm the Medical Director for Clinical Pathways and Value Based Cancer cures and animal practicing medical oncologist and I've been involved with this committee for several years.

We'd like to first welcome all of the standing committee members. We also welcome those online. We want to thank the staff for all the hard work getting us prepared for this meeting today and the developers who are going to be present today, either in person or online, as well as anyone else

participating online. So, thank you for being here and we look forward to a very busy day.

Shelley Fuld Nasso: Hi, I'm Shelley Fuld Nasso. I'm happy to be here. I'm the CEO of the National Coalition for Cancer Survivorship, a patient advocacy organization based here in the D.C. area. I've also been involved with this committee for a while and was formally on the Consensus Standards Approval Committee, CSAC, which you'll hear about a little bit later, because one of the measures we're reviewing is one that they - (deferred back to us) after our last review of it. And I can tell you a little bit more about what happened there.

But I'm just really happy that we're mostly all here in person today because it's so much nicer to be able to see everybody's faces when we're working on this. It's hard to do by phone. I think we've done a great job of it, but it's nice that we have the chance to be here today. So, thank you everybody for traveling from all over the place to be here and to all the NQF staff.

Ava Clarke: Hi everybody. My name is Ava Clarke and I'm the Acting Vice President of Quality Measurement. And so, I would just like to sort of piggyback on what the Co-Chair sand and say a big thank you to all of you for traveling and being part of our work. You - we couldn't do our work without your expertise and your volunteering of your time and you make our work really so much better.

So, I really wanted to say a big thank you for all the time that you spend, both coming here and discussing, but reviewing measures, having dialogue with our developers, things like that. So, I very much appreciate that.

We're going to combine introductions and Disclosures of Interest together, so, you should have received the Disclosure of Interest form from us. We asked you a number of questions about your professional activities. Today, we'll ask

you to orally disclose information you provided that you believe is relevant to the committee. We are especially interested in Grants, Research, or Consulting related to the committee's work.

Just a few reminders. You sit on this committee as an individual. You do not represent your employers, or anyone who may have nominated you for this committee. We are interested in your disclosures of both paid and unpaid activities that are relevant to the work in front of you. Finally, just because you disclose does not mean you have a Conflict of Interest. We do oral disclosures in the spirit of openness and transparency.

So, I'll ask you to please state your name, where you're from, and if you have any disclosures. So, we'll start with our Co-Chairs. Karen Fields.

Karen Fields: Karen Fields. I have no - nothing to declare.

Ava Clarke: Shelley.

Shelley Fuld Nasso: Shelley Fuld Nasso. I have no disclosures.

Ava Clarke: All right, we'll start with our committee members. (Oksana Varsey).

(Oksana Varsey): (Oksana Varsey). (Unintelligible) no disclosures.

Ava Clarke: Yes, just a reminder, if everybody could use your microphone. If you want to just press the button at the bottom, the red indicates that it's on. (Gregary Bocsi.

Greg Bocsi: Good morning, this is Greg Bocsi…

((Crosstalk))

Ava Clarke: And I just will apologize in advance if I mispronounce your name. Fred Brakeman.

Fred Brakeman: Good morning, Fred Brakeman, University of Texas, MD Anderson Cancer Center. No disclosures.

Ava Clarke: Steven Chen.

Steven Chen: Hi, I'm Steven Chen. I'm a surgical oncologist in San Diego, California. In the interest of full disclosure, I'm also the Chief Oncologist of a startup biotech company in San Diego that does not have a relevant disclosure to any of these, but we are in the oncology space.

Ava Clarke: Heidi Floyd.

Heidi Floyd: I'm Heidi Floyd. I'm a patient advocate and I'm from Atlanta, Georgia. No - nothing to disclose.

Ava Clarke: Brad Hirsch.

Brad Hirsch: Brad Hirsch here. Medical Oncologist with Texas Oncology, as well as CEO of a startup called SignalPath in the clinical research space. No disclosures.

Ava Clarke: Great. Jette Hogenmiller. (Lenore Johnson).

(Lenore Johnson): Hi, I'm (Lenore Johnson). I'm from the Chicagoland area. I'm a patient advocate and I have no disclosures at this time.

Ava Clarke: Great. Leonard Lichtenfeld.

Len Lichtenfeld: Good morning. Len Lichtenfeld, American Cancer Society. I've no relevant disclosures other than the fact I'm a member of ASCO and I've had no relationship currently, nor for many years, with their quality reporting committees.

Ava Clarke: Okay, great. Stephen Lovell.

Steve Lovell: Good morning, Steve Lovell. I'm from the Seattle area, a patient representative and no disclosures.

Ava Clarke: Jennifer Malin. Jodi Maranchie.

Jodi Maranchie: Hi, this is Jodi Maranchie from the University of Pittsburgh. I have nothing to disclose:

Ava Clarke: Great. Denise Morse.

Denise Morse: Hi, can you hear me?

Ava Clarke: Yep.

Denise Morse: Okay. Hi, Denise Morse. I'm from the City of Hope Cancer Center in Duarte, California. And I apologize for not being there in person and I have nothing to disclose.

Ava Clarke: Great. Ben Movsas.

Ben Movsas: Yes, hi. I'm - also apologize I cannot be there in person. I'm from Henry Ford Health System in Michigan and I'm a Radiation Oncologist. And I have nothing to disclose related to these Quality Measures.

Ava Clarke: Great. Robert Rosenberg.

Robert Rosenberg: Hi, I'm a Diagnostic Radiologist in private practice, in Albuquerque, New Mexico. I'm on the Board of a National Accreditation Program for Breast Centers, which is part of American College of Surgery, but not related to any of their measures or to the COC.

Ava Clarke: Great. David Sher. Danielle Ziernicki.

Danielle Ziernicki: Good morning, this is Danielle Ziernicki. I apologize I can't be there as well. Thanks to the committee for having everyone in this meeting. I am a Doctor of Pharmacy. I am the head of Medical and Regulatory at the Dedham Group, and I have nothing to disclose.

Ava Clarke: Do you have something else?

Len Lichtenfeld: Yes, actually I do, since Rob made a comment. This is Len Lichtenfeld and I am also on the Board of the National Accreditation Program for Breast Centers, not involved in any Quality Metrix for them as well, so I don't view that as a great conflict.

Ava Clarke: Okay, great. Is there anybody whose name I didn't call? Or anybody who might have joined after I started role call? Great. I'd like to remind you that if you believe that you might have a Conflict of Interest at any time during the meeting, please speak up. You may do so in real time during the meeting, or

you can talk with anybody on the NQF Staff or to the Chair or just send a note to anybody.

If you believe that a fellow committee member may have a Conflict of Interest, or behaving in a bias manner, you may point this out during a meeting, send a message to the Chairs or to anybody in the NQF, or talk to any of us. Do you have any questions, or anything you'd like to discuss based on common disclosures today? All right, great. Well I'm going to turn it back over to…

((Crosstalk))

Katie Goodwin: Thank you both. Good morning everyone, I'm Katie Goodwin and I'm going to give you an overview of the evaluation process for today's meeting. As standing committee members, you are acting as a proxy for the NQF Multi Stakeholder Membership.

And today, you will be evaluating each measure against each NQF Measure Endorsement Criterion, ultimately making recommendations regarding endorsement to the NQF Membership. We do ask that during discussion, you indicate the extent to which each criterion is met and the rationale for your ratings.

Some ground rules for today's meeting. During the discussion, we ask that you base your evaluation and recommendations on the measure evaluation criteria and guidance. Please attend the meeting at all times and remain engaged in the discussion. Keep comments concise and focused and indicate agreement without repeating what has already been said.

So, the process for measure discussion and voting, we are fortunate as mentioned earlier, to have our measure developers with us today. They will be introducing the measure and will then turn it over to the lead discussants to lead discussion of the measures - in relation to the measure evaluation criteria.

We ask that the lead discussants provide a brief summary of the premeeting comments that were submitted by standing committee members, emphasizing areas of concern or differences of opinion.

We also wanted to note that the preliminary rating by NQF is intended to be used as a guide to facilitate the committee's discussion and evaluation. The developers will be available to respond to questions or provide clarifications as needed during the standing committee's discussions. So, first the full committee will discuss, then vote on the criterion and then move onto the next criterion.

The votes will be taken after the discussion of each criterion, so we'll be starting with Importance to Measure and Report, which is must pass. And that includes both evidence and performance gaps. We'll then move onto Scientific Acceptability of Measure Properties, also must pass. And that includes reliability and validity.

After Scientific Acceptability comes Feasibility, followed by Use, which is now must pass for Maintenance Measures. The last criterion is Usability before we move onto Overall Recommendations for Endorsement. As a reminder, you are voting on the measure as specified, as it was submitted, and as it is before you today.

Here at NQF, for the purpose of this Standing Committee Meeting today, we define Quorum as 56% of the Committee and we do have Quorum present.

We needed a minimum of 13 Committee Members present in order for us to vote. To achieve a pass or recommended rating on a measure, we need greater than 60% yes votes of the quorum. And we do add high and moderate together to come up with that greater than 60%.

Consensus not reached is when the votes fall in between 40 and 60% of the yes votes of the quorum. Does not pass or not recommended is when the votes are less than 40% of the yes votes of the quorum.

When we do have the case of consensus not reached, these measures will move forward to the public and NQF member commenting period. The committee will then review the comments that were submitted during that period and revote during the post comment call in May. Any questions about the process for today's meeting? Okay. I think now, I am turning it over to Hannah and we'll walk through a test of the voting.

Ava Clarke: So, just really quickly before we do that, I just want to make sure - I think we found a couple of other committee members join the call. So, I think - so if you have not completed your disclosure, I ask you to disclose. So, I think - Jennifer Malin, did you join?

Jennifer Malin: Yes, I did.

Ava Clarke: Okay, could you just state your name, where you're from and if you have any disclosures.

Jennifer Malin: Good morning everyone. I'm Jennifer Malin. I'm from United Healthcare and I have no disclosures.

Ava Clarke: Great. Anybody else that might have joined that has not done disclosures? Okay, great. Let's move to the test vote.

Hannah: Hi everyone. So, yesterday a voting link was sent out, this is how we'll cast our votes for each measure. I'm going to go ahead and activate the vote and go ahead. There's no right or wrong answer here. If anyone's having issues or needs me to send them the link directly, please let me know?

Ava Clarke: Hi, April.

Lenore: Lenore, I'm sorry. Quick question for us. We're into the (unintelligible) the first measure that's coming up. Can I just sit back and kind of see how it goes? Or the example? Or…?

Ava Clarke: So, if you're new to the committee, we'd like you to participate in a discussion and also vote on the measure. So, feel free to vote your conscience.

Hannah: Looks like we received everyone's vote so we're ready to proceed.

Ava Clarke: Okay. So, I'd like to call the committee together to start and considering the first Measure which is NQF 1858 Trastuzumab Administered to Patients with AJCC Stage 1 T1C - III and Human Epidermal Growth Factor Receptor 2 or HER 2 Positive Breast Cancer Who Receive Adjuvant Chemotherapy.

The developers are the American Cancer Society - or excuse me American Society of Clinical Oncology and I will first ask the developers to give a brief introduction. I'd like to say that the lead discussant will be Dr. Lichtenfeld. Additional reviewers are Greg Bocsi, Brad Hirsch, Lenore Johnson, Jodi Maranchie, and Danielle Ziernicki. And we'd note that there's no committee conflict.

So, one other thing we didn't remind the discussers are, that when you want to make a statement, we usually flip our cards so we can tell and try to go in an orderly fashion. And again, always - I think we can only have three of these on at a time, so be sure to turn it off. We want you to turn on your light, but also turn it off. But, thank you for the developer's presence today. And we'll let you get started.

Dave Ryan: So, I guess we'll just introduce ourselves first. So, I'm Dave Ryan from - representing ASCO. I'm the Chief of Rheumatology Oncology at Mass General in Boston and also the Clinical Director of the Cancer Center there. I'm a Professor of Medicine at Harvard Medical School.

I have no relevant disclosures to the Measures that we're discussing today, although at a full disclosure, I do have an interest at NPM Capital, which is a venture capital firm in Boston. I have a major grant in pancreatic cancer research out of (center for) Cancer and I do write the up-to-date guidelines or review for pancreatic cancer, rectal cancer, and anal cancer.

Caitlin Drumheller: Good morning everyone, I'm Caitlin Drumheller, I'm the Measure Developer with ASCO and, we also - I don't know…

((Crosstalk))

(August Malfi): And I'm (August Malfi) also Measure Developer with ASCO.

Caitlin Drumheller: So, just for the (unintelligible) meeting do you prefer the Clinical Expert kind of given (unintelligible) of the measure? Okay, all right.

Dave Ryan: So, you can all read this measure. So, we think that the reason this measure matters, is that the process evaluated in this measure are for two positive patients being administered Trastuzumab within 12 months of a breast cancer diagnosis.

As you all know, multiple randomized control trials have demonstrated that the administration of Trastuzumab improves (over) two positive patients. the V3 survival and overall survival. Additionally, this measure is directly supported by recommendations in NCCN, Cancer Center Ontario, an ASCO.

Clinical Practice Guidelines. Both their analysis of Meaningful Difference and Performance, as well as the Chief (Unintelligible) 17 EPP Experience Reports indicate that this measure is topped out.

Nevertheless, our aim is to secure reserve endorsements to ask for this measure, while we work to deploy a replacement measure on appropriate treatment for HER 2 positive breast cancer patients, which has been developed and is currently undergoing testing.

Ava Clarke: So, why don't we turn the floor over to the discussants?

Len Lichtenfeld: Well that was a succinct and brief presentation and I wish I could be as good as you are in doing that. It's Len Lichtenfeld for those on the telephone. The bottom line is, that this is a (status) of measure of the administration of Trastuzumab to women who are HER 2 positive, with breast cancer.

Interestingly, I - there's a comment that says 15% of the patients - I've heard various numbers, but that's the lowest one. I don't know if that is really a factor number, I was more - I thought it was 20-25%, but I can always stand to be corrected by people more knowledgeable then I.

But that's not really relevant to the discussion. From my perspective and I think from the perspective of the comments, this is a standard of care measure. It was first introduced in 2012. It was endorsed - it was approved in 2012. It was endorsed again October 2015 and now is present for - being offered for re-endorsement.

It is a standard of care; however, as you noted in your comments, that standard of care is shifting and could lead to some question or some confusion. Having said that - Karen, if you don't mind me sharing - we did - Karen and I did discuss that this morning and the reality is that no matter what new regimens or new approaches might be available, Trastuzumab remains a critical part, whether (CO) adjuvant, post-treatment, what-have-you. And knowing that another measure is forthcoming, that certainly is relevant information.

The evidence is strong. It's been the subject of a number of articles from - and it's clearly the standard of care - so I don't believe there's any question there. If there are questions, the one that was brought up in the review was that of reliability.

Based on the unit of measurement, whether - what level - and perhaps someone could help me understand that a little bit better - it was actually measured at the facility level but reported at the group and practice level and not the individual level. And it was considered - the reliability was considered insufficient. Validity was considered high and acceptable.

I think the key question before the committee is the issue about being topped out. The reported rate of appropriate care was 97.1%, if I recall. That suggests that there's a very limited amount of improvement that could be anticipated.

In fact, I think for most things (unintelligible) if we got to 97%, we'd be pretty happy. Maybe with 3%, could be patients were not being properly treated, who don't have appropriate exclusions. That is a possibility, although it could also reflect the fact that people didn't record the exclusions whatever the reasons may be.

So, I don't know Karen, if you want any further discussion or (Unintelligible) want further discussion or commentary, from my perspective the measure is appropriate. Reliability is a question. Validity is acceptable.

And the primary question is the fact - the noted by the discussant - that this measure has topped out and whether it goes into reserve status or we continue I think that's, again, a matter of technical consideration. I don't know whether from staff and others in the committee.

Ava Clarke: And so - I'll call on you next Steve. So, I guess what we'd like to do is follow the order of where we need to vote. So, the first area that we like to finish the discussion is on whether or not the evidence is adequate. We do recommend that the committee vote again on evidence, because it was updated.

So - but if there's any - Steve, if you had a question, you (unintelligible) your question.

Steve Lovell: Yes, and - this is Steve Lovell. And, you know, I guess maybe - and I'm somewhat new to doing the measures that are maintenance measures - but if something is topped out - and even knowing that there's another measure may be coming adjacent to that - is it standard to kind of keep that open if it is already capped - topped out? I'm just kind of curious from a background and a history perspective.

Woman 1: I'll let you take that.

Woman 2: So, I think that's - (unintelligible) as part of evidence of a conversation about sort of performance gap. And so, I would - I think that would be a good time to have this conversation. I think that's a conversation that the committee needs to have, right?

So, I think the idea is, is that you as a committee and clinical expertise, need to talk about - well, okay, this measure might be tapped out, but what are the effects of that if not endorsing that and kind of things like that - I think that's actually a conversation that the committee should have.

So, we have - there's sort of two parts of evidence. There's a sort of importance to measure and then there's the gap. So, I'd start with the first one and then sort of have that conversation about gaps. Thank you.

Woman 3: I just have a quick question for the developers and I'm sorry if I wasn't here and I missed it. But I would assume that now that there's a (unintelligible) (approved for a test) (unintelligible) that if someone gets the (unintelligible) it will still count? Okay.

Ava Clarke: And just a housekeeping - if you want to speak, just put your (TEM) Card up and that lets us know and then the…

((Crosstalk))

Len Lichtenfeld: Just a quick question. You using 12 months as the guideline for being administered? It seems long. Is there a tactical reason why you're using 12

months? I mean would that relate to being topped out or would that change with your update?

Woman 4: That's a great question. So, I have been involved with the breast technical expert channel that's been working on the (unintelligible) measure that we developed to replace this. There was some extensive discussion about what is the time - the appropriate timeframe for administering Trastuzumab? And my understanding - and I'm not a clinician - but my understanding is that its variable based on the chemotherapy regimen.

And so, essentially the panel felt that with the replacement measure, that we should focus on appropriateness of treatment rather than timeliness of treatment. And so, because there was a lot of variability with potential chemotherapy regimen, we didn't want to kind of, artificially restrict the timeframe for each regimen because that may vary. The timing may vary based on the regiment. And so, I think we, sort of, deferred for a wider-timeframe to allow for potential variation in chemotherapy regimen.

Ava Clarke: (Dr. Bungi).

(Dr. Bungi): I do have two questions. One is, how is the neoadjuvant Trastuzumab accounted for in this model? And number two is - for some of the other measures, there's an age cutoff. And, you know, as we have older patients (the competing cause mortality) would take over the cancer mortality - why this one does not have an age cutoff?

Dave Ryan: So, I can't answer the last question. I think that that - I wasn't involved in the discussions and I'm not quite sure if that came up during their discussions for the age cutoff, so, I can't answer that one.

In terms of the first question, for neoadjuvant chemotherapy - so, Herceptin is still standard post-surgical resection and so it is still being captured in that patient population, despite the use of it being a neoadjuvant.

I believe there is some discussion at the task force around whether or not we should be capturing neoadjuvant in addition to adjuvant, meaning post-surgical resection.

At the present time, given the length of time people are on Herceptin, they're getting captured anyway and it was felt that they should continue as is.

Woman 4: With respect to the age cap, that was another thing that came up in discussion with the breast technical expert panel, simply for consistency, because this measure is in use in this, we decided to leave it as is for now.

But the replacement measure that we do have, I believe, caps patient's 18 to 70 and I think that's following NCCN Guidelines. So, we are looking to address that, but for consistency, we tried to leave it at this for now…

((Crosstalk))

Ava Clarke: So, with regard to evidence is there any more discussion related to the quality of evidence? Or is the team ready to vote now? So, I guess what we'd like is to call the vote for Evidence.

(Hannah): Voting is now open for Evidence on Measure 1858. Options are A for High. B for Moderate. C for Low and D for Insufficient. Voting is now closed for Evidence on Measure 1858. We have 12 votes for High. 5 votes for Moderate. 0 votes for Low and 0 votes for Insufficient. This Measure passes on Evidence.

Ava Clarke: Len, do you want to take us back to discuss the Gap - the Evidence of a Performance Gap.

Len Lichtenfeld: So, the performance gap - and let me just bring it up here if you don't mind, while we're talking - I don't want to go completely from memory - I will - just in the interest of time, it'll take me too long to find the spot.

The concurrence with the measure is about 97% - 97.1%, if I remember, which is pretty high. That leaves a Gap of 2.9 and approximately 3%, which in the world I live in is outstanding. You know, that's basically complete compliance for the reasons I already mentioned, my thoughts on that regard.

So, from my perspective, this is a measure that has achieved its goal. Whether - and I would like to hear some discussion about what the various options may be. Do I think it's an important measure? I think that's really the question. Do we need to have a measure in place in order to ensure that we stay at that high level, again pending another measure? And I don't have an answer to that.

I think that there's no question that this is a standard of care with understanding the various permutations. But this is standard of care. It's important. It makes a real difference in outcomes to women who are HER 2 positive. It's a huge impact on cancer mortality - breast cancer mortality, which has actually dropped 40% from 1991 to the present time. And frankly I think some of that - a good deal of it - has to do with this the research reported in the early 2000s, so. I don't have any good answers, but I do want to hear what others think.

Ava Clarke: So, the mics are open for discussions about importance (of the) measure. If not, I'll go first. I'm a clinician who treats breast cancer patients, that for disclosure, I guess. And so, I agree that the performance - overall performance standard of the clinicians that are reporting on this measure is very high.

But there's a large body of clinicians that aren't using this measure and reporting on this measure. And so, the developers also cited evidence from the literature that there still remains a broad variability that's obviously evidence from the literature, but the range was surprising, 9% to 100%.

Even though this has been a standard of care for many years. And so, given that kind of data, I would assume that although it's the standard of care, there's still broad uptake in that - there's surprising data clinically since - as Len indicated - the outcomes in breast cancer improved dramatically with the addition of these drugs for HER 2 positive patients.

So, I would tend to recognize that if there's still among the - on the basis of literature abroad - variability. I think it still remains important to keep this as an option because we don't have a standard of care.

I also am happy to hear that there's the next measure coming. Because how we use Herceptin, how we approach patients with HER 2 positive disease, has changed dramatically over the last 5 to 10 years and not recognizing that and creating another measure to look at the role of other adjuvants like (Trastuzumab) et cetera, timing of administration is critical and the standard of care continues to evolve.

I do think that this measure captures that Trastuzumab needs to be part of all of the therapies, whether it's before or after the initiation of (adjuvant

therapy). And so, I'll stop and endorse that I think that the Gap persists from the medical literature.

Woman 5: Thank you (Unintelligible). I'd say, just as a patient advocate and cancer survivor myself, just statistics alone to speak and so, just a positive aspect for this measure. So, I'm totally in agreement with this just as a cancer survivor. The ability for something rather than nothing.

Man 2: Yes, this is coming from my hat as a Clinical Director of a large system, at Cancer System, that I don't think that there's an oncologist out there who doesn't know that if a woman is HER 2 positive, that you're supposed to use Trastuzumab.

I do think that - and this is just a theory from watching this at our own institution and other institutions and having been on the measures committee, that this is a Systems' issue, where the physician doesn't know in a timely fashion, that the patient is HER 2 positive because the surgery or the pathology was read outside of the system or practice, that that oncologist is in. And it's a matter of getting that information and creating that system.

I think this is an important measure because it drives home that point to everybody that from an administrative perspective, that the records have - when you're treating somebody with cancer, it is your obligation to fix whatever issues you have in getting those medical records into your chart, whether it be electronic or still paper.

And I think, when we see gaps - when you see gaps in the literature compared to the gaps in folks who are in our measure environment, I think it's probably largely due to system issues as opposed to physicians not practicing according to the guidelines.

Woman 6: So, one part of this analysis report actually caught my eye. an excess of 250-unit NPIs indicated results of the team level analysis. A majority of NPIs are small denominators and all at 100%. What about the minority? And so, how this measure should be performed? Should it be based on the volumes? And low performer of the providers, meaning NPIs, not at the (team level)?

Woman 4: Yes, that's a good question. We did have some limitations in terms of the data availability. And so, because this is an older measure, we had to essentially get mixed performance data, I believe from CMS.

And so, we sort of did the analysis that we were able to do with the data. There was no question about the granularity of the data and whether we could do NPI analysis. So, that's a valid question. I'm not sure that the data that we have is sort of granular enough to give more robust...

Woman 6: The reason I bring it up is the 97% performance may be masking the fact that there are providers out there that are not delivering to this standard.

Woman 4: I would agree with that.

Len Lichtenfeld: So, I have to share - if you read - I'm just trying to figure out how we - depending on what decision we make - memorialize the discussion? The analysis by staff says - offers the 97% concordance number, right? And then it says the measure is topped out. I mean, that's what it says.

And the discussion around the table is the various - the variability, or why that variability exists. And I don't disagree - I mean, obviously what you - you reflected what's in the evidence document.

So, then the question becomes, depending on how this committed votes - let's hypothetically say the committee says a significant gap persists, the measure is not topped out, there's a variance with summary document and the recommendation - how does one memorialize that concern going forward? Because then there should be a record that the committee did not agree, specifically why the committee did not agree. I'm assuming there's a summary that's provided.

Woman 7: So, we would share - publicly we share our measure worksheet, that gives, kind of, the initial staff rating. But then once the committee votes, it would show clearly within the votes that your votes would be different than the staff rating as an example. And then we do provide a summary of this discussion. Does that answer your question?

Len Lichtenfeld: As long as - that commentary is critically important because I think there's - these are all public records. People can look at them. They should know why we did what we did. You know, we have two competing interpretations. One based on the information that's available and measured and monitored and the other based on literature to suggest that it's not the case.

That then becomes the dynamic that has to be reflected going forward I would think. Otherwise, people would say you're just - the risk in any measure is that you show such good performance, that people are going to ask the question - people who are trying to respond to these measures are going to ask the question - why do I have to keep on doing this if everybody's doing so well.

And it has to be very, very clear that there's a need in the eyes of the committee that everyone is not doing very, very well, notwithstanding some of the number that are top line, they're standing out here - on

recommendations that have been made. I think it's a - just a - shall we say not a burden, but I think it's important, let's put it that way.

(April): Yes, and this is April. I mean, I think we ought figure a meeting somewhere in the (unintelligible) and I think that is the place where we could sort of memorialize this discussion to make it clear that there was lots of discussion about what the variability of the performance gap actually is and leave it at that.

Ava Clarke Any further comments? Any members from the (unintelligible)? Then I think we should move forward with voting on the gap.

Hannah: Voting is now open for Performance Gap on Measure 1858. Options are A for High. B for Moderate. C for Low and D for Insufficient. Voting is now closed for Performance Gap on Measure 1858. We have 0 votes for High, 12 votes for Moderate, 5 votes for Low and 0 votes for Insufficient. This Measure passes on Performance Gap.

Ava Clarke: So, let's move onto discussing, Reliability.

Len Lichtenfeld: So, I mentioned before, the issue on Reliability is, I guess what the level of measurement, whether it's facility or group practice. And reliability was considered insufficient by staff. I would like to hear, perhaps a bit - so I understand a little bit more about what differences we're really talking about here. I personally struggled with that a little bit. Maybe I shouldn't have, but I did and I'll share that with the committee, so.

Woman 4: Yes, (unintelligible) we also (unintelligible) had a question about that. It did seem like we provided the same, sort of, type of testing for a number of our measures and there seemed like there were some inconsistency with NQF

Staff evaluation of that. For example, for the colorectal measures, there was one where the (unintelligible) liability question came up and another where it did not.

So, we did provide some clarification back to NQF, I believe towards the end of January. We think, honestly, that this is maybe a semantic difference. In some of our description of the testing that was used, we did use the word facility, but the data that we used was essentially mixed data that is collected at the group and practice level. So, that's why we indicated that the testing was done at the group and practice level.

I think we did try to clarify and update our form so at Section 1.5 we emphasized the practices were identified by (least number of TINS) - Section 1.6 we talked about an (NPI) level analysis.

So, I think it may simply be a semantic issue where we, you know, essentially the contractor that we used to conduct some of our data analytics used the term facility in describing some of the tests that were done. But we do feel that the testing was done at the group and practice level.

Len Lichtenfeld: If I may, for some of the people around the table who may not understand the lingo, do you want to explain what a TIN is?

Woman 4: NPI National Provider Identifier, I believe. And TIN, Tax ID Number for Practice.

Len Lichtenfeld: So, a TIN could be - what kind of - if a TIN level, can you share examples of what a TIN might be? I think I know the answer, but I'd rather hear it from you all.

Woman 4: So, I believe, correct me if I'm wrong please, but I believe that there may be multiple physicians within a TIN. And I think that larger systems may also have multiple practices that have their individual TIN - Tax ID Numbers.

Len Lichtenfeld: That's - then I'm not sure that I understand. Because a TIN from my perspective is a single Tax Identification Number for an entity, which basically the entity can be a large institution. It can be a whole entire system. Or it might be an individual practice.

And it's not, most institute - unless I'm wrong and I don't know the answer - most large institutions, I don't think have separate TINS for each practice. So, there's a lot of stuff that gets rolled up into a TIN that would be different than it might be in Len Lichtenfeld's medical oncology practice, of six docs somewhere. There's variability - am I correct on that?

And I don't know - so that's where the level of size, you know how much you're reporting under a single number and how that reflects on quality of the individual provider?

Woman 8: I agree with that.

Woman 4: There's a point of clarification from in separate I believe that the mixed TINS may just be (unintelligible) physicians rather than workers (unintelligible).

Ava Clarke: So, I think the question at hand for the committee is that, although there's the disparity in language, is there a disparity in how we interpret the results of reliability? Because the data that was provided was high level reliability and we've heard some discussion from the developers about how to clarify those disparities. But I think the committee needs to feel comfortable with those clarifications.

And I would want to ask the staff one question. Is this reflective of the change in the standards from the last evaluation period to this evaluation period?

Nicole: No, I don't think so. We just were simply reviewing the measure based on certainly what the developer gave us and brought that question to them for clarification, which I think they provided. So, there's no change in our…

((Crosstalk))

Len Lichtenfeld: I didn't hear - from Nicole - I couldn't hear what you were saying.

Nicole: All right. The question was whether there was any change in our criteria which may have prompted a different rating on reliability for this measure versus the last time it was reviewed?

And I had informed the group that there was no change in our criteria. That we did follow up with the developer to seek clarification and they've also provided some of the clarification verbally about the testing…

((Crosstalk))

(Craig Buchell): Hello, this is (Craig Buchell) on the phone. I didn't know how to indicate that I wanted to speak, so I tried raising my hand on the teleconference, but I don't know if anybody notices that? Can I barge in?

Ava Clarke: And for the group, we can't see that, so if - feel free to tell us you're there. Thank you (Craig) for that.

(Craig Buchell): Okay, sorry about that. I would just say that my reading of what has been submitted and the analysis of the staff is that the insufficient rating is based on the developer submitting information that used the term facility to describe the testing, when in fact, the testing wasn't done at the facility level and I think they said that their contractor had used that terminology.

And I would tend to agree with the analysis that the reliability rating is actually fine and that the preliminary rating for Reliability was just based on - I don't want to say typographic, but inappropriate use of terminology in what was submitted, not reflecting what was actually done. Thank you.

Ava Clarke: So, those of you on the phone, you can continue to use the raise your hand function on the conference call and NQF staff members will monitor that and let us know if you have - you can also pipe up and talk, but if you want to use that we will - somebody will monitor that and let us know. Are there - oh

Woman 10: I think this discussion is interesting from the perspective of what we are trying to measure. We are trying to measure individual physicians delivering the right shares. And if the granularity for measuring that the NPI (level) is not there, then this measure is really not reliable. I am not measuring their group practice. I think the intent is to make sure every individual physician is delivering the right shares.

Distinction is important. The reliability is not good for NPI level. Perhaps good for the TIN level.

Woman 4: We tried to anticipate that our data might not be granular enough for that and that's why we actually changed. I believe when this was originally endorsed, it was endorsed by the individual level and so we changed that to a group

practice level endorsement to try to reflect the granularity of the data that we had available to us.

Ava Clarke: My question then for the staff is, then what we're comparing is the way the measure's written and what the intend to endorse, rather than something that was - that there was a change in the measure basically.

(Nicole): That is correct. But since the developer is here to provide clarification, that should be to help inform your impression and vote.

Len Lichtenfeld: So, that raises an important question. Did you say that the original intent back in 2012 was that this be reported through NPI - through individual practitioner? And now we're coming to - and then - and I'll -pre-state my next question - I understand what you're saying. I'm with you - I get it.

I don't know - and the question - the other secondary question would be does - so we have the - I hate to say it this way - so the perfect to the enemy of the good. Namely, are we capable of looking at - reporting at an NPI level instead of at a facility level or a group level? So, the points well taken. I don't know what the practicalities are. So, two questions.

Woman 4: Yes, so I think - I wasn't an actor when the measure was originally developed back in 2012. I do believe that probably the intent was to assess individual clinician practice. It is my understanding that the measure was previously endorsed at the individual level.

I don't believe that reliability testing was done on the previous submission. So, we did do updated testing and just because of the granularity of the data we did, obviously tried to anticipate some of this and changed the endorsement to the group and level practice.

I think within this, you probably see a combination of reporting. There are individual reporters, there are also group reporting. And so, virtually it's hard for us to…

Len Lichtenfeld: So, I think that's an important point. If the (MIPS) 0 - under (MIPS) standards, it can be done either way. It's not required that it be done at the individual or the group practice level. It can be a choice by the reporting entity. Is that correct?

Woman 4: That's my understanding.

Len Lichtenfeld: So, technically, this would meet that with all within that…

Woman 4: Yes, and we did use mixed data for…

((Crosstalk))

Ava Clarke: Any other questions, or comments, or comments from the members that are on the teleconference? Is the committee ready to move to a vote?

Hannah: Voting is now open for Measure 1858. We are now voting on Reliability. Options are A for High, B for Moderate, C for Low and D for Insufficient.

Len Lichtenfeld: I'm having technical difficulties. I've got to reboot. So, my vote won't come in. I got it. I'm good.

((Crosstalk))

Hannah: Voting is now closed for Measure 1858 on Reliability. We have 1 vote for High, 13 votes for Moderate, 3 votes for Low, and 0 for insufficient. This measure passes on Reliability.

Ava Clarke: So, Dr. Lichtenstein - or Lichtenfeld - can you send us through the validity discussion, there's several components that we need to discuss in validity, but some of them are similar to the last discussion.

Len Lichtenfeld: Well you're obviously prepping me here because I'm bringing it up and I just remember that the staff thought the validity was high if I remember correctly. But I'll - unsure which - they call that individual steps because there was an overarching question about whether or not - let me just go down to the point. Bear with me for a minute as I (unintelligible).

Yes, I'm not sure what level your concerned - there was a - the results for establishing validity, the correlation was strong. I guess I'm not - I'm not following what you're asking because it says the overall rating was high. That was the staff recommendation.

Ava Clarke: Yes, I think maybe the issue was the gap between Reliability and Validity and if there were any threats or concerns to the Validity.

Len Lichtenfeld: Well there were no concerns about the validity. I mean, the comments from the committee were all - in fact on Reliability, there was - but we've already had that conversation. But on Validity, I don't see any - I don't see the threats to Validity. But those are pretty much the standard, you know, somebody may cover that paper. I'm not following where you're headed.

Ava Clarke: The gap. I mean the different - and validity from what we were presented in part of our discussion. I don't think there's any threats to Validity either. I just wanted you to lead the discussion.

Len Lichtenfeld: I - I'm not following you. I mean, I was prepared to say that - yes, I don't see a problem with it. So, am I missing - I'm always worried am I missing something - are you trying to send me a subtle message of some kind to…?

Ava Clarke: I'm trying to make sure we have a discussion about Validity, as a group and get feedback from the committee. That's all. So, I apologize.

Len Lichtenfeld: That's okay. I - this is where the validity was high staff rated and high. The measure developer performed - there's weren't any outstanding - and the comments were generally all consistent with that.

Ava Clarke: Is there comment on the phone? All right, no comments. Any other discussion from the group? Are we prepared to vote for validity?

Hannah: Voting is now open for Measure 1858 on Validity. Options are A for High/ B for Moderate; C for Low, and D for Insufficient. Voting is now closed for Validity for Measure 1858. We have 14 votes for High; 3 votes for Moderate, 0 for Low, and 0 for Insufficient. This Measure passes on Validity.

Ava Clarke: We have a new member that joined us, Jette, and we need to do our disclosure. So.

Jette Hogenmiller: My apology, my information says 9 AM. In writing too.

Ava Clarke: Well we apologize for that. But what we've been asked for us to do is sort of give your name, where you're from and then if you have any Disclosures of Interest. Can you use your mic?

Jette Hogenmiller: Oh, my apology. Oh, I pressed (unintelligible). Anyway. Jette Hogenmiller, Denver, Colorado. No disclosures. But there's a lot of snow, come skiing.

Ava Clarke: Thank you and apologize for the miscommunication on the date - on the time. So, it looks like the Measure passed. So, the last, we have a few other areas to consider. The next one is Feasibility. Any comments from the reviewers?

Len Lichtenfeld: The consensus from staff was preliminary rating was Moderate on Feasibility. And one of the questions about Feasibility is which is required date and what is not routinely generated (unintelligible) during care. Which are not available on (unintelligible) format? The general consensus is that feasibility is routine. The information is available.

I was a little bit surprised to see people say it's not easily available in HER. And those who deal with the every day may have a different experience. One would think that this is one element that would be - would be available. But I don't think that's (a fundamental issue). This is not an E measure at this time, so it's important, but not necessarily relevant. There were not concerns - significant concerns regarding Feasibility raised by the committed.

Woman 11: So, when you get to this data for the measure developers, how do you look at the delivery of the Trastuzumab per se? If the patient is seeing a practice and treated in that practice, obviously that's visible - what about if the patient is seeing one practice for an opinion and treated at another practice? How is that physician's record is going to be viewed?

Dave Ryan: So, it's not included in the analysis. So, it's by the treating practice or treating physician.

Ava Clarke: And is that satisfactory?

Woman 11: It is. But that creates a feasibility issue. So, what do we know about who has the patient? Who is the treating physician? How is that recorded?

Dave Ryan: I believe and I could be wrong about this - at the logistical level, I believe it's based on if the practice is administering the Trastuzumab in their infusion unit. So, if the practice isn't administering Trastuzumab in their infusion unit, and the patient is being seen by somebody else, then it's just a consult and it's not - it's coded by the coders as not - that they're not eligible for this particular measure.

Woman 11: In other words, this requires some level of abstraction. Because whether we are seeing the patient for consult or for treatment is not a - is something that's usually embedded in our documentation. Not a unique field.

Dave Ryan: Yes, but there is no way - other than some level of abstraction - there is no way to get at that issue of consults.

Len Lichtenfeld: I'm having a bit of difficulty scrolling through all the documents that we've been provided. There are exclusions on this - on the definition of - and it depends what - the other question is what's included in those - included in those exclusions.

So, I think the question then, could devolve, that based on the exclusions does require manual abstraction to get there. But if exclusions are stated, then - but

I don't know if it's strictly for consultation only. It did say people who come into the practice after they've been started on chemotherapy - I think that's in there, but I'm here scrolling to try to find it. Maybe you have a better familiarity then I do.

Woman 4: Yes, so just (unintelligible) that we do have a denominator exclusion for patients who transferred to the practice - to the reporting practice after the initiation of chemotherapy. So, if their treatment has already started somewhere else, they are meant to be excluded from the measure.

And, you know, we do - we do acknowledge that this is a registry measure. It is not an (EQM) so there may be some degree depending on this reporting practice - some degree of abstraction required anecdotally, because the measure is in use and mix, we haven't heard problems of abstracting or reporting the measure. But, of course, there is always the potential that there is some degree of manual review required.

Jodi Maranchie: This is Jodi Maranchie, sorry for the delay here. I was just wondering if a person was referred after one year, would that penalize the person who eventually gave the Trastuzumab?

Ava Clarke: I assume that's a question for the developers.

Woman 4: And I think - at the (unintelligible) the question is a good one. I think the intent of the measure is to capture Trastuzumab's administration along with the adjuvant treatment course. So, technically the measure is only looking at the 12-month reporting period. Afterwards, they eventually would not be included in the measure.

Ava Clarke: Len, did you have a comment?

Len Lichtenfeld: So, a situation could theoretically develop - theoretically develop where someone is seen in a community facility and they decide to get another opinion 13-months after their diagnosis. They go to the Cancer Center - recognized cancer center, who decides to redo the pathology and converts it from a HER 2 negative to a HER 2 positive, maybe a quality defect.

So, how that person is HER 2 positive and I - that's a problem with Karen and y'all have to deal with. But let's assume that if you discovered that, in fact, original test was in error, they would then take - they would then be eligible to get receptive, I guess. I don't know what you would do.

Having said that, that person would not be included in the denominator. They would be excluded by virtue - you know, one of the things that comes up - and we were talking a little bit earlier about all the committees that we sit on - I've sat on a couple of them where the more guys that were memorializing some of these questions - it is important for people to understand how a measure can be applied.

I think that's another good question about what would happen in the scenario where somebody is identified at a later date to be (in fact) a HER 2 positive or it's after the 12-month reporting period, that they are not part of this measure. So, there's a body of information that people can rely on through interpretations. That happens with standards, it happens with payments, it happens with a lot of things. It would be helpful (here) as well.

Ava Clarke: Steven.

Steve Chen: Yes, Steve Chen. As far as for the the cut-of-edge cases, I mean, I think they're all covered under the denominator exception which, if anything, is probably too broad as you go to treat your successor Quality Measure.

I mean, the denominator exception basically is if you write down any reason for not giving Trastuzumab, you're excluded. It doesn't say - they give some examples, but you could literally write almost anything and be excluded. And I think that would be - in all honesty, my only concern is that someone wrote down some incredibly specious reason like, I don't believe Trastuzumab works.

In theory, they could except themselves from the denominator. I don't think anyone believes that. It's something I'd recommend to you as you develop your successors to take that up.

Ava Clarke: And then, is there someone online? Someone online?

Greg: This is Greg. I guess my point, with regard to feasibility, is that you know, of course feasibility is speculative for new measures. But for measures that come up for endorsement like this, where people are already reporting the measure, like a decent number of people or whatnot - I think that speaks to the fact that it has a certain baseline level of feasibility for the people who are reporting.

And, you know, at least for me, takes away some of the concerns. I mean, there may be particular people who have difficulties with it, but it's certainly plausible. Thanks.

Ava Clarke: Thank you. Any other questions or comments? Or are we ready to vote? Vote. Sorry…

((Crosstalk))

Hannah: Voting is open for Measure 1858 on Feasibility. Options are A for High, B for Moderate, C for Low, and D for Insufficient. Voting is now closed for Feasibility on Measure 1858. We have 10 votes for High, 7 votes for Moderate, 0 for Low and 0 for Insufficient. This Measure passes on Feasibility.

Ava Clarke: And then onto Usability. We want to focus on Use. Len, do you want to lead that?

((Crosstalk))

Len Lichtenfeld: You know, I'm scrolling back through it. You know, we have a couple different forms here going back-and-forth. We're almost there folks. I should have followed - Lenore pointed out earlier, but she printed out all of her things, unfortunately I was traveling and I couldn't print them out. That would make life a little bit easier for me wouldn't it?

Thank you very much, I appreciate it. I'm almost there. So, Usability. This is a Measure that is not currently publicly reported. I believe there was a comment that I read in some of this - that, that may change in time, that it may become public reported, but not currently.

It's used in several different measures. We've talked about (MIPS) as part of (COPY) so, that is further indication of usability. And, the rating that was offered by the committee was that the Usability should pass. I don't recall - and let me just double check - there was no concern raised about Usability and

frankly in terms of benefits versus harms risks, it was all basically positive statements.

The one question that came up again, was the issue about being topped out? But we've already discussed that.

Ava Clarke: Dr. (Barse).

(Dr. Barse): I go back to the developer's rationale, which is stated as (unintelligible) envision the use of this measure willing to (concordance) with recommendation (unintelligible) by the administration for the (station) population.

And I, as much as I do agree that the evidence for (unintelligible) exists. I don't think this measure is Usable in terms of improving the concordance because of the issues that we discussed that you cannot get to the (level) of the provider and you know, how are we going to use this measurement to make any difference in concordance and that, you know, MIPS is not (a teachable) to everybody. It's only a (subset) of practices that are part of MIPS.

So, I think this measure, as much as it's needed and meets many criteria, it's not going to be a usable measure. For the intent that the developer had stated.

Woman 4: Second, in terms of the individual physician performance, I think we do have some of that data. We have that in the MIPS data. Some of that is group practice. Some of that is individual reporting. We don't have the granularity that we wanted to give out exactly the difference between who reported as an individual, and who reported as a group.

I think it's a bit of a philosophical question as to whether - do measures actually change physician behavior and improve concordance with recommendations? I hope - yes.

Dave Ryan: I would come back to the statement I made about systems and I would use an example. In my world in colorectal cancer, where it became a part of measures to look at 12 lymph nodes in the specimen, it was clear that the medical oncologist wasn't the one leaving out whether there were 12 lymph nodes present in a resection specimen. But the medical oncologist has a tremendous amount of influence on the delivery of care within their system.

I think, yes, I agree with you at the level, because we can't look at this routinely, I should say, or consistently at the level of the individual practitioner, doesn't mean that this isn't a useful measure from a system standpoint. It's clear that there is still variability within the United States as to patients getting Trastuzumab who are HER 2 positive.

We believe it's a systems issue and leaving that measure in there is incredibly important at making that point and having the medical oncology community, in particular, I would say followed closely by the surgical oncology community - making this case to the individual systems, which - individual practitioners, increasingly are subject to the systems they are in - making the case that adequate recording of this measure so that appropriate actions can be taken - is incredibly important.

So, from my viewpoint, speaking as an individual, that's the reason why this measure is still an incredibly important measure.

Woman 13: I don't disagree with you. It's important. I think the rationale needs to be stated differently if it is a system issue. It is not really to improve

concordance with recommendation. It's to improve the access to this care at a population level. These are two different things.

Ava Clarke: Steven.

Steven: Yes, so I guess from my perspective, ironically because it is so close to being topped out, I worry less about this, sort of, individual clinician accountability. Because even a medium sized practice or a large practice, if they have any one clinician that's falling down, they're going to drop them from the 90th percentile to the 10th percentile with just a few cases. Which has really been the problem with topped out things is that the compression of the range.

And so, I feel very confident that any reasonable practice that is hitting 91% is going to be scrambling to figure out why they're at 91 and not at 99 if the median is 100.

That even the one at 97% is going to be scrambling because it's going to be such a DRASTIC change in their MIPS criteria. And so, ironically because it is so close to being topped out, I have virtually no worry about reporting (unintelligible) practice or site level.

Ava Clarke: Len, did you have a question?

Len Lichtenfeld: You know, I forgot to take it down. But I've been here mulling something that from my perspective, forward thinking a little bit, it's not just the physician, it's not just the system, it's also the patient.

And I think that in this particular measure, it's, you know, I think it's important to set down a marker that I think as our system moves forward, patients are

going to be monitored along the best of autonomic care. But that's a longer discussion for another day. It's more theoretical.

So, I don't think we can leave out the fact that patient's family's need assurances as well that the right things are being done. I think it's part of a broader discussion.

Ava Clarke: And then I think there's Greg on the phone.

Greg: Yes, I guess on the question of Usability and Use. You know in the NQF criteria, it's stated that, you know, one of the requirements for, you know, continued endorsement of measures was, of course, the use and accountability programs. But also, the public reporting.

And so, they described their, you know, plan for public reporting. I guess this is more of a question for the NQF staff. You know, the criteria is that, you know, after endorsement it should be public reported in six years.

Since to some extent, you know, the - unless the developer is collecting the data and publicly reporting it themselves, they're kind of subject to the whims of the people who may be able to publicly report it. Is that a deal breaker for us? Because, as I said, they have a plan to do it. It just - it hasn't been done yet. And yet the NQF criteria are that it should have been done for endorsement - continued endorsement at this point.

Nicole: Yes, this is Nicole. So, I think, you know, the challenge is that it is being used in the Accountability programs and so, just to be perfectly honest, I don't know if we have a great answer to that. But that is certainly a requirement, as we look at a Maintenance Measure for re-endorsement, that it is publicly

report. But then also taking into account that it is part of the big Accountability program.

April: Yes, this is April. Nicole's right that it is definitely part of our criteria. I will say I've been in several of these meetings and I think Standing Committees have various thoughts on whether that's a criterion that should continue. And so, they've had various conversations about, you know, Accountability versus Public Reporting versus (unintelligible) cases, even Quality Improvement Measures that may be not publicly reported but are used by a large number.

And so, I think what we've allowed them to do is have that kind of conversation. And so, I think it sounds like you guys are having that conversation. I think if you feel like it's then an Accountability Program, it's really important to measure I think, that you should sort of think about how that - how you vote, as opposed to always making it very structure-like criteria.

Ava Clarke: Okay. Any other questions or comments, or should we move to votes? Go ahead.

Hannah: Voting is now open for Measure 1858 on Use. Options are A for Pass, B for No Pass. Voting is now closed for Use, Measure 1858. We have 17 votes for Pass, 1 vote for No Pass. This Measure passes on Use.

Voting is now open for Usability on Measure 1858. Options are A for High, B for Moderate, C for Low and D for Insufficient. Voting is now closed for Usability on Measure 1858. We have 2 votes for High, 15 votes for Moderate, 0 votes for Low and 1 for Insufficient. This Measure passes on Usability.

Ava Clarke: And I think we have one more vote, which will be Overall Suitability or - of the Measure. So, let's just move to that.

Hannah: Voting is now open for Overall Suitability for endorsement for Measure 1858. We have A for Yes, B for No. Voting is now closed for Overall Suitability for endorsement for Measure 1858. We have 18 votes for Yes, 0 votes for No. This Measure passes.

Ava Clarke: Thank you. And we'll move onto the discussion, so. All right, so one down and eight to go. We did that in a little over an hour, so - we have time for a meeting next week on the phone if we don't get through. So, we're going to try really hard to get through it. So, Measure 1859 is (KRAS) Gene Mutation Testing Performed for Patients with Metastatic Colorectal Cancer Who Received Anti EGFR Monoclonal Antibody Therapy.

So, Greg is our Lead Discussant. On the phone we also have Brad, (Lenore), Len, Jodi, and Danielle as our additional discussants and I hope that everyone will feel free to chime in as well. But Greg, do you want to start us - oh, I'm sorry, we start with the Developers giving us a couple minutes and then we'll turn to Greg.

Dave Ryan: So, this Measure - I think it's worth going through the Numerator and Denominator quickly because we have another Measure that is similar to it, coming up on the heels of it. So, the description of this is that it's a percentage of adult patients 18 and over with metastatic colorectal cancer who received Anti-Epidermal Growth Factor Receptor Monoclonal Antibody Therapy from whom RAS, either KRAS or NRAS gene mutation testing was performed.

The denominator here is adult patients with metastatic colorectal cancer who received Anti EGFR Monoclonal Antibody Therapy. There are no

denominator exclusions. The numerator is that RAS gene mutation testing performed prior to initiation of Anti EGFR Monoclonal Antibody Therapy.

So, we think this matters because, the process being evaluated in this measure the completion of RAS testing to identify those patients who will not benefit from anti EGFR therapy. Evidence now supports expanded RAS testing, including testing NRAS in addition to KRAS mutations.

Clinical trial data shows that the benefit of using EGFR inhibitors and treating metastatic colorectal cancer, either as monotherapy or in combination with other treatments, is limited to nonexistent and actually recent data to suggest even harmful in patients with RAS mutated tumors.

This data strongly suggests that patients with RAS mutations are better served with other therapies, especially considering the harms and costs of anti EGFR treatment. Additionally, this measure is directly supported by recommendations in the American Society of Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, the American Society of Clinical Oncology, and the NCCN Clinical Practice Guidelines.

This Measure - as I stated earlier, is related to another measure undergoing re-endorsement, Number 1860, Patients with Metastatic Colorectal Cancer and RAS Gene Mutation Spare Treatment with Anti-Epidermal Growth Factor Receptor Monoclonal Antibodies.

Ask of use of relationship between the quality actions of this Measure and the patient's outcome are the following: First, the target population is identified, namely patients with metastatic colorectal cancer. The process of RAS mutation testing is evaluated in that and appropriate administration of monoclonal antibody therapy targeted agents.

The outcome is that patients are spared toxicity associated with contraindicated treatments. And finally, the outcome - that intermediate outcome is that - and the overall outcome is that you have improved quality of life, improved progression free survival, improved overall survival, improved response rate, and reduced resource utilization costs.

Ava Clarke: Great. Thank you very much. So, Greg, do you want to kick us off with the - as the Lead Discussant.

Greg: Sure. One sec. Okay right. So, this is Measure 1859 as we discussed. It's a Measure that was originally endorsed in 2012, that was actually the most recent date of endorsement was 2012. It's level of analysis is clinician or the group practice as a process measure. And so, we're looking at it for maintenance of endorsement.

And so with regard to the evidence, in fact the evidence has evolved somewhat since the time of the initial endorsement and as mentioned, that is reflected in the American Society for Clinical Oncology, College of American Pathologists, Association for Molecular Pathology, and American Society of Clinical Oncology Practice Guideline that was formed after a systematic review of the evidence. And it's also supported by the NCCN Guidelines.

So, that's new evidence. Of course, since it was previously endorsed, we know that it previously passed that criteria. Now, interestingly the preliminary reading for evidence by the staff was low and to be honest, I guess I would personally disagree with that. I thought actually, the evidence was at least moderate.

And you might even be able to argue that it was high quality evidence. But in any case, I think we'll have some discussion about that and determine whether or not other people feel that way.

You know, in terms of the premeeting comments, there was some concern that maybe the measure doesn't exactly align with the evidence. And, you know, the evidence might not be the highest quality. So, maybe that was reflected in the preliminary staff rating. But again, once we open up discussion of evidence, I think people can weigh in.

Overall though, the comments I think were favorable. And of course, this is something that had, you know, passed the evidence criteria previously and since that time, the evidence certainly hasn't diluted that information. So, as I said, I think that overall the information that we have is favorable.

Now with regard to opportunity for improvement, you know, registry data provided from CMS showed a persistent Gap. You know, there wasn't specific evidence presented of disparities, but there was a (SERE) study that suggested that there may be some disparities.

But I think it was - I mean there's overall consensus already in the comments and the preliminary analysis that a Gap exists. So, I think our discussions actually will probably be quite short.

In terms of Reliability, you know, previously…

Shelley: Hey Greg, Greg…

Greg: Yes.

Shelley: This is Shelley. Can - let's just - let's talk about the evidence of the Performance Gap before we move on. So, if you could hold that. I want to go back to evidence. Would somebody from staff want to talk about want to talk about why staff rated it at low?

Woman 14: Sure. And I think this was addressed a little bit by the developer, but essentially the evidence that we reviewed, we noted that it was based on expert opinion and felt that it was about the EGFR inhibitors as opposed to the RAS mutation status. The other thing - and so looking at that in relation to the algorithm that we follow when we choose our ratings, that's how we got to a low rating for evidence.

But I think the developer's comments at the beginning addressed some of the challenges that were noted in the - you spoke a little bit about the Numerator and Denominator. So, that helps (give) that position.

Woman 15: I think this Measure can be very confusing if the context - the right context is not given to it. So, there are drugs that we can only use if we have the results of this testing and if the RAS testing is defined as (unintelligible), let's put it this way. The question you're answering is what is the evidence for testing? And I think that's where the discrepancy comes in.

And we obviously have high level of evidence for testing, because otherwise we will not have the ability to use the drug. I don't see any other way. I actually think - I think the problem is how this measure is positioned. It's not a direct measure independent of the downstream of this measurement - or RAS measurement, I should say, not the Quality Measure.

And I think that's where the confusion comes in, in terms of assessing and addressing the Gaps and the need for such measurement. That's what I think.

I think the problem is if it convoluted the (space) on how the evidence is (looked) at may come across different.

Karen: So, I agree. I think that the problem with interpreting the data is, we have the neg - we know the negative data - if we know that giving these drugs in patients that have the mutation doesn't work. So, conversely, we need to screen to find out who shouldn't get the drug, as much as who should get the drug.

And so, I think that I agree that the indirect - this represents indirect data evidence, whereas when we see the next measure, it's more correlated with direct evidence. So, I think that's where the confusion for having to rate it is. But I don't think that diminishes the importance of the connection between testing and appropriate drug utilization.

Woman 15: So, it's a negative predictive measure, basically? Or test?

Woman 16: Can I ask a question just from a clinical practice perspective? I mean, don't you have to do the testing in order to get the drug approved for insurance? So, is that - I guess that's where I wondered - maybe it's like if insurance becomes a - the Quality Measure because we can't get it if you don't - right? So, I don't know how that relates? I mean, maybe that comes more into play with the Gap and it's not the Evidence.

Woman 15: And so, I think the issue of the insurance is more relevant to the next Measure, which is (1860) which is delivery of the treatment which is mostly about performing the testing. So, is there a need to perform a test? And is there evidence to perform a test? And my simple answer would be, there is evidence to perform a test.

And in fact, the Measure should be worded more tightly that this assessment should be done in a timely manner, not in the last minute when you decide to use or not use the…

((Crosstalk))

Greg: Yes, I guess I'm confused - it seems like there's two very related - well there are two related proposals. And I'm wondering why there's a need to be two rather than just one? Because if you're going to use the drug, you should test in advance of whether the drug's affective. And I guess my experience would be HER 2 in the old days, which is always tested for. And then you deliver the drug if the test is positive.

So, I guess from a clinical perspective, is there controversy over performing the test in advance?

Dave Ryan: I actually sat on the GI Committee when this was being discussed and there's - we felt in the GI Committee that there's two sides to this coin, both of which are very important. The first is, is that people who are receiving cetuximab, panitumumab, which are the drugs in question here, actually have the RAS mutation and had it tested for. We chose to leave it as it is, because it was already in process and doing very well and we though accomplishing what it needed to accomplish.

There is this growing evidence, though, that - was alluded to in the comments - that those who are RAS wild type - I'm sorry - those who are RAS mutant, who receive this drug actually are getting harmed. And so, we felt in the GI Committee at ASCO that we needed a Measure to address both sides of that same coin.

In truth, we just went through the Trastuzumab discussion. You could make the same argument for Trastuzumab. It's probably not as harmful to give Trastuzumab to people who don't need it. There is harm, don't get me wrong. But there is immediate harm that gets done to people who shouldn't get cetuximab or panitumumab and wind up getting it.

And that toxicity is - in the broadest sense of the word - it's not only physical toxicity, it's - it could be toxicity in terms of their overall outcome. And it could also be toxicity in terms of cost and survivorship.

So, for all of these reasons - and we debated whether we could hone this down better into one measure between the two - we decided to keep it at two measures for the moment. As you may or may not be aware, there's even more data coming out suggesting that only patients with left sided colon cancers - in fact it's not suggesting, it's getting stronger by the day - only patients with left sided colon cancers who are RAS wild type, are the ones that benefit from cetuximab and panitumumab. Those issues are going through the Guideline Committees, kind of, as we speak presently.

So, we can't come out with a Measure ahead of a Guideline Committee. It gets us - the cart ahead of the horse, so to speak. So, we're waiting for that to shake out and then start to come up with our new process measures around this particular issue.

Ava Clarke: Len.

Len Lichtenfeld: I always get a little worried when I know there's a new process Measure coming down the pike and we're doing this one. I don't know - well whatever. There was a lot of discussion in this measure about - and I'm not a GI oncologist and I do not know the literature; like you all do - there's a lot of

discussion about various mutants that have to be tested for evolving knowledge and so forth and new guidelines. That point was made.

And yet the Measure descriptor is the same we've had for a long time. Is there a possibility that somebody could slip - I mean, clearly what you're trying to accomplish is to make sure that they get the appropriate testing, that's what we all want? But there's no guarantee in this measure, the way it's stated that - I mean, could there possibly be a lab out there that's not doing the full compliment of KRAS, NRAS testing.

And they could slip by as passing this measure without having met their intent. Because the intent is clearly new guidelines have new information and yet we're still using the same descriptor and the same Measurement. I mean, someone could go to a chart and say, yep they had the tests, okay. But that's not what this is asking us. That's not what the commentary in here says.

Woman 17: So, going back to your question, the difference between RAS testing in colorectal cancer and HER 2 testing in breast cancer - the HER 2 testing is standard at the time of diagnosis. That is not the case for RAS. It is only relevant to stage 4 and our pathologists will say we cannot test RAS unless we have an order from the oncologist.

Going back to your issue, the question is what is the oncologist ordering? Are they ordering KRAS or RAS, or maybe even a (broader panel)? So, I think that the question is - is this testing necessary? The answer is yes. The process associated with this testing is not well established, let's put it this way

Len Lichtenfeld: So - oh, go ahead.

Dave Ryan: So, let me expand on the HER 2. So, HER 2 testing gets done almost like a point of service. But local pathology departments do the HER 2 testing. This is something that either - I would say most practices in the United States have to be sent out to one of the big next generation sequencing platforms and that takes three to four weeks, sometimes longer to come back. There's a whole process that a practice has to go through to secure the Slides, get the Slides sent out.

There are local practices - or usually if they got (unintelligible) centers who have Departments of Molecular Pathology, that insist on doing this internally - most of those, I would say, are at the cutting edge or even further beyond what the big companies are doing. But not all of them.

And some of them are testing less - in the vernacular, less amounts of the KRAS or NRAS or BRAF genes that are necessary to make the decision. Essentially, any activating mutation in the (Unintelligible) Pathway renders the use of cetuximab or panitumumab useless and more than useless, nowadays, we would say harmful.

So, having said that, it was the committee's - our committee's feeling at ASCO on the GI panel that this was still a very valid test - a very valid measure to look for, because it gets still at the basic component of, is RAS testing being done and are there decisions being made on the basis of that.

And so, for that reason, we felt that it was appropriate to continue. There are always - to your question about evolving guide - evolving measures - there are always new data coming down the pike. And there are some controversial things around this.

You know, most notably - and for those of you who aren't GI oncologists, I'll just throw it out there - it seems like something simple, that we should be able to identify - but what is a left sided colon cancer? You know, what is a right sided colon cancer and what is a transverse colon cancer.

And these are really - they're not necessarily easy definitions when you go and try and apply them to a measure like this, then try and abstract that data successfully. So, there's a lot of evolving issues around this measure. I don't think it makes the measures any less good than (unintelligible).

Len Lichtenfeld: Well, I live in a world of skepticism and, you know, being - just because you're a (unintelligible) doesn't necessarily make you right, okay. We know, there's a whole situation going on, on the west coast and a book written about it - you know, what was going on over there - what people said they were doing wasn't what they are doing. But that's - I'm going off target.

I'm becoming a little concerned that a global measure of RAS testing absent statements in the Numerator - or the Statement of Purpose, whatever the word is, specific - I mean, because it's clear to me that in this measure, that those new guidelines were considered important and I don't see if reflected in what we're actually measuring.

And I worry - I do worry - now I may be wrong - I worry that there may be a lab out there that uses a kit. They may not use all - I mean if there's a kit test available for example - I don't know, there may or may not - I understand what you're saying. I hope that you all do - but maybe there's some kit in some lab in the hospital saying okay, we tested for KRAS and it's okay. But in fact, there were 10 others they didn't test for, whatever the number is. I'm getting a little bit concerned.

Now you also said you have another measure coming down the pike. Is that measure intended to address those specific questions?

Dave Ryan: Yes, it will. But it's - I would say it's further - it's further down the pike in development then the HER 2 measures. In part because the evolution of RAS testing is still evolving. It's not as mature as the HER 2 stories.

Shelley: I guess, I mean, the accuracy of the testing is not really the scope of this. It's just this measure. Are we making sure we're not giving treatments to patients who haven't been tested to ensure that they would benefit from it? And especially, if they would be harmed by giving - by receiving the treatment. Right?

Not that the accuracy of the testing isn't important, but it's not…

((Crosstalk))

Len Lichtenfeld: No, it's more than that Shelley. I've - actually - and again I'm not an expert in this - there now are a number of RAS mutants that can be detected. And that's the point of the guideline. Any of those RAS mutations as it was counted, anything in that pathway, will invalidate and potentially harm a patient.

So, in the past when this first came out - when it was first reported, it was simply mutation, no mutation. Now it's a whole class of mutations and the failure - and what I'm hearing is the - it's not just the generic high level that you do the test or not.

But as you get out into the periphery of what we know is generally accepted as appropriate medical care, we could harm patients. So, I think the granularity in this case - and this is my concern - is becoming more important

because it could cause harm if it's not done. And we have no assurance that it's - unless somebody can tell me - unless somebody can tell me there's only a single - across the United States, under every circumstance, somebody does a test.

So, for example, I don't know if - is there a blood-based test for this yet? Because some of those only have a certain number of mutations included and is in development and an FDA approval goes out.

Dave Ryan: It's coming.

Len Lichtenfeld: What?

Dave Ryan: Yes, it - it's coming.

Len Lichtenfeld: Okay. But the point is harm could occur for any of the downstream mutations. Not just at the high level, yes, no the testing was done.

Ava Clarke: I was just going to ask if there were more comments on Evidence. We can - do we feel like we can vote on evidence? Do we have more discussion? I mean, I think through the comments have brought it into evidence question.

Greg: Yes, I mean, from my perspective, most of the commentary was around the Reliability question. So, it almost seems like people are - have accepted the Evidence, but my opinion is, that we're probably ready to vote on that.

Ava Clarke: Okay. Let's do that.

Hannah: Voting is open for Evidence on Measure 1859. Options are A for High, B for Moderate, C for Low, and D for Insufficient. Voting is closed for Evidence

on Measure 1859. We have 4 votes for High, 13 for Moderate, 1 for Low, and 0 for Insufficient. This Measure passes on Evidence.

Ava Clarke: Okay. So, let's talk about the Performance Gap now and I think Greg, you already kind of kicked this off on that. I don't know if you have anything more you want to add or if we open it up to other maybe members for comment on that?

Greg: No. Other comments I already plowed right through it.

Ave Clarke: Anyone else…

Greg: I will make one additional comment and that is, we have identified some of the (code ons) and KRAS and NRAS that we feel have to be included in that. Are they complete to the point of the previous comments? The answer is probably not as complete as they possibly could be.

But there is some debate. Remember that as the test - the molecular genetics tests get more complete and you go deeper on the analysis, then the evidence becomes retrospective. You're waiting for (someone to go back) and retrospectively look at that particular mutation. Specific mutation of a specific RAS (code on) to see if it is associated in those clinic trials.

What is becoming evident though, is that any activation - any activating mutation in the broad (MAC) (unintelligible) pathway probably renders use of an anti EGFR monoclonal antibody as useless or potentially harmful.

Ava Clarke: Any other comments on Performance Gap? All right. The we can vote on Performance Gap.

Shelley: Actually, if I make one note to further support the Performance Gap is the data that you provided from MIPS is before the Measure was expanded to include (unintelligible) testing and (unintelligible) care.

Hannah: Voting is now open for Performance Gap on Measure 1859. Options are A for High, B for Moderate, C for Low and D for Insufficient. Voting is closed for Performance Gap on Measure 1859. We have 7 votes for High, 10 votes for Moderate, 1 vote for Low and 0 for Insufficient. The Measure passed on Performance Gap.

Ava Clarke: Okay, great. So, now let's move into Reliability. Greg, you want to start us off.

Greg: Sure. So, well we - I think we've already moved into Reliability because there was a (unintelligible) discussion about the specifications, which - you know, the current submission for Maintenance of Endorsement - I guess I didn't really - they didn't really loop back on that. So, whatever concerns people may have had, I guess they'd have to take that into consideration when they vote on Reliability.

For Maintenance of Endorsement, there was a question of the Reliability testing. And Reliability testing was performed and the initial rating was Insufficient, I believe, based on the, you know, letter-for-letter discrepancy between the - the use of the term facility in the submitted documentation versus the actual level of analysis specified in the measure.

My reading of that was that was just a terminology discrepancy and not a deal breaker for the Reliability testing. And if you accept that interpretation, then the Reliability testing was satisfactory and showing that I had had acceptable Reliability.

I believe that was also the opinion of others based on the pre-meeting comments where, you know, it wasn't - people didn't pile on that it was insufficient. They actually thought that it was - the comments suggested that the Reliability was fine. So, I guess I'll stop there for Reliability and let other people weigh in light of our more recent discussions.

Len Lichtenfeld: Just a brief comment and I absolutely agree that my comments before were probably - I mean I'm worried about definitions and how they follow through. But Reliability is - can we accurately measure what the measure's stating is one thing?

Validity, however, is perhaps the area where the concern may be. Are we actually measuring what we should be measuring based on the submission that was - comments on the submission (about the new guideline)?

Woman 4: I would like to note that there was a little bit of an error in the preliminary analysis numbers that were provided to you. It actually was just for (unintelligible) acceptability numbers the data for 1860 was transposed in. But as NQF is aware, we're getting that sorted out on the back end and the numbers are very similar.

So I guess feel free to speak to this, but my understanding is that the kind of grouping isn't impacted by the change in numbers, but (unintelligible) get through the specifics if you want to know them.

Views or comments?

Female 3: Do we have ideas (unintelligible) reliability then?

Female 1: Okay.

Female 2: Voting is open for reliability on Measure 1859. Options are A for high, B for moderate, C for low and D for insufficient. Voting is closed for reliability on Measure 1859. We have 2 votes for high, 14 for moderate, 2 for low and zero for insufficient. This measure passes on reliability.

Okay. So now, let's talk about validity, which (Lynn's) already kicked us off, but (Greg), do you want to speak to that?

(Greg): Okay. Sure. So, again, this is for maintenance of endorsements. So the expectation is that they completed validity testing. They did that in various (unintelligible) validity testing of the measure score and they did this by coordination analysis with the other KRS measure that we're going to be looking at shortly.

And the results of that was, you know, a correlation coefficient of, I think, it was .49 for this one based on 28 practices which, you know, is, you know, moderately good correlation, but that's all you need.

And so again, reviewing the pre-meeting comments, nobody expressed really any concerns. I don't have any concerns. And so - but it's maybe people have evolved in their thinking.

So again, I'll leave it to everyone there to discuss further.

Female 3: Dr. (Barse)?

(Dr. Barse): It's from my own learning I do have a question. As I vent with these measures each one of them are assists on different data sources for reliability and

validity, some from this, some from (unintelligible), some from registries, whatever.

How do we rate these sources in terms of their relevance? I think that's important than we talk about a reliability and validity. It is - it's beyond the scope of a single measure, but we have to rate how we see each one of these sources.

Are they are reliable source for assisting these properties of these measures?

Female 1: Yes, I think that's an excellent question and probably one that we would have to talk about overall at NQF and maybe something that can be brought to our leadership and some of the bigger committees like the CPAC or the board to talk about (unintelligible) perspective, but I think it's a really excellent question.

(Greg): Yes, I mean, it's just - if I can opine on the matter, you know, there are many measures in the world that exist that have never been tested and any testing is non-mathematical proof that the measure is truly valid.

It's just evidence that we can use and compared to measures that have undergone no testing whatsoever, tests - measures that have gone sort of, like, through this type of accepted testing, you know, have, again, some body of evidence to suggest that they're more likely to be valid than measures that have not done the testing.

And so in this case, they've gone through a prescribed method of validity testing. The results have been, you know, generally favorable. Again, I - in my opinion, it doesn't approve that it's valid, but the measure developer table they did the right thing, they did the right type of testing.

And so unless we have specific reasons why we think that the testing is, you know, grossly a representative of reality at least from my perspective they did as good of job as possible given the fact that all of these measures are based on heterogeneous groups of individuals submitting data of varying quality, you know?

You can only go so far. But they went as far as they could and they provided us with some evidence. They did the testing and, you know, overall it looked pretty good to me and at least to people who had submitted comments prior to the meeting.

Female 1: (Lynn)?

(Lynn): I'm on the horns of a dilemma. I think that I agree absolutely that the measure has reported on the validity testing assigned. I have no problem with that. I am having a problem with what we should be measuring and that - and not only there's a such a space to my personal knowledge, that was in the submission that the intent was, you know, that the new guideline to have a broader panel.

This measure came out of an era where it was yes/no. There was a report - I think it was from Australia if I remember. I'm now trying to remember this (unintelligible) journal a number of years ago, a couple of years ago, where we just said this is an observation.

But now, the science has advanced and the dilemma is I think we need to have a measure because I think it's important. The problem I have is are we measuring what we should be measuring and the statement, the submission itself, says what we should be measuring and I'm not sure we get there.

So I don't know what the - I don't know the answer. I would hate like heck to see this not go forward because of the validity question. So I - but on the other hand I cannot make the observation of that - what the submission says versus what the measure is looking at.

The measure still is a yes/no measure. It's not - it's granular is what the submission says it should be. I just - so I don't know what you do. But can somebody help me out of my dilemma on what we should do?

((Crosstalk))

(Dr. Barse): So in - I do agree with you. I had the same dilemma. I think in the reality of their - and their quality measure is meant to measure that reality. And are we putting the right metrics to measure what the reality is out there? And that's what comes to bridge the gap, what's the measure? Is it reliable and is it valid?

I do agree with you that there is no perfect measure because our science is evolving and if you wait for our science to mature, we will never have a measure.

I think this measure - there are some issues with the measure itself, but in general it probably as (Dr. Ryan) stated it probably provides an emphasis to the system to have something to look at the issue of the right mutation when it comes to colorectal cancer patients. That's how I see it and that's what I think it should be sustained.

Man 1: So you're pointing out a real big philosophical question and we struggle with this on the measure's panels across the board and that is we have a process measure.

And so we're measuring process; however, that process is relying on a particular task. In this case, a molecular task which is may not be defined to the level that we want to be defined at.

And when we've tried to do that and define those molecular tests -- and I would say we're all struggling with this at the level at a granular level -- in each of the disease centers now, you know, G.I., breasts, thoracic and so forth.

The - we're kind of waiting for the molecular pathology community to get their act together if I could be blunt and I don't think they've got their act together quite yet.

And so we as oncologists who are experts in the process of delivering care whether surgery, radiation or medical oncologists can't tell the molecular pathology community who are the experts in measuring these mutations exactly what need - what they need to do.

It's a real - yes, I don't know how to solve that problem in our system, but I agree with you that from a philosophical standpoint it is the one - it is the huge flaw, I would say, in both this measure and the next measure we're going to discuss.

Female 1: Okay. Shall we go ahead and vote on validity? Voting is open for validity on Measure1859. Options are A for high, B for moderate, C for low and D for insufficient. Voting is now closed for validity on Measure 1859. We have 2

votes for high, 13 votes for moderate, 3 votes for low and zero for insufficient. This measure passes on validity.

Okay. So now, let's move on to feasibility. (Greg)?

(Greg): Okay. Well, I'll be succinct. The preliminary rating for feasibility was moderate. Everybody who commented prior to the meeting agreed with that. And based on the fact that people are currently reporting the measure, it's certainly feasible for a subset of the measured community.

So I don't think anybody has any doubts, but if they do, they should probably talk about them now. Thanks.

Female 1: Anyone? Okay. Then we can (unintelligible) then. Voting is open for feasibility on Measure 1859. We have A for high, B for moderate, C for low and D for insufficient.

We're just waiting on one more vote. Voting is closed for feasibility on Measure 1859. We have one vote for high, 17 votes for moderate, zero for low and zero for insufficient. This measure passes on feasibility.

Okay. So next is use and usability. (Greg), would you like to kick us off?

(Greg): Yes. So this is a very similar to the previous measure we discussed. It's for endorsement. Since it was endorsed, it's been used in several accountability programs. It is not currently publicly reported and you have criteria required that it be publicly reported within six years of initial endorsement.

Despite that, the preliminary rating was passed. I didn't see anybody raise into those concerns in the pre-meeting comments and I personally, again, feel that

you're somewhat subject to other people about, you know, other people decisions about whether to use your measure and to collect the data and publicly report it.

And so the degree to which a measure developer can make that happen I think may be limited.

So as I said, it's in the accountability programs. It's not publicly reported. And they indicated that they haven't received any specific feedback from people being measured or people using the measurement related to the usability or use of the measure. Thanks.

Female 1: Thank you. Any other comments from (unintelligible) for usability? Yes?

(Dr. Barse): For my learning, we talked about user and usability. From whose perspective are you talking about?

Female 1: I mean, so we try to focus on the folks who are implementing the measures. So one of the things we get back from the developers sometimes are comments from people who are using the measure in the field. We try to include some of that information in our analysis.

(Greg): Well, since that was brought up, I mean, my analysis was based on use as you said, you know, just be - is it used by a program? Like, have people found that this is a measure that they want to collect the data on and report it or measure people on?

And then the usability relates more towards the people who are being measured and can they receive feedback that is helpful to them if you change

their practice such that how they're being scored isn't just an observation, but actually a tool that they can use to make changes.

Female 1: Anyone else? All right. We can vote on these.

Female 2: Voting is open for use on Measure 1859. Options are A for pass, B for no pass. Voting is closed for use on Measure 1859. We have 18 votes for pass, zero votes for no pass. This measure passes on use.

Female 1: Do we want to move to - okay.

Female 2: Voting is open for usability on Measure 1859. Options are A for high, B for moderate, C for low and D for insufficient. Voting is closed for usability on Measure 1859. We have 4 votes for high, 12 votes for moderate, 1 vote for low and 1 vote for insufficient. This measure passes on usability.

Female 1: Great. So before we move onto - well, now, I think we need to talk about the overall suitability, but I'd love to hear from any of our patient advocates on the committee if you have thoughts on this ever hearing the discussion that we have? (Heidi)?

(Heidi): I wasn't sure when to chime in. I'm - this is the first time I've participated, but I've been on set for other grant review boards. This is the first time I've ever heard a comprehensive consideration for all types of toxicity and they were celebrating that over here up to and including financial taxes due which I have not heard represented anywhere else.

So I applaud that and just find it wonderful. So thank you for allowing me to know that and report that back out to the community. So especially with this measure, I mean, it's truly important. Thank you very much.

Female 1: (Bernard)?

(Bernard): Hi. I'm in agreement with (Heidi) as well. For me - and I'm a colon cancer survivor - I'm just happy to see a measure there. So something - again, and I - it's just being relatable as a - the patient, something's there. Somebody wants to trust something so that we have an opportunity for choices of treatment.

Man 2: Well, you know, from my perspective it was a little eye-opening and I really was encouraged by the discussion that happened among, you know, the fact that the technology is moving so quickly and how you can have a measure that over time does not take in the complexity of what really is going on and I think that kind of shows that there's a challenge that I think is here to really keep up with that and how to evaluate really - really get into the details of how the technology has changed.

And I think the second point that I saw -- which I've seen many times before -- is some of the siloes that are happening within health care and that have always been there with molecular biology versus the clinicians who are trying to implement, you know, what the standards are.

So I thought the discussion was good and I certainly didn't have any concerns. I was encouraged by the fact that, you know, we were bringing up some of these details and some of the complexities with the technology.

Female 1: Thanks. And I will chime in. I'm not a survivor myself, but representing patients that - I think, you know, some of the discussion is a little bit alarming about how we, you know - it's good that we're - the science is moving quickly and we're learning what works and doesn't work, but also a little bit of learning to think that you're not getting the right testing because patients want

to assume that - first of all that if there's testing that is done correctly, but that they're getting the right testing to guide the treatment decision making.

So while - I mean, I think having this is better than not having it on balance because it does - I think it's also really important to (Heidi's) point that we have a measure that's really looking at avoiding harm to patients as to taking a treatment that's not going to benefit them and looking at that harm holistically in terms of, you know, the - well, it's the opportunity cost of treatment that - being on the right treatment, but also the difficult toxicities and the financial toxicities of taking - of being on the wrong treatment.

So I think on balance even though there are these concerns we have, it's really important for patients to have this test to make sure that they're getting the right treatment and that they're not getting the wrong treatment.

Man 2: And I would just - and I would add it really - I mean, you know, good for the quality forum to trying to come in and do this, but that really is the challenge is figuring out how do you evaluate this measure that takes into account all of that.

Female 1: Thank you all. And other comments on overall suitability? Are we ready to vote? Okay. Great.

Female 2: Voting is open for overall suitability for endorsement for Measure 1859. Options are A for yes, B for no. Voting is closed for overall suitability for endorsement for Measure 1859. We have 16 votes for yes, 2 votes for no. This measure passes.

Female 1: All right. We made up a little time. We're only 15 minutes behind, but let's just make the break 10 minutes instead of 15 so that we can try to stay on track. So we'll be back at 10:40.

Female 3: Okay. Why don't we go ahead and get started again? So we were going to - why don't we go ahead and get started again. So we're going to move on to companion measure, I guess we'll call it, for the last discussion, 1860, patients with metastatic colorectal cancer and KRAS mutation spared treatment with anti-epidermal growth factor receptor model clone antibodies.

And I then like to first turn the discussion over to the developers and then we'll move on to our group.

Man 3: So I - this is much of the same conversation we just had, but I think this one is much more straight-forward and let me just focus on the denominator and then the numerator.

The denominator is adult patients with metastatic colorectal cancer who have a KRAS gene mutation and the numerator is anti (unintelligible) monoclonal antibody therapy not received.

So it should be 100%. And so I think - I won't go into all the reasons why I think this matters because I think we went through it, but I will say about this particular measure that it's very much straight-forward even if the passive inadequate of no one debates right now a positive result.

So if you have a positive result meaning if you are a KRAS mutant by whatever measures being used, you should not be getting a monoclonal antibody if epidermal growth factor receptor and this measures that particular

issue in a very straight-forward manner and we will have gaps in this particular area that needs to be addressed.

So I think given all the prior discussion, we could just move straight on to what you guys need to do.

Female 3: So (Danielle Zuniki) is going to be the lead discussion. (Danielle), are you back on the line?

(Danielle Zuniki): I am…

((Crosstalk))

Female 3: Thanks. Go ahead.

(Danielle Zuniki): Okay. So - sure. So as described - thank you for describing it. I think this measure is way more straight-forward than the other measure. It's a companion measure. This measure's 1860. We've already read the title.

As far as the description, it's patients with metastatic colorectal cancer and KRAS gene mutation spared treatment with anti-epidermal growth factor receptor monoclonal antibodies.

So again, it's basically not treating if the patient is positive for the KRASG mutation.

As far as the level of analysis, it's a group - it's a clinician/group or group practice and as far as evidence, there's preliminary (unintelligible) relayed - rated high and it is an ongoing maintenance measure.

So this is not for new endorsement. It's a maintenance measure. It was last endorsed October 2012 and since then there's been new evidence and we'll talk about that in a moment.

This is a process measure, not an outcomes measure. And the new evidence as briefly discussed -- and I'll try to be brief -- but basically it includes halting the use of anti-EGRF monoclonal antibody therapies in patients who will derive any benefits.

And the body of evidence addresses the relationship between KRAS's status in patients with metastatic colorectal cancer who underwent anti-EGFR monoclonal antibody therapy specifically Cetuximab or (Pantoshinab) and the outcomes of the tumor response profession through survival and overall survival and that it was not as effective as we would've hoped.

So therefore, this evidence is based on also guidelines specifically for guidelines. The American Society for Clinical Pathology, the College of American Pathologists, Associations for Molecular Pathology and (unintelligible).

So those specific guidelines talk about colorectal carcinoma patients being considered for the anti-EGFR therapy and receiving KRAS mutational testing.

And furthermore, the other guideline that supports this evidence, NCCN guideline. It's a Level 2A guideline and it talks about patients with any known KRAS mutation or NRAS mutation (unintelligible) 234 and for NRAS it's 234 as well.

It should not be treated with either Cetuximab or (Pantoshinab). And also, the FDA labels have been updated in the indication section for both products to reflect this.

So that's the new evidence. And as far as the pre-evaluation comments are concerned, there was a high level of agreement that there was sufficient evidence.

So I'll stop there for any conversations.

Female 3: So any conversation from the room right now I see no cards turned up. I think we did have a robust conversation about the evidence already.

(Danielle Zuniki): Agreed.

Female 3: Anyone else on the phones? Otherwise, I'm going to move that we go ahead and vote on this section.

(Danielle Zuniki): Thanks.

Female 2: Voting is open for Measure 1860 on evidence. Options are A for high, B for moderate, C for low and D for insufficient. We're just waiting on one more vote. Voting is now closed for evidence on Measure 1860. We have 11 votes for high, 6 votes for moderate, zero votes for low and zero for insufficient. This measure passes on evidence.

Female 3: So (Danielle), let's proceed to the gap.

(Danielle Zuniki): Okay. Great. Thank you. As far as the gap, so there is an opportunity for improvement here. On the data presented by ASCO is based on 2017 MIPS Performance from Registry Data provided from CMS.

The data was based on 158 providers representing 43 practices and 495 individual patients. It appears the gap is 91% suggesting improvement.

As far as any pre-evaluation comments, this was also agreed upon -- at least within the comments -- that there is opportunity for improvement and the preliminary rating from (unintelligible) was high on this particular category.

Don't know if anyone has any...

Female 3: So - yes.

(Danielle Zuniki): ... information.

Female 3: So again, any comments in this area? Do they - I think you - the - you - the rate was 91%. You described the gap was 91% and that's an error. Otherwise, that was statistically significant in performance. So I think…

((Crosstalk))

(Danielle Zuniki): Yes…

((Crosstalk))

Female 3: Yes. Any other comments? Because we're on the path to new record for evaluating evidence, but we'll - but we did have the throw discussion. So I

don't have a - and so why don't we move - right. Exactly. So we'll go ahead and move to vote on this performance gap.

Female 2: Voting is open for performance gap on Measure 1860. Options are A for high, B for moderate, C for low and D for insufficient. Voting is closed for performance gap on Measure 1860. We have 15 votes for high, 3 votes for moderate, zero votes for low and zero for insufficient. This measure passes on performance gap.

Female 3: So we'll move on to reliability.

(Danielle Zuniki): Okay. So as far as the reliability, we'll discuss the specifications and testing together. So first, it has to do - let's talk about the changes. So the measure has been expanded from its last endorsement to KRAS mutational testing based on guidelines, which I've already talked about what those guidelines are to include NRAS in addition to KRAS.

And just detecting for mutations in KRAS fast (unintelligible) 12 and 13 as recommended previously before to treatment with anti-EGFR antibody therapy. Patients to metastatic colorectal cancer should have their tumor tested for mutations in additional mutations including KRAS Exon 3, Codon 59, 61 and 4, Codons 117 and 146 and then NRAS, NRAS Exons 2, Codons 12 and 13, three Codons 59 and 61 and four Codons 117 and 146.

Additionally, the measure (unintelligible) removes exclusion for patient transfer to practice after initiation of chemotherapy and receipt of anti-EGFR monoclonal antibody therapy as part of the clinical trial protocol. And those were the substantive changes to this particular measure versus when it was endorsed in 2012.

So I wanted to see if there - to stop and see if there's any comments on that. Otherwise, I'll go into the reliability testing.

Female 3: No comment.

(Danielle Zuniki): Okay.

Female 3: Oh, (Lynn). (Lynn) does.

(Lynn): Yes, you would - you went through the entire list of, you know, what should be tested for. Help me out. Is that in the measure? I mean, is it written in the measure?

Female 3: Yes.

(Danielle Zuniki): So I actually did have a question about that for ASCO. It's - could you comment on that if you will?

(August): Hi. Yes, this is (August) from ASCO. So that is in - within the specifications within the numerator guidance and that's for both this measure 1860 and also within 1859. So we do go into specifically which Codon based on evidence today should be included and that list can be updated as we move forward in the future.

(Danielle Zuniki): Okay. I thought…

((Crosstalk))

(Danielle Zuniki): Okay. Thank you very much.

(August): And so previously the measure - I think there's - some of the Codons for KRAS were added since the last submission and then also NRAS all of those Exons and Codons were added...

(Danielle Zuniki): Yes.

(August): ... into the…

((Crosstalk))

(Lynn): Okay. This reflects my ignorance. Where is that in the measure descriptors? Is there a pending - help me out. Is there a body of material that details that so there's someone who is reporting on this measure can say, yes, this is what I should have and - because that was the whole discussion we had the last go round?

(August): And...

(Lynn): I just don't know. I'm looking at a measure to - a numerator and denominators thing here, it's one sentence.

(August): And so within the preliminary assessment that you guys received, only certain components are copied over from the measure information form, the full form.

So then that longer form there is a place to include the general statement of the numerator and denominator, but then to also provide specific guidance. So it's provided in that specific guidance.

And how that normally translates when it's implemented into a program is that it is abstract or health checks. So when (unintelligible) measure they can

(unintelligible) depending on the program, but, you know, there's always that little question mark, right, that you can hover over and see what does this mean specifically for my patient.

(Lynn): Thank you.

(Greg): It's on Page 33 of the preliminary analysis.

(Lynn): What page number? I'm sorry. Thirty-three?

(Greg): Thirty-three.

(Danielle Zuniki): Thirty-three. Thanks, (Greg). May I move on to the testing? Thank you. Thank you for explaining and clarifying. So as far as...

Female 3: One second.

(Danielle Zuniki): Oh, sure.

Female 3: We have a question. (Steve)?

(Steve): One question for the developers. I saw that - it mentioned that you removed the exclusion for - that they transferred over to a BRAC after they'd already initiated the therapy.

Can you talk about the rationale for removing it in this circumstance? Because I know it's still in some of the other ones. So is - was there something specific about this one?

(August): So I'll be transparent here. Part of the limitations when we're going through this endorsement process is that we have to have the data available to support the measure.

And so in this case they were using mixed data where those exclusions were not included in the measure specifications. They do have some programmatic guidelines around similar things when you are reporting submits. So it's not in your specific measure.

So to align with the data that we have available that we could use to support a measure from a testing standpoint, we have by necessity need to remove those to match what was available.

Female 3: Okay. (Danielle), I think you can proceed.

(Danielle Zuniki): Okay. Thank you. So as it relates to testing, the mean reliability is 9% with a median of 100%. And also, I think we did just talk about the measure specifications; although, obviously it's a lot longer than what is needed in our packet and I won't go into all of them.

But as you know, in the numerator there were a lot more mutations listed, but the numerator as it states is anti-EGFR monoclonal antibody therapy not received and the denominator statement is adult patients of metastatic colorectal cancer who have a KRAS (unintelligible) includes KRAS or NRAS gene mutation.

So that is the specifications. As far as the preliminary comments that it did not look like there were any concerns regarding reliability testing and Q.S. rating was preliminarily high on reliability.

Female 3: Okay. Dr. (Barse)?

(Dr. Barse): I actually have a question about the testing panel. So if we have an older patient which met all the criteria for Exon 23 and that's positive, do we need to do testing that's compatible with the current or it's - if you have one mutation that's established and is enough, I think that's a problem because I don't want to retest somebody that I know on the old testing was positive. That percent is done.

Man 3: Yes, so the old testing is inclusive of one of the Codons.

(Dr. Barse): Right. So if you have the old testing...

Man 3: Yes.

(Dr. Barse): ... that's added (unintelligible) right?

Man 3: Exactly.

(Dr. Barse): It's - the testing does not need to be inclusive of all of this current recommendation?

Man 3: Correct.

(Dr. Barse): Okay.

Man 3: Correct. So once you're positive, you're positive.

(Dr. Barse): Yes, that's what I wanted to...

Man 3: To clarify the question for the rest of the group, so there would be a - if you were negative, let's say it's in 2015...

(Dr. Barse): Yes…

((Crosstalk))

Man 3: ... you had metastatic colon cancer and you went on chemotherapy and it got tested and then it got receptive in the liver and you didn't need any more chemotherapy for the next three, four, five years and all of a sudden you had new (unintelligible) seizure, liver metastases, we would recommend that you go back - not we as ASCO, but we as experts in G.I. cancer would recommend you go back and get retested under a more complete genetic testing.

(Dr. Barse): Yes, but for the sake of the measure, if you have one positive...

Man 3: Think of the measure.

(Dr. Barse): ... it's counted, right?

Man 3: Correct. Correct. Once you're positive - once you're a RAS, you're essentially always a RAS…

((Crosstalk))

(Dr. Barse): In - but I mean in the coding.

Man 3: Correct. In the coding.

(Dr. Barse): It is appropriate to reflect it that that's counted not as (unintelligible)?

Man 3: That is correct. Yes.

Female 3: So (Lynn)?

(Lynn): This may be a comment to be differed to a potential harm's risks, whatever, of the test, but I think you said earlier you're eluding to the fact that current testing is NGS-based?

Man 3: Yes.

(Lynn): Okay. So just as an FYI since this is - I'm using the Medicare population, I think we have to bear in mind that Medicare will only pay for one NGS test per patient per tumor.

So there may be a financial issues for the retesting that is not - that will not be covered. Now, I'm not saying that is the way it will happen, but I do think that - and I think it's important that we put that in at some place because from a policy perspective it's - pardon my expression - is atrocious that that rule is in place particularly in a medically vindicated situation that you just described and I agree with what you - I mean, from what I know I agree certainly with - not just with colon cancer.

So that's a potential risk for this measure that a patient may not get it done because they can't afford to get it then.

Female 3: Any other comments? So (Danielle), do you have further comments on reliability testing?

(Danielle Zuniki): I do not.

Female 3: Anyone else on the phone before we move to a vote on the reliability? So why don't we go ahead and vote?

Female 2: Voting is open for reliability for Measure 1860. Options are A for high, B for moderate, C for low and D for insufficient. Voting is closed for reliability on Measure 1860. We have 10 votes for high, 8 votes for moderate, zero votes for low and zero for insufficient. This measure passes on reliability.

Female 3: So we can move on to validity.

(Danielle Zuniki): As far as validity, this is, again (unintelligible) reliability is high/moderate, the empirical testing has been for validity and there were no threats to validity not in the comments, in the committee, nor in an interest assessment.

So I don't know if anyone has any additional data to add.

Female 3: Okay. I don't believe there's any comments in the room. So I guess we'll proceed with voting.

Female 2: Voting is open for validity on Measure 1860. Options are A for high, B for moderate, C for low and D for insufficient. Voting is now closed for validity on Measure 1860. We have 12 votes for high, 6 votes for moderate, zero for low and zero for insufficient. This measure passes on validity.

Female 3: We'll turn to feasibility.

(Danielle Zuniki): Okay. So as far as feasibility, this - all data elements were collected through electronic data or using keyword searches. Preliminarily, this was redid as

moderate and the comments from the committee, there were not a lot of comments regarding feasibility.

The only one comment was there is - that it's not structured in the EHR. So that could potentially - I do have a question, I guess, for the measure steward. In the same vein of conversation, I guess, ASCO is not - it sounds like based on the data we have ASCO is in the process of assessing the feasibility of developing an electronic clinical quality measure.

So I guess my question is more for ASCO on this piece.

(August): I think for the time being the reason that we are continuing to endorse this measure as a registry measure is because right now it does require some level of abstraction as, you know, electronic health records continue to evolve and there are new discrete data fields that is the dream, so to speak, as to great new measures, but right now we're just really limited and to make sure that we have the right population and are, you know, looking at the right evidence is more important than us being able to pull everything electronically at this time.

(Danielle Zuniki): Okay. Thanks. So I did see us abstracted and also electronically when you can. Got it.

Female 3: So any other comments or questions or concerns about feasibility that the group needs to air or anyone on the line? So we can move to voting for feasibility.

Female 2: Voting is open for feasibility on Measure 1860. Options are A for high, B for moderate, C for low and D for insufficient. Voting's now closed for feasibility on Measure 1860. We have one vote for high, 17 votes for moderate, zero

votes for low and zero votes for insufficient. This measure passes on feasibility.

Female 3: Use and usability?

(Danielle Zuniki): Okay. As far as usability, the preliminary rating was moderate as well. It is actually - this measure is used in various accountability programs. So MIPS, the (Unintelligible) Qualified Clinical Data Registry, (COPE), et cetera.

So it is being used. And additionally, as it relates to benefits outweighing harm, the benefit does outweigh the harm according to preliminary evaluation and also the NQF committee comments.

Female 3: We have a question or comment?

Female 4: Actually, a quick question. It says publicly reported and it says no. Can there be just a little bit more elaboration on that?

(August): Yes. So with public reporting right now as a plan for CMS is to put their measures that they would be publicly reported and our measure is a myth, but there have been continued delays in them actually setting up that public reporting.

So it's kind of we don't have our own platform for public reporting. We use these measures at ASCO within our ASCO programs for quality improvement. They're really not intended to be shared, you know, as publicly and externally on an individual practice level.

So at this point we're kind of waiting for CMS to follow through on what they said that they're going to do and still plan to do. It just hasn't happened yet

and I think that's something that other measure developers in addition to us are kind of struggling with at the moment.

Female 3: Any other comments? So we'll proceed with the next two votes use and then usability.

Female 2: Voting is open for use on Measure 1860. Options are A for pass, B for no pass. Voting is now closed for use on Measure 1860. We have 17 votes for pass, one vote for no pass. This measure passes on use.

Voting is now open for usability on Measure 1860. Options are A for high, B for moderate, C for low and D for insufficient. Voting is now closed for usability on Measure 1860. We have 2 votes for high, 15 votes for moderate, one vote for low and zero votes for insufficient. This measure passes on usability.

Female 3: We didn't previously discuss related in competing measures, but the last measure was a complimentary measure. Otherwise, there weren't any conflicts or alignment that needed to occur.

So I think we should move on to overall suitability for endorsement. Oh, wait. Excuse me.

(Dr. Barse): So when we were discussing the last measure, she brought up a good point that if these measures are being used by payers, insurance company for providing (unintelligible) why should it be an NQF measure in case to measure such a thing?

And I think for this measure, not the previous one, this measure is something that is asked for. Then you order the drug, they ask do you have

(unintelligible) results. And as much as I do agree with the evidence gap, reliability and validity of the measure we are doing, I don't think this measure is usable.

Know something that should be adopted given that the peers are pursuing this and the resources in the public domain should not be used to measure something that payers are already pursuing.

(August): I will add part of our motivation as a society in speaking NQF endorsement is actually so that payers will use the measure and to promote consistency hopefully across payers in using the save it measure rather than using all of their own individual versions.

So one of the current users of this measure is the Core Quality Measure Collaborative which was originally convened by America's Health Insurance Plans and now (unintelligible) overseeing the group with the intent of bringing together a bunch of payers to decide on core sets of measures for different diseases including cancer.

So part of what we've seen, it's not a retirement in a lot of these programs that you have NQF endorsement, but it is strongly preferred and I think that's normally the first place they go to look for measures because, you know, that does mean that the measure has gone through rigorous testing and all of these other things.

So I think the goal is to no longer need NQF endorsement because all these payers have already adopted it, they're all using it and hopefully there's no longer a performance gap and they can just get rid of the measures and go on to moving - to measuring something else, but that's kind of our motivation

when we come to NQF is so that we can increase the reach of these measures and you get more people to use them.

(Dr. Barse): So the previous measure discussing the previous session can be used as a venue for bringing a, you know, standard to the measurement by payers? How is that? There is no payer that's not going to ask for these. These drugs are very expensive and remain very expensive.

So I think regardless of what we think in terms of the measure, they're going to ask for the test results.

Female 3: So there's someone on the line that has a comment?

(Jen Maylin): Yes.

(Greg): Oh.

(Jen Maylin): Okay. This is (Jen Maylin). Yes, I just wanted to clarify I don't think that it's so much that payers are using this as a quality measure. There may be some that do, but I think for the most part I think it's being articulated. But, you know, many, if not most payers will have a prior authorization requirement where they will not reimburse for the drug if the mutation status has not been checked.

Female 3: Thank you. There was another person online too.

(Greg): It's (Greg). Well, I guess, as - just for the purposes of our job here and endorsement by NQF, I think it's impossible for us to know because it's not an NQF criteria to survey all payers to see to what extent they make coverage determinations based on the measure.

And so it may be true that many of them do make - require this type of thing, but then that being the case, I mean, it raises the question of the performance gap and, you know, how are these people for the other measure getting the drug if that requirement exists.

And, I mean, if you can imagine any number of scenarios, but I guess I would just say that maybe we step away from that particular concern unless NQF feels that that's a part of their endorsement criteria that we need to examine.

Female 3: Any comments from NQF?

(Nicole): Can you - this is (Nicole). Can you repeat the question?

Female 3: Well, I - to rephrase it, does NQF think that part of the endorsement process should include payer …

((Crosstalk))

(August): Yes. So I think to the point the developer made, we do have the CQMC process, which is looking at many things related to that. So I wouldn't weigh it higher in your decision process as we talk about this overall endorsement.

Female 3: And so I wanted to make a comment as well which is although many commercial payers use preauthorization and require standards, there's large non-commercial payer that doesn't have a preauthorization requirement and that's a large group of patients would fall under this umbrella.

And they can deny payment afterwards, but there's no preauthorization process. So I think that when you think about the patient population it's probably not being mandated until retrospectively. So...

(Lynn): Just briefly on that one, that applies to intravenous drugs administered in a physician's office as a (unintelligible) care I don't know that I'd - and my question would be that that's not the same necessarily for the outpatient medications. It goes through a different process.

Outpatient oncology medications because, again, we're talking about cancer-standing committee larger questions than just this drug.

Female 3: So the - I can't answer your question completely, but I - unless there's an intermediary that's going to…

((Crosstalk))

(Lynn): Well, Plan D. Plan D has an intermediary Medicare...

Female 3: Right.

(Lynn): ... the Medicare Advantage has an intermediary for sure.

Female 3: Right.

(Lynn): So...

Female 3: But straight Medicare?

(Lynn): Straight Medicare does not on physician administer medications in the office.

Female 3: Right. Right.

(Lynn): Yes.

(August): I guess I - I just think we should be looking at it from the performance gap standpoint. If insurance preauthorization makes a measure topped out because you need insurance approval in order, you know - then we may not need it.

But if there's a performance gap, then whatever insurance is doing whether it be because Medicare doesn't require authorization or whatever, then I don't think we should necessarily have that be the overriding factor. It should be something that we think about but look at the performance gap.

Female 3: Okay. Any other questions or comments? Because we have one final - Dr. (Barse)?

(Dr. Barse): On this data for the gap is from 2012? This is 2000 - no, 2015, yes? The gap data.

Female 3: Oh…

((Crosstalk))

Female 5: I think it's from 2017, yes.

((Crosstalk))

Female 5: Although, ASCO, please correct me, but I thought it was pulled from 2017 MIPS reporting.

(August): That is correct. And at the time that we are preparing the submission, that was the most recent data that we had available.

Female 3: So if it's okay we could proceed with overall suitability for endorsement.

Female 2: Voting is open for overall suitability for endorsement for Measure 1860. Options are A for yes, B for no. Voting is now closed for overall suitability for endorsement for Measure 1860. We have 17 votes for yes, one vote for no. This measure passes.

Female 3: Okay.

Female 1: So we - our plan is to have lunch at 11:30, but one option would be to continue if we - if the group agrees. And what we would do is we would change the order a little bit, ASCO.

Our plan is to review 0383, the Oncology Plan of Care Measure before we go to 0384. So we think that flow of that discussion would make better sense for the group overall.

And so if the committee's okay with continuing, we'll jump right into the review of 0383. Any objections to that? And then we'll break for lunch before we continue with the other two measures in that group.

(August): So no objections from a time standpoint. Just a preference. So it's 385 not in numerical order, 384 is obviously, but if the - talking about the pain quantification is sort of first in the process of care followed by the plan of care for pain, I don't know if that's a preference from that standpoint.

Female 1: So here's what I think we need to go - we need to do this first. So we as a committee reviewed the quantification measure 0384 in the last cycle and recommended it for endorsement.

The C-SAC -- which I had on at the time have now rolled off -- asked for the committee to reconsider that and they did - and we can get into the details of it. It was part of our evidence. A lot of it had to do with that it would (unintelligible) measure.

They were ready to just say just to override our decision we eventually through - and (Karen) wasn't on the call. We didn't actually think this was going to be controversial than it turned out to be and I was home sick calling in to the call. So I wasn't here in person to help make the case.

But one of the questions that was raised was did you look at this in conjunction with the related measures and we said no because we had not. And so that was sort of the hook to say let's send it back to the committee through you.

So if we start with 0384 and don't have that discussion, it - then we're not accomplishing what we needed to accomplish. That's why I think we need to start with this first because I do think that, you know, we have - need to have the discussion that if - do we need both measures.

And I think the way we start - we have that discussion is to start with is the plan of care in place. So that's why we're switching it up and I understand your point that it's really, I think, if we're going to reconsider the quantification measure, we have to have this conversation first. Okay.

So we're starting with - and changing the order, we're starting with 0383, Oncology Plan of Care for Pain. And so ASCO is the developer on this, so we will - and this way you don't have leave and come back. You can just stay seated and...

((Crosstalk))

Female 1: ... give us an overview.

Man 4: Okay. So the denominator for this measure is all - it is for patients regardless of age with the diagnosis of cancer currently receiving chemotherapy or radiation therapy who report having pain.

There are no denominator exclusions. The numerator is patient visits that included documented plan of care to address pain. A documented plan of care may include use of non-opioid analgesics, opioid psychological support patient and/or family education referral to a pain clinic or reassessment of pain at an appropriate time interval. There are no numerator exclusions.

So this measure assesses the percentage of visits for patients currently receiving chemotherapy or radiation therapy who report having pain who have a documented plan of care to address that pain.

Proper pain management is critical to achieving pain control. Pain has a severe impact on a patient's quality of life and cancer pain is associated with numerous psychosocial responses.

The most recent NCCN guidelines state that, quote, "Unrelieved pain denies patient comfort and greatly affects their activities, motivation, interactions with family and friends and overall quality of life."

One-third of patients described cancer pain as intolerable aspect of cancer and adequate pain treatment has been shown to result in clinically relevant improvement in health-related quality of life.

ASCO views the relationship between the quality actions of this measure and a patient health outcome (unintelligible). The patient with cancer receives chemotherapy radiation. The patient is then queried about pain level, a plan of care is created to address pain at each visit and improve quality of life and hopefully survival results.

This measure was originally paired as we stated with PCPI's NQF 0384 which was the pain intensity quantified but has since has become unpaired.

When these measures were originally developed, the intent was that these measures would be used together to represent optimal care for patients with cancer.

Although these measures are currently unpaired, our aim is for both measures to be re-endorsed because we think both clinical actions of pain quantification and developing a plan of care to address pain are required to deliver optimal patient care.

Anecdotally we will comment that we have heard from programmatic staff that it's difficult to implement 0383 which is as we stated the plan of care for pain when the companion is not included in the program as was just discussed.

And so we anticipate continued barriers to real world implementation in circumstances for 0383 is implemented in a reporting program without 0384.

Some way can I - I'm going to be - given the prior discussion this is like having a molecular task with no action associated with it.

So we divided the molecular task. In this case, pain (unintelligible). It's not a molecular task. It's a task that you're being asked how - what your pain level is like your blood pressure and now, we're at - we're saying what should be done about it, which is report a plan of action.

Female 1: Can the folks on the phone (unintelligible) if you're not speaking please? Thanks. Okay. So we can start with Dr. (Braidman) (unintelligible) discuss it?

((Crosstalk))

Female 1: Unmute your mic.

(Dr. Braidman): I'm sorry. Sorry. This original endorsement was in 2008, the most recent endorsement in 2012. It's a process measure at the clinical and group practice level and to start out with discussion in terms of evidence.

So the measure is - the evidence is driven by the NCC practice guidelines Level 2A. So based on lower level evidence, the initial recommendation, however, is that the evidence is insufficient because the evidence or practice guidelines don't address the measure head-on itself.

Want to continue or pause for...

Female 3: I think we need it. We need it - yes, and we just - let's stick with evidence first (unintelligible). Yes.

(Dr. Braidman): So where we are seems pretty straight-forward. It's not - I've been in a situation like this before and it's not exactly clear to me how we proceed.

So there's clear evidence that having pain addressed and a plan of care is important, but the evidence for that doesn't relate directly to the measure as stated.

And so according to the evidence algorithm it falls out as insufficient.

Female 3: So can I ask a quick question as clarification with the staff? So the preliminary rating is insufficient. We had this discussion before, but I think that the committee needs to also understand that we as a committee and a body of experts can override insufficient.

So I think that's where the conversation needs to go about whether or not we value this measure from a clinical expertise standpoint or other important measure.

Female 1: (Lynn)?

(Lynn): The - I don't want to get into the whole issue about pain and the evidence we have and people with (unintelligible) that's not uncommon. I mean, the pain will at least (unintelligible).

But I want to go specifically to on Page 3. It says there's no evidence, but it does include in-depth discussion of the (unintelligible). So I'm confused. Is it because - so the - there's an overarching statement. So there's not evidence.

And then the little bullet there says there's a whole discussion about the evidence. Is there or is there not evidence? I mean, I don't know if I'm making myself clear.

And with one hand we say there's no evidence and the other hand we say there's lots of evidence that's discussed. Somebody help me out.

Man 2: Well, and I - let me just jump in real quick. And I just was curious maybe to hear from staff their rationale for insufficient with the evidence, just a little bit more detail on that. That - I was (unintelligible) confused on. Is that okay, (Lynn)?

(Lynn): We're all trying to get to the same point. I mean, I - if we say it's insufficient because of some high level, the evidence wasn't presented, but yet the secondary level there's adequate evidence within the guideline.

That becomes a matter of, you know, how it's stated and I don't want to - you know, insufficient's pretty important to conclude if we say that. I'd rather not have to say that.

Female 1: So we would like to refer you to the evidence algorithm which is what we applied to (MAND) on this insufficient rating. And based on what was submitted with no systematic review of the evidence specific to this measure, a lack of quality, quantity and consistency, on the specific what the measure is specifically looking at was lacking.

However, we found the clinical practice guidelines submitted don't support that specific measure focus. So the guideline does provide the in-depth discussion on the evidence of benefits and harms of specific therapies and interventions.

So I think that the committee here has a couple of options using this algorithm and the deciding factor would be in Box 10. Before we get to Box 10, there's a question is empirical evidence submitted, but without systematic review and grading of the evidence?

If we say no, that leads us to Box 10 and that question is are there or could there be performance measures of the related health outcomes or evidence-based intermediate clinical outcome or process?

And that is what we would look to U.S. experts to give us input on. If you said yes, there would be - it would remain insufficient and there wouldn't be an option to say insufficient with exception.

If you answered no to that question, there are not performance measures of a related health outcome. We could go down this insufficient evidence with exception room which would lead to a pass of the evidence.

So I know it sounds a little confusing. Basically, the questions you would be answering if you went down this room is there evidence of a systematic assessment of expert opinion that the benefits of what is being measured outweigh potential harm?

If you answer yes to that question, the final question would be does the standing committee agree that it is okay or beneficial to hold providers accountable for a performance in the absence of empirical evidence of benefits to patients?

You - yes, sure.

Man 5: So I think - so speaking as an oncologist, I guess an expert in G.I. cancers where there's a lot of pain, I think we're caught in a loop and the loops is I had previously served on the NCCN and I am no longer serving on the NCCN's - not on the pain management guidelines, but I was on the colorectal guidelines.

The loop is, I believe, that Level I evidence is randomized controlled trial demonstrating benefits. It is considered unethical to have a randomized control trial of addressing pain management in somebody who is in pain.

It would be like - I mean, it's kind of obvious, right? There are clinical trials trying to figure out which type of pain management is the best, but there is broad agreement among all the experts in the field that if a person is in pain that that pain should be addressed in some capacity.

We're not stating which capacity that pain needs to be addressed. We are just stating that if you're going to measure pain, you need to address pain and we do think that measuring pain is a fundamental principle of caring for cancer patients.

That being said, I think that because there's not Level I evidence we relied in the measure's steering committee on the fact that the NCCN uniform recommendation by experts was that pain should be addressed and that there is a growing body of evidence, some of it from my institution at MASS General which combines a quality of care intervention with - concurrent with an early diagnosis of metastatic lung cancer that did show and has shown consistent improvement in quality of life.

As a component of that quality of care intervention is addressing somebody's pain. So there is a growing body of literature that quality of care intervention which includes addressing somebody's pain has meaningful benefit

consistently in quality of life and not consistently in terms of overall survival yet, but at least one study overall survival.

So we in the measure steering committee felt like this was - that there was sufficient evidence that this was the right thing to do for patients.

Female 1: Okay. Dr. (Barse)?

(Dr. Barse): I do agree with you that addressing pain is only a humane thing. None of us would ignore pain or should ignore pain. So could we say that the evidence is in the fact that physicians do not have time or attention to pain and therefore justifying this measure rather than discussion whether addressing pain is something that we should or we shouldn't do?

I mean, we should really look at the evidence on a different perspective. I don't think anybody around this table or anywhere else would argue that we should address pain if it is there.

And that's not - we shouldn't look for evidence that is it important to us (unintelligible) obviously it is, but we should ask for evidence is there enough evidence that that is not being attended to adequately.

Female 4: It would seem indeed the measure as plenty of evidence pain is important to measure, you know, it impacts individuals, but rather the evidence does not support that a plan of care makes a difference in that pain, but plan of care would be considered a minimum standard of care for individuals.

And so by default it ends up potentially the only thing that's current and so one doesn't have evidence for what makes a different specifically as it outweighs others. Potentially it's better to be doing something than nothing

that really is, you know - I mean, it's standard of care for a plan of care to exist.

(Lynn): I think I'm going along the same line you are. Are we - and I need to understand something here about parsing the question. There are two questions: One is treating pain appropriate and I suspect that an expert panel would uniformly come to the conclusion of treating patients with cancer pain is appropriate.

So are we parsing out the question that the evidence is insufficient to show that documenting a plan of care is insufficient evidence? Is that - when we use - when you came to the conclusion it was insufficient evidence, was it around treating pain? Was it around documenting a plan of care, either or both? Okay.

Female 5: I would try to chime in and say probably the - I think the documentation of the plan of care is almost a proxy for following the guideline, which is act on the pain, do something about the pain.

So I think from a sort of abstraction or reporting perspective, we're trying to capture some indication that the clinician is addressing the pain. I think it gets a little bit - also, the measure is broad to include any type of pain, any patient-reported pain. That could be low pain.

It could have a low pain score and there should still be a plan of care to address that pain even if it's to say we're going to follow up at our next appointment or we're going to reassess or have you tried, you know, over-the-counter analgesics.

So I think that proxy is what we're kind of doing the best we can to functionally capture whether the clinician is trying to address the pain.

(Lynn): So that answers your intent. I'm interested in knowing what our staff colleagues had to say when they rated it insufficient. What were they thinking?

Female 5: So we were looking apparently at the documentation aspect of it, but the evidence before the documentation as opposed to the treatment and management of pain.

(Lynn): So that in turn, if I may, leads us to the point where we have the assumption - I mean, ASCO's provided their intent, you're focusing on the documentation making a difference in evidence, and then the question before the committee is not whether or not we treat pain. The question before the committee is does the - is - do we think the documentation is an acceptable and effective proxy for getting to that point if I may?

Female 5: Yes, that's correct.

Female 1: Okay. There's a comment on the phone and then (Karen) and then I think we need to talk about how we will question (unintelligible). Yes. So who is it on the phone? Who has the comment?

(Jen Maylin): Hi, this is (Jen Maylin) again.

Female 1: Okay. Great, (Jen).

(Jen Maylin): I mean, I think, you know - I mean, I think the - I think everyone - you know, I would say almost everyone probably agrees conceptually that it's important

to treat pain. I think the challenge, right, is that this is - it's a percent of visit measure, right?

So someone who comes in with more visits would have more episodes than the denominator and that you would need to document it every visit. And then, you know, the - I think as the measure developer maybe was commenting, it doesn't take into account severity, you know?

It's not - at least, like, I know kind of just in personal experience with patients, you know - and this is even - the example I'm going to give is kind of laughable, but it actually was a - so if you have documented pain, the patient with metastatic breast cancer, pain level just described at seven; although, she didn't have any analgesics at the time, so I asked her what the pain was and she - I mean, this was - you know, this is actually a true story.

So it may sound silly, but it's actually a true story. She had stubbed her foot on the way in to the clinic that day and when they asked her pain, it was a seven because she just stubbed her foot, but she didn't have, you know - she didn't have chronic pain and it was resolved at that point.

But, you know, in accordance with the measure, you would have to document it. And the other kind of point that I'll just make is that, you know, that does (unintelligible) put pressure to prescribe medications like opioids potentially when they're not warranted, you know?

The easiest way to document something, just to prescribe something rather than (unintelligible) you know, long-hand (unintelligible) group plan as - and there is - I can't remember exactly now remember where I saw it, but something on the order of, you know, 30% of patients with early stage breast

cancer are still on opioids or early stage cancer are still on opioids over a year after completing surgery.

So, you know, I think that the, you know - while the intent is completely on target the fact that the measure is relatively broad and, you know, there aren't necessarily any studies that demonstrate that people who do a good job documenting pain have better outcomes for their - or benefitting the plan of care for pain have better outcomes, you know?

The - I think the measure leads to a lot of questions in terms of the evidence that the process describes improve outcomes.

Female 1: Okay. Thanks, (Jen). Now, did you say - was (Greg) also - hey, (Greg)?

(Greg): Yes. So - well, I guess my comment isn't quite as relevant. Just exactly what she said. I mean, I think kind of where we are is that there is an evidence for this specific intervention, but although the evidence is insufficient, we might choose to overrule that based on our judgment about the importance of it.

Female 5: I'll just add - I just have a few additional comments. I think that the gold standard would be an outcomes measure and maybe this will become more important for the next phase of our discussion and this phase of our discussion because we would like to see patient reported outcomes or patient recorded - reported outcomes measures give us a more exact answer on how we're dealing and responding to the pain.

Having said that, those measures don't exist or aren't in cancer at least those measures exist in other areas and they're currently being developed, but I don't think that we should - the state-of-the-art now is those measures aren't readily applicable for us.

And throwing out a process measure that is valuable for clinical practice and is also important to the stakeholders, the patients, because it's not a perfect outcome measure it's not necessarily an appropriate way to think about the existing measure in front of us.

So I would say that expert consensus and guidelines that were cited to create this measure seem to be inappropriately to look at evidence as well.

Female 1: Okay. So do you want to suggest out what our options are, (Karen)?

(Karen): So for voting we would first ask you to vote on evidence looking at the high, moderate, low or insufficient rating. We'd ask you to kind of apply that algorithm and apply our criteria.

Looking at the evidence submitted and whether or not it's supporting the specific intervention that it's being measures, in this case, the documentation, it sounds like the committee will probably lean towards insufficient.

I don't want to assume that, but assuming that is the case and the majority does vote insufficient, we'll then ask you if you'd like to apply the exception option.

And if that is the case, the committee agrees to apply the exception option, the measure will pass evidence and we'll move on to performance gap.

Female 1: Okay. Before we take a vote, I just want to ask (Heidi), (Lenora) or (Steve) if you all have any comments you'd like to add? Because I think when we discussed this the last time, and I know, (Lenora), you weren't part of it, but

your opinions really helped us think through this. So I'd really like to make sure you have a chance to weigh in.

(Steve): I'll - I can jump in because I was involved when we were discussing this on the phone and, you know, this is a, I think, pretty critical measure for patients and I really appreciate, you know, (Katie's) comments about a way to proceed because I really feel like we need to have, you know, some more discussion about it and I would like to get past the evidence part of it.

And I fully understand how it came up to be insufficient, but I would certainly - I feel pretty strongly that I want to move forward past the evidence to have more discussion.

Female 1: (Heidi), (Lenora), do you have anything to add?

(Heidi): Sure. I'll jump in too. My concern is that the statements were made that questions being asked to people who were not - I forget who was on the phone who said, you know, she wasn't - the patient was not an active cancer patient and her - she was responding to some other pain, it was kind of a holistic thing.

You can be a not active cancer patient and still in pain, you know, and it could be a residual from surgery. It could be from anything. So I think everything kind of needs to be taken into perspective, but I don't know how those questions can be asked more directly, you know...

Female 1: Right.

(Heidi): ... cancer-related pain or something as - you know, if you make it more directed.

Female 1: Well, and...

(Heidi): It might be more beneficial.

Female 1: Yes. And when you're in cancer treatment that your (unintelligible) is your physician. And so if the pain maybe it isn't because of your cancer, but you're there and that's just help to treating you more holistically.

(Heidi): Yes. And...

Female 1: We should be…

((Crosstalk))

(Heidi): ... opinions from different - all different medical, you know - when I was going through my chemotherapy, I was pregnant at the same time. So I had several teams of people kind of checking in at once and the pain was really due to several different things, going through chemo while you're pregnant incorporates a lot of different aspects.

So no opioids. So just so you know none of that went down, but nonetheless, I think it's important to kind of look at things holistically and not just a drug (unintelligible) tight lines of just this specific type of cancer and what else is going…

((Crosstalk))

Female 6: As a patient that really what's foremost is quality of life. If this pain measure - so it's going to affect my quality of life where I can't work or I can't

effectively involve myself with everyone, then it's an issue as a patient, you know, if you - it's almost, how can I say, a catch-22 because if you solve my pain issue then I may have a side effect from it. And if you don't solve my - I - you know, I'm still having an effect from the cancer or so.

So it's a catch-22 when I - and I appreciate just trying to be able to talk it out to see which direction that it goes.

Female 1: Okay. All right. Thank you. All right. So can we go ahead and vote?

Female 2: Voting is now open for evidence on Measure 0383. Options are A for moderate, B for low and C for insufficient. Voting is now closed for evidence on Measure 0383. We have 3 votes for moderate, 4 votes for low and 11 for insufficient. This measure does not pass evidence. We will now move to vote on insufficient with evidence.

Voting is now open for evidence with potential exception to evidence. Options are A, insufficient evidence with exception and B, no exception. Voting is now closed. We have 16 votes for insufficient evidence with exception, 2 votes for no exception.

Female 1: Okay. Thank you. Okay. So I think we've earned a lunch break. So we can either take a half an hour or we can take 15 minutes to get our food and work - continue working while we eat. Does that sound - yes, working on - okay?

So right out here, 15 minutes, and then we'll come back and continue this. Agree. So hopefully we'll just…

((Crosstalk))

Female 1: Hi, everybody. I think we're getting ready to get started again. Do - hopefully, the folks on the phone are still hanging in or maybe you've gotten your lunch and they are back with us.

So since we've - we (unintelligible) through this evidence that was basically the hard part. I don't know. Or maybe it's the - maybe that was just the beginning.

But the next category to discuss is performance gap. So (Prince), do you want to start us off?

(Prince): So…

((Crosstalk))

Female 1: Can you put your mike on?

(Prince): Oh, I'm sorry. Although the measure is included in the MIPS program, that program is yet to make this (unintelligible) data available.

And so the fellow provided a summary of data from literature that clearly demonstrates that patients with cancer care receive desperate treatment across groups. And the preliminary rating for opportunity for improvement is moderate.

Female 1: Okay. Does anybody have any comments on this or do you all (unintelligible)? No? Do we want to go ahead and vote then on this performance gap? Okay.

Female 2: Voting is now open for performance gap on Measure 0383. Options are A for high, B for moderate, C for low and D for insufficient. (Unintelligible) two more votes. Two more votes. Are the folks on the phone able to vote?

Female 7: Yes.

Female 1: Thank you.

Female 2: Yes. Voting is now closed for performance gap on Measure 0383. We have one vote for high, 13 votes for moderate, 3 votes for low and zero for insufficient. This measures passes performance gap.

Female 1: Great. Thank you. Okay. Next, reliability.

(Prince): There aren't really significant concerns noted, some varying elements about the fact that the measures apply to two different populations, but no significant concerns and the preliminary rating for liability was moderate.

Female 1: Does anybody have any questions or comments on reliability? Okay. Well, we can go ahead and vote on that then.

Female 2: Voting is now open for reliability on Measure 0383. Options are A for high, B for moderate, C for low and D for insufficient. Voting is now closed for reliability on Measure 0383.

We have one vote high, 13 for moderate, 3 for low and zero for insufficient. This measure passes reliability.

Female 1: Okay. Next, validity.

(Prince): Doesn't seem to be any significant concerns. Two members of the committee did identify that they had confirmed, but they just stated yes and didn't provide any information about what their concerns were.

So I would throw it out to those people to share, but initial rating is moderate.

Female 1: If you remember how you answered the survey and you said yes, can you share those...

(Prince): Yes.

Female 1: ... concerns?

Female 6: Or maybe they were saying yes, it's valid.

(Prince): The question is do you have any concerns...

Female 1: Okay.

(Prince): ... of the testing results and two people wrote, "Yes."

Female 1: Anyone on the phone? No? Okay. And I guess we can go ahead and vote on it.

Female 2: Voting is open for validity on Measure 0383. Options are A for high, B for moderate, C for low and D for insufficient. Voting is now closed on validity on Measure 0383. We have one vote for high, 14 for moderate, 2 for low and zero for insufficient. This measure passes on validity.

Female 1: Okay. So feasibility?

(Prince): Feasibility, again, no significant concerns and some varying comments about the level of abstraction that we required to collect data, but most elements are going to be reported in EHR's.

And so original preliminary rating is moderate feasibility.

(Dr. Barse): This is actually is very difficult measure to extract from EHR because there's no place - no specific place for documentation of the plan of the - because I know when we implemented that we had to put additional measures in place. So if any patient of pain that was more than zero it brings up a window, asks the provider for putting input and that's how we are measuring how providers are responding.

You had to build in additional (unintelligible) to actually collect that information. So it's a rather difficult measure.

(Greg): So how - and maybe to say that again, how did you have to document?

(Dr. Barse): (Unintelligible) built a logic in the EHR that if anybody has pain level more than zero, one or higher, then it alerts the provider, a window pops up the minute you open the chart in the EHR that your patient has pain this level and some description. There's (unintelligible) description.

And then it opens up the (unintelligible) but you view the options in terms of what you're doing. (Unintelligible) refill new prescription, continuation of the content monitoring and alternative methods of (unintelligible) five or six options and then you pick one of them and that's how we gather what the providers are doing.

And also making sure they're not missing the pain as you called it.

(Greg): So it's nothing very standard...

(Dr. Barse): No.

(Greg): ... at all?

((Crosstalk))

(Dr. Barse): … developed to build support.

(Greg): Okay.

Female 7: I would just comment what I think (unintelligible) I would just comment though that it's traditionally done through audits and most of the medical records.

And the harder part is that there - that is there a treatment plan that's subject to a little bit of judgment when your abstracting those charts? Like, I addressed the patient by having them come back again or I prescribed pain medicine and (Hopey)'s very broad in what equals a treatment plan.

Having said that, we've had this measure for a long time and able to consistently measure (unintelligible). So I don't think it's going to be electronic anytime soon. You're one of the few places that's worked on an electronic solution.

(Denise): Hi. This is (Denise) on the phone.

Female 1: Hi, go ahead.

(Denise): Can you hear me? Hi. Thank you.

Female 1: Yes.

(Denise): I agree with Dr. (Barse). This is also the way that a number of different centers, those specifically that had to report this as part of the PCHQR program has switched to electronic means due to the requirements.

It is manual and electronic and ends up being more of a checkbox than a plan and the plan is not very well defined to what counts and constitutes a plan. So this has been a very challenging measure to measure consistently.

Female 1: Any other comments here in the room or on the phone?

((Crosstalk))

Female 8: It would seem extremely difficult to get accurate percent of visits when one doesn't know what that denominator really probably is.

(Brad Hirsch): And this is (Brad Hirsch) on the phone. I would just add that I think our - my practice is known. I think maybe the practices have already gone through putting electronic infrastructure in place to capture this.

I know across the oncology network there's been in most - both generations of the electronic health record they ran in and they put in the (unintelligible) cores that are upgraded as well as reviews for documentation in different areas.

So I know that this has already had that impact.

Female 1: Okay. Great. Thank you. Okay. So are we ready to go ahead and vote on feasibility?

Female 2: Voting is now open for feasibility on Measure 0383. Options are A for high, B for moderate, C for low and D for insufficient. Voting is now closed for feasibility on Measure 0383. We have zero votes for high, 13 votes for moderate, 5 votes for low and zero votes for insufficient. This measure passes on feasibility.

Female 1: Okay. So now, we move on to use and usability.

(Prince): So the measure is currently being publicly reported. The (unintelligible) program details is currently used in MIPS and in the PCF-exempt cancer hospital quality reporting program.

Female 1: Okay.

(Prince): And so preliminary rating is passed.

Female 1: And what about usability?

(Prince): And usability, no significant signs. People know (unintelligible) everybody indicates that the benefits outweigh the harms, some anticipated concerns over whether or not there can be a promotion of opioid (unintelligible) but it's recognized that that concern is unavoidable.

Female 1: Thank you. Any comments? Dr. (Barse)?

(Dr. Barse): Yes, I think this issue with regards to the opioid is actually very real and to be very honest with you, I've become more kind of cautious in terms of prescribing patterns.

So in the future if measure is going to go forward, it would be good to build in for patients an incurable cancer to measure this and those with curable cancer kind of be less pursuing the rounds of over-prescribing pain medication.

I think that may be something that should be considered for the future.

Female 1: Yes. (Heidi)?

(Heidi): I think in addition to that I think that's a brilliant suggestion. Terminal cancer is a great option; however, I think also having been a recipient of several surgeries in cancer myself, more details about - instead of just handing you a vial of medicine, here's what the risks are.

It would be great if that could be explained and for someone like myself who's absolutely terrified of taking them, the explanation that it might be okay to take one or two, but not - I mean, there's got to be some give and take and I think more education given to the patient would be really beneficial.

Female 1: Great. Any other comments? Okay. Can we vote?

Female 2: Voting is open for use on Measure 38 - 0383. Options are A for pass, B for no pass. I think we're waiting on one more vote. Voting is now closed for use on Measure 0383. We have 18 votes for pass, zero votes for no pass. This measure passes on use.

Female 1: Okay. Thank you.

Female 2: Voting is now open for usability on Measure 0383. Options are A for high, B for moderate, C for low and D for insufficient. Voting is now closed for usability on Measure 0383. We have one vote for high, 13 votes for moderate, 3 votes for low and one for insufficient. This measure passes on usability.

Female 1: Okay. So next we should talk about the related and competing measures and I think there's several that were listed as - that aren't necessarily - or (unintelligible) to cancer in addition to the ones that we'll be considering right after this.

And do you have any comments?

((Crosstalk))

Man 6: … So the relator or compete measures are 0524, Pain Interventions Implemented During Short-Term Episodes of Care, and 1628, Patients with Advanced Cancer Screening for Pain at Outpatient Visits.

Female 1: Okay. And so part of - I'm sorry. She asked a question. I missed the last…

((Crosstalk))

Man 6: 1628, Patients with Advanced Cancer Screening for Pain at Outpatient Visits. There are two competing measures the way I did them.

Female 1: Okay. So I have a question for the developers on this because we will be looking at the other two measures and you said that 384 is important to - that

there will be concerns if (unintelligible) about this measure and can you please elaborate on that? I'd like to understand that.

But I fully understand you have to ask about pain in order to have a plan in case - plan of care in place, but my question is does there need to be a quality measure for the asking about it if you have the quality measure for the plan?

(August): Yes, great question. I think this is a one anecdote here, but we have heard from someone on the implementation side of the PCH to our program where I think the plan of tempter pain measure remains, but the chain quantification measure is not in there anymore and they're having some more difficult concerns about how they can implement one without the other.

It's not something that we've been in communication with them on. I think it might be sort of a discrepancy between implementation logistics versus the kind of periodical questions do we actually need to individual measures.

I can't answer that definitely for the group. I think, you know, back to when the measures were developed, they were intended to be used together as kind of companion repaired measures.

Technically, they're unpaired now. I think it's our understanding I think part of that may be because there is an e-Measure or the pain quantification measure and we didn't have an (unintelligible) e-Measure for the plan of care for pain measures.

So I think that might've necessitated some of the split. But that's something that we definitely, you know, kind of considered. We've had some preliminary discussion with PCPI about how might we harmonize these, could we create, you know, a unified measure.

That's easy to talk about kind of at a superficial level. There are obviously a lot of considerations for how we would test a new measure, you know, what we would do with the E version of the pain quantification measure.

So something we're definitely aware of and thinking about.

Female 1: Okay. Thanks. Any other comments or questions on related and competing measures? Okay. So what about on the - oh, go ahead.

(Greg): Yes. And so I guess this is just - I don't know if it's a recommendation, whatnot, but - I mean, is the measures are internally related to one another, are these the types of things that would be suitable for making a composite measure?

I mean, it - if it - if it's a paired measure that either one of them requires the other - like, you don't want to use either one of them individually. I mean, that - my impression is that's the whole point of a composite measure, but that's just a though.

Female 1: Yes, I think we definitely could do. I think the question is just procedurally for NQF endorsement and re-endorsement we're trying to - you know, we don't have data collected on a composite measure that we could use for testing in order to kind of meet these criteria.

So that's our sort of limitation there. In theory, we absolutely could do that and it kind of - it does make sense to look into doing. Okay.

So if we're - let's talk about overall suitability for endorsement. Are there additional comments or discussion we need to have before we vote on that?

All right.

Female 2: Voting is open for - oh, I'm sorry. This is the wrong slide. Voting is now open for overall suitability for endorsement for Measure 0383. Options are A for yes, B for no. We're just waiting on one more vote.

Female 3: Sorry. My computer died and I'm having to reboot. So maybe you should proceed without me.

Female 2: Okay. Voting is closed for overall suitability for endorsement for Measure 0383. We have 15 votes for yes, 2 votes for no. This measure passes.

Female 1: Do you want to do the e-Measure first?

((Crosstalk))

Female 1: Okay. Thank you very much.

(Greg): Thank you.

Female 1: Thanks very much, everyone. I appreciate it.

Man 1: Thank you.

((Crosstalk))

Woman 1: …he wanted me to — yes, okay.

So, while everyone changes seats here, I'll introduce the next measure. We're going to go out of order again and look at 0-3-8-4-e, the electronic measure and I'm realizing that they're both paired; 0-3-8-4 and 0-3-8-4-e.

So, before we — the name of the measure is Oncology Medical Radiation Pain Intensity Quantified, and the sponsor's PCPI. Before we talk about the measure, I think that NQF staff would like to give us some preliminary information about the new data that was presented to us.

(Elvia): So, there is some updated data and information that was shared by the developer. Unfortunately, that information was not uploaded in our system correctly, and so the staff preliminary analysis, we used the data from the previous review of this measure, which happened all in 2018.

And so, what we've tried to do, which was an email that we sent out to the group yesterday, is to highlight the updated information within the measure worksheet. That measure worksheet is a PDF document, and so we inserted essentially the new information and comment bubbles. So, our intention was to allow the Committee to see side-by-side the updated information on performance gaps and updated testing information on reliability and validity.

I'd also like to say that this staff rating on the reliability and validity would not be accurate, because again, we reviewed it under looking at the wrong data, and we can talk through that in a little more detail when we get to that section.

And lastly, there may have been some comments in there around data element testing requirements. That is a new requirement that we have requested as of August of 2019. This measure was submitted prior to that requirement, and so that isn't a requirement we'll be looking at again this e-measure.

Our developer is here as well, and so we're giving them some additional time to talk about some of the differences that are present in the data and to offer additional guidance as we walk through this measure.

Man 1: Just for my benefit, as well as others, we have an e-measure and the exact — well, similar — non-e-measure. Would you help us understand the difference and why we would have two?

(Elvia): Are you asking why were there two submitted, or what…

Man 1: There were two submitted, and what's the difference between the two?

(Elvia): Right. So, it's from the TCPI. And so, before the incorrect requirements were just to submit one measure. And then, to submit the information on what the modality was, or whether it was registry claims or EHR.

At some point, the requirements changed and we were required to submit them separately. So, that is when well, 0-3-4 was given in "e" for electronic version of the measures.

So, that, it's partly because of NQF requirements, and partly because the coding is just different. So, while we have the same enumerator and denominator, one has coding for (instruction) from EHRs and one has coding for (instructions) from registries and/or claims.

Woman 1: Before we get started, did everybody get a chance to review the data that was submitted to us last night and understand where there were updates to the data? So, (JET), getting here on an airplane from California.

(Elvia): Correct.

((Crosstalk))

Woman 1: So, I think that one of the reasons that we're so happy that the developers were able to come today is because they can spend a little extra time making sure that we understand the questions.

We're also going to do a brief introduction, but probably we'll really want to spend a little extra time with you also during the statistical (violations).

(Elvia): Yes, absolutely, and at the PCPI, it's a team. At first when we put together measures and as we maintain them, so we do have staff on the phone for testing and for questions about the coding and specifications.

And then I'd like to introduce Dr. Paul Wallner, who is our expert, who will be presenting the measure, and then I'll just add a couple of things after he's done. Dr. Wallner?

Woman 1: You may be on mute.

Is there any way we can ping him or anything?

(Elvia): Yes. He's on the line. Yes, Dr. Wallner, are you on mute?

Dr. Wallner, I believe that it's showing that you have your headphones on. Either maybe the microphone is not on, or maybe you could remove the headphones and use the speaker instead? Yes, I think that would help. Thank you.

Dr. Paul Wallner: Hello. Is that better? I just called in on my phone.

(Elvia): Hi, Dr. Wallner. You scared me there for a second.

Dr. Paul Wallner: Sorry for the problem. I was working through my computer, so I called in on my phone. I must apologize for the delay. Okay.

(Elvia): Very good. Thank you.

Dr. Paul Wallner: Sorry. Thank you. I'm Dr. Paul Wallner, a radiation oncologist and Chair of the PCPI Technical Advisor – Technical Expert Panel or TEP. I appreciate this opportunity — I'm going to turn off my computer; I'm sorry; I'm getting feedback. I'm sorry.

I appreciate this opportunity to open the discussion on PCPI Measure 0-3-8-4-e, and I'll be discussing at the same time 0-3-8-4. And as noted, my PCPI colleagues are available to respond to any additional questions that might arise regarding the measure.

When I finish, I believe that they'll have a few moments to present some clarification, as alluded to by staff regarding testing and rating of the measure.

0-3-8-4 is a process measure. The percent of patients' visits, regardless of age, with a diagnosis of cancer receiving chemotherapy or radiation in which pain intensity is quantified.

There are currently four generally accepted vital signs in clinical medicine: body temperature, blood pressure, pulse rate and respiratory rate. A fifth, oxygen saturation, is sometimes considered and measured.

All of these are easily obtained, documented and followed but, except for acute clinical episodes, are of minimal interest or efficacy – clinical efficacy – to cancer patients who overwhelmingly indicate that pain is among their greatest concerns as the most significant impact on their quality of life and perhaps even as a dimension on overall survival.

In proposing this measure, we believe that for cancer patients, pain intensity must be a measurable and trackable vital sign, and that by measuring and tracking that pain, quality of life and survival of cancer patients will ultimately be improved.

Assessment of pain also requires direct interaction of patients, facilitating a patient-centered approach to its management. Quantification and documentation of pain should also provide a greater rigor in its management, and we're delighted by the reception that the previous measure received regarding pain management.

In 2019, NQF staff analysis raised concerns regarding precise definition of gaps and permissibility of differing measurement tools. We believe that this current proposal clearly demonstrates those gaps and pain documentation and adequately explains the rationale for not attempting to move forward with any significant signal pain metric.

One question that might arise during your deliberations is that of pain quantification. This is critical for cancer patient management. Why is this proposal limited to patients receiving only chemotherapy or radiation?

The TEP is actively considering how all anti-cancer interventions could be included, and we anticipate adding other modalities to the measure in the

future but do not want to delay consideration of the current modalities at this time. We very much appreciate consideration of this important quality measure. Thank you.

Woman 1: Thank you. I think that we'll begin the discussion, unless you have — oh, you do have something to add to the discussion?

(Elvia): Yes. I just wanted to add — just in going from what (Nicole) mentioned — we have submitted just based on some — we had submitted both 0-3-8-4 and 0-3-8-4-e at the same time, but then, 0-3-8-4-e was deferred just because of some questions related to the testing. So, we did provide data to mitigate those concerns from NQF. And we had submitted that information, so we'll go through it, and it has different information than was provided to you before.

So, when we get to those sections, we'll go through more of the detailed information on what they are, as you had mentioned. Hopefully that will be helpful.

Woman 1: Great. And (Shelley) and I both were – conferred with the staff today and felt that the team needed to see the updated data and a thorough explanation of that so we could vote on it because that data wasn't available to us when we did our initial review.

So, as you'll see, some of the recommendations, some of the ratings might – will – change because of the new data. And so, we appreciate extra time from the Committee getting — and your willingness to get new data on short notice, and we definitely appreciate PCPI for helping us walk through that.

But I'll turn it over to the Discussant.

(Rob): I'm not sure I'm appreciative of the late data, but the group is. So, this is quantification of pain, as noted, and with patients undergoing radiation therapy and/or medical oncology. This was previously approved in 2008 and 2012, I believe — or related measures.

There's been significant discussion in the past and in the Fall Meeting. I think basically patients believe this was critical, and there was also discussion of whether or not there was a linkage between asking the question and quantifying it and having a patient outcome.

Although, as noted in the prior discussion, unless you do the test, which in this case is what's your pain level, you can't really create an outcome. And so, this was viewed, and we did an override of a necessity to ask the questions in order to get a patient outcome.

Is that a reasonable summary?

Man 3: Yes.

(Rob): So, that was a problem with was there sufficient evidence. And I don't know whether you want to go into that discussion again. There's been new data presented which, as I look at it, suggests that there is an increasing – there's an increasing gap in performance from the data as presented.

I don't know whether the dates 2016, a decrease in the mean from 2015, suggesting that there is a performance gap. I don't believe we have new data for disparities.

So, where do we go…

Woman 1: Why don't we go back to the discussion of evidence and go to our logical progression. Did the developer want to add some information?

(Elvia): In regards to the evidence, so for the gaps in care, as you mentioned, the updated data that we presented shows quite a bit of a gap in the 2016 data, because as our ASCO colleagues had 2017 data when they submitted the measure, they used the same program data from CMS. When we submitted it, the 2016 was what was available.

And then we see in terms of the gap in care, yes, it is in fact decreasing from 2013 at 82.7% to now 2016, which is the new data that was provided, which is at — the median was, I believe, 60 – in the 60s — so obviously it has increased.

And then, in terms of the evidence, it's similar to what our ASCO colleagues as well, that the empirical evidence is lacking in terms of the connection to the linkage to outcomes. So, I think it's a very similar issue, Dr. Wallner.

Did you want to add anything?

Dr. Paul Wallner No, I think that's correct, that we see a real gap, and we believe that both measures will improve that gap.

Woman 1: So, I'd like to open it up for discussion. Any other questions or comments? I also would remind us that we went through a very similar discussion of the relationship of evidence to this measure and we can also consider the previous way we approached evidence, voting on is the evidence insufficient, and therefore we want to override that with (unintelligible).

So, any discussion? (Jen)?

(Jen): I just had a question. This may or may not be the appropriate time. Was the same sample utilized for both e- and non-e to look at the difference in — I mean, there's no gold standard, but were there significant differences between e-approach and a non-e approach in terms of measurement?

So, as we look at this, do we have any sentiment that one is better than the other, or that there's some difference, pros and cons?

(Elvia): In terms of the reliability and validity, right? I can jump in, unless, Heather, you prefer to do so.

Heather Tinsley: No, go ahead.

Woman 1: Can I interrupt? Since that moves to testing, and we might want to spend more time concentrating on that, if you don't mind, let's put that on hold, and come back to it so that we don't go back and forth and miss our point. (Len)?

(Len): A point of information. Not relevant to the discussion here, but important. I'm just scrolling through this new information as provided and see that I'm listed as a member of an expert panel that was – apparently commented on this back in 2011. So, I'm going to recuse myself from the discussion. It was so long ago that I didn't realize I was on that panel until…

((Crosstalk))

(Len): I guess my name is there. It also has (Paul) being somewhere in New Jersey. I think he's in California now. So, it's a little bit dated, but in order to avoid any appearance of conflict I'm going to recuse.

Heather Tinsley: So, this is it, a (for your interest). So, our policy is that if you had interaction with the measure within the last five years. So, 2011, if I can do some math, it would be longer than that. However, if you feel like you have a conflict or things like that, you can always recuse yourself.

(Len): I don't remember it, so I (unintelligible) you say within five years. Now, I don't know if I'm listed as being a member currently, if that was the case — but I'm not. I've not been actively engaged in PCPI for quite some time. I'm okay. I don't think I have a conflict.

Heather Tinsley: Okay. Well then…

(Len): But I want to make sure everybody — my name is on this list, and so I don't want to cause a problem.

Heather Tinsley: So, maybe I could just confirm with PCPI that you haven't — that (Len) hasn't been involved in this measure in the last five years that you are aware of.

Woman 1: As that I am aware of, no. We have a current TEP, and he's not listed, and the 2011 TEP was one of the measures, the first being…

((Crosstalk))

Heather Tinsley: All right so, then you're good to go.

(Len): Thank you.

(Len): So, we're going to now do the questions for Committee? I mean, that's kind of an issue. What's the relationship between documenting pain intensity and

outcomes of patients in terms of quality of life and survival? So, I guess that's the…

Woman 1: Correct. That's the…

(Len): That's the current thing we're going to talk about.

Woman 1: And then, just to reframe the discussion, (Shelley) should probably remind us again about the discussion that occurred at…

(Shelley): Right. So, I mentioned a little bit about it earlier. And it was specific to 384, not 384-e. But it applies to both.

So, the CSAC reviewed this in June of last year, and as I mentioned, (Sharon) wasn't on the phone, and I was only on the phone and not in person.

And they had a really negative reaction to our recommendation on this measure and were ready to completely overturn it. And eventually we were able to sort of ask them to not overturn it but let the Committee review it.

But even then, because I went back and looked and it was like 9-to-5 (high), and people even after that discussion wanted just overturn our decision. And they were reacting very strongly to the idea of it being a (check-the-box) measure. And that – it was more about that than about concerns about the evidence or about (check-the-box) measure. That it doesn't actually improve care.

So, we did talk about opioid overuse, and that wasn't of primary concern, as my recollection. I don't have — I went back to see if I had Minutes from it — but that was something that we discussed.

The evidence exception was something we discussed. But it really was about this is a (check-the-box) measure. We as NQF should be doing – shouldn't be endorsing (check-the-box) measures. And there was a really strong feeling on the Committee about that.

So, you know, we did — the ultimate decision was ask the Committee to reevaluate this with the related measure that we just approved for the Plan of Care in place. So, I think that — you know, I heard, I mean, that's why I asked a question as the ASCO developers. And I liked their perspective on it too.

Do we need this measure in order to have that measure? From, you know, that one is not as — there still will be some who say that's a (check-the-box) measure, that you have a Plan of Care in place, but at least that's beyond asking the patient if they have pain.

But the question is, do we need both? And I'm really not clear; I mean, I heard that answer. It seems to me from the way you all have described how you're – in clinical practice have built around this is that you are asking and documenting the Plan of Care. So, do we need a quality measure for both? Obviously we need to ask in order to have a Plan of Care, but do we need both quality measures?

And so, that's just something we need to be able to articulate back to the CSAC with whatever recommendation we make that we considered that.

I think they will look more favorably on the one that we just approved, and I think if we come back with the same recommendation then I don't know how

they'll review it. Sometimes it's personality-based and depending on who's in the room.

((Crosstalk))

Heather Tinsley: Can I re-jump in? (Unintelligible). (Shelley)'s actually right. I think the CSAC had a long discussion about — I would sort of say this is an example of what we would call a documentation measure. Something that is satisfied through some test documentation.

What is the role of those measures in endorsement? What's the role of those measures in quality improvement? What's the role of them in our endorsement portfolio?

That's actually a conversation that we're actively having with the CSAC, and I think within our organization, as well, I think it's an important discussion to have. I will say that, you know, our measure — just to sort of be clear, and we will make this again very clear to the CSAC — our measure criteria has not changed. And so, the task before you today is to review the measure as part of our measure criteria.

So, the fact that it is a documentation-only measure is not something that should come into play unless you want to ask other specific questions as it related to our criteria.

This is also — you know, to (Shelley)'s point, I think this will be something that we'll also make very clear to the CSAC — should you decide that this is something that you want to recommend for endorsement and it goes through the process, we will also make that very clear to the CSAC that our criteria

has not changed. If it meets our criteria, that's, you know, they have an idea to talk about it. But if it meets our criteria, they should be thinking about that.

So, I just want to make that clarification and just, you know, also just to sort of say we've talked a little bit about sort of larger, big-picture items in our measurement. I think this is one of them too that we'll be sort of thinking about in tackling. But just wanted to clarify, please review the measure as part of our criteria and look at it through that lens.

(Elvia): So, I also listened in on the CSAC discussion. And just, I wanted to add just a little bit of a nuance to what you guys have mentioned.

One of the things that — it was not only (check-the-box), but I thought they were also reacting to the fact that it was on its own. So, if you assess pain, great, then what? Without the benefit of knowing that then once you assess pain, that informs the other, the second measure, 0-3-8-3.

So, I think that was also part of it, that it was only one measure, yes, so that was, yes, that was an issue.

(Shelley): Yes — oh. Add to that there was a lot of philosophical discussion about having a less-than-ideal measure is that a step in the direction of getting to the ideal measure? Which, the ideal measure would really be that you manage the patient's pain, not that you'd have a Plan of Care in play. So, we're still even — we're not there.

So, the question is, do we use our resources on measures that aren't getting at the — that aren't measuring what we actually care about? Or is that – do we take that as step in the right direction or does continuing to endorse measures that aren't what we really want to get to crowd out, I guess, the resources that

could be used toward developing the ultimate measures that are more important. And that was a philosophical discussion that the CSAC had. I'm not sure that that's what our question right now, because we're looking at the criteria, but that was the discussion they had.

And there were some who felt strongly that as long as we continue to endorse measures that aren't achieving our actual outcomes, we are using resources there that could be used elsewhere to actually create measures that are more meaningful to patients, or that are better at capturing the ultimate outcome.

Heather Tinsley: Yes, I would agree. I think that was the discussion at the CSAC. I think that's actually the discussion on the larger kind of issue about documentations that we want to have with CSAC, I think.

Are there other instances where on documentation measures do make sense? What are those criteria? Those are all kinds of things that we will be talking about with the CSAC. So, definitely those are things that would raise big-picture items, and if you want to — if you have an opinion, come talk to me at the break. I'm happy to take all opinions.

But just want to make sure that you're, you know, being fair to our process of reviewing it.

(Shelley): And just to add, as ASCO mentioned, we did discuss — we had an initial discussion — about that issue. In fact, in about whether it would be better for us to put our heads together and bring these measures together and maybe see what others – what other elements could be included to get to those outcomes.

Dr. Paul Wallner This is Dr. Wallner. Can I make a short comment about this issue of the philosophy?

I think to suggest that this is simply a (check-the-box) measure is incredibly simplistic, and presumes that it's a yes/no, a bi-modal answer. And that nobody's doing anything about it.

Well, this measure is quantifying the pain, and by definition, when one quantifies it, one effectively must act upon it. It's hard for me to conceive of a physician or health care provider who would quantify, depending on a zero-to-ten scale, let's say a 5-6-7 or even above and not act upon it directly.

So, we believe very strongly that by documenting the measure of pain, the metric of pain, that this will lead to improved quality of life and improved outcome because it will drive the management.

Woman 1: Thank you. Well, I think we have some discussion that we can get started on. (Stephen) has been ready to make some comments. So.

(Stephen): So, — and I say this acknowledging that I voted for both (RAS) testing and (the e) – the other one. Given what we did with 383, it would seem to me that you're going to end up having to do this step to accomplish that.

And so, the only thing that I think's in here that isn't necessarily already almost explicitly encompassed in 383 is the scoring – the scoring the pain. Because 383 just says whether they have pain or not.

And so, I think that's the crux of 384 is whether you think this specific quantification at some discreet level — although I would argue that most plans have some level of discreetness within them.

And so, I guess that gets back to, at some level, this philosophical question of if things duplicate each other but they're valuable, do you approve both? And my instinct is usually try to be as parsimonious as possible because of the burden of compliance.

And so, from my perspective, and then perhaps this isn't the point to talk about it at the evidence level. It might be the point to talk about at the harmonization level — that while this standing on its own, I probably would have voted for. As a pair, I almost think that the other one subsumes it because it's so inherently part of it that you couldn't do it without it. And so, my instinct is that this probably — is it worthwhile, given that you do have 383?

Woman 1: Any other comments?

Dr. Paul Wallner I mean, I'd like the clinicians to join in the discussion, because they're the ones really involved. Could you have a Plan of Care without quantifying the pain upfront?

Woman 1: Go ahead.

Woman 3: So, I'd like to say something. So, it's these processes are dependent on each other, but executed by different members of the team. The team assessment is usually done by the frontline staff – nursing and nurse assistants – who (know) the patient. The Plan of Care is usually done by physician extenders.

The second part can — so you can under-detect pain if somebody's not bringing it to your attention: listen, this patient is reporting the pain. So, these two are dependent, but kind of in different pathways. And what I was thinking

is that — and I commented about this earlier today — is you really have to think what and who we are measuring.

When we are measuring the Plan of Care completed, we are measuring a provider. But when we are measuring whether the pain is assessed and documented, we are measuring a health system. So, these are two different measurements: inter-related, but of two different things and different processes. And as much as they're related to each other, they're not exactly the same thing.

Woman 4: So, I'd like to comment as well. I agree that you're measuring is the system intact and is the physician responding? An unintended consequence is since the physician needs to respond to create the Plan of Care, and this is just a personal experience talking to physicians, why present that data when frequently it's hard to interpret that data.

The reason we don't automatically want to quantify pain is pain is a very subjective issue. As (Jen Mainlen) pointed out, you can have pain for unrelated causes that need a completely different kind of intervention.

So, it really starts to — once that data gets presented to a physician, the physician has to respond and act to that pain. And the unintended consequences to disconnect recording of pain from physicians' responsibility. So, I just don't see how you could decouple those activities, and that's — as a treating physician, I'm giving you that perspective.

(Jamie): This is (Jamie) with the PCPI, and one thing that I just wanted to point out for the Committee, if you don't mind my interjection here, is just that for the measure that we have here on the screen, 0-3-8-4-e, this is a part of all the EHR data source, and 0-3-8-3 and 0-3-8-4 are reportable via registry or CQM.

So, the — I think perhaps the point of pairing is a bit more logical from that particular data source, but for 0-3-8-4-e, there is not an e counterpart for the ASCO measure. So, just wanted to bring that to the top of mind of this Steering Committee for this discussion.

Operator: There's someone on the line with a comment. (Greg)?

(Greg): I know it was already covered — sorry. I left my hand up.

Operator: Thank you.

Woman 1: (Jedda)?

(Jedda): Definitely, you know, concerns about what are we measuring and whose process. But I would suggest that we have lots of measures that have multiple individuals involved in the outcome or the actual measure itself and impacting it.

\

So, maybe it – you know, – seems like there's plenty of reasons that one would put them together because very seldom are any outcomes solely dependent on one single type of provider, though it certainly has recognized its reliability in that.

(Danielle Zernicki): I have a comment. This is (Danielle Zernicki).

Woman 1: Yes

(Danielle Zernicki): The only comment that I might have is regarding how otherwise would a treatment plan be executed if we didn't know the severity of pain? I'm sure it would be documented somehow, maybe outside of a quality measure.

However, treatment — pain level would dictate the actual drug treatment, so that's kind of why I think uncoupling might be difficult.

Woman 1: Okay. Any other — (Len)?

(Len): This is a question. The intensity, obviously — maybe I'm missing the discussion — but intensity plays a big role here. Multiple different people evaluating pain have different interpretations of intensity. I'm assuming is there a validated pain score, aside from the smiley faces that we use? And because you get into the situation clinically where there's differences of opinion about pain and intensity.

So, I'm just curious about how that — or maybe it doesn't play into this at all. But the word intensity, absent of measure, you know, is just sometimes problematic. I think the concept is very real. I think the issue that docs talk about is — I can't speak to nurses; you may have more experience on that one — but it's a problem.

Woman 4: And PPI, there is a designated — you know, out of (NOLSAC) and so on. But to your point, what measurement for intensity is it, is that across the board? Probably not.

(Len): There are validated measures that can be used in research settings, but you know, just in the routine clinical setting, I'm just scratching my head a little bit.

Woman 1: PCPI, is there — I think it makes reference to there's validated scales to use, but I don't think it dictates that a validated scale needs to be used?

(Jamie): Right. Because different of physicians have been using different tools. We leave it up to them. So, as not to — as you mentioned — not to force the use of one over another one. And then have an impact on the workflow.

Woman 1: And (Robert), you had a comment?

Woman 3: So, if one is going down the route of measuring the pain severity, which is a valid thing to do, why pursuing that what did you call it in the chart and not pursuing a patient-reported outcome? I mean, that to me sounds like the measure to pick. Patients are reporting, providers are responding. And it kind of puts this measure in a very different space. Why not pursue that route?

Woman 1: Right. And that's an important issue, I think, and I think what you're talking about is a patient-reported outcome performance measure. And where this one obviously uses patient-generated data, but that it's interpreted by the physician, and then it's acted upon by the physician or the team.

So, I think we're not there where we have developed PRO-PM. That's certainly something that we could certainly consider and Dr. (Anwalder), who's the Chair of the TEP, may wish to bring this before the technical expert panel, or we can just determine with ASCO whether that's a possibility. There is someone on the — (Jody)?

(Jody): Yes, hi. I just wanted to comment. I don't deal with cancer pain as often as I do with surgical pain, but I absolutely agree that quantitation is tough to do. And patients will give one number to the nurse, another number to their family, another number to the physician.

It's very subjective, and I find talking to the patient is more valid than any quantitative (pain). So, I never felt that that quantitation was particularly effective, and I agree with the statement that it is part of the workflow.

Patients come in. It's the nurse. It's the M.A. who asks those questions and says where would you rank your score? And half the time when I confirm that with the patient, it's not the same score. So, I think they can be uncoupled. I think that you can address the patient's pain without having a number.

Woman 1: (Len), did you have a comment? I saw you reaching before that.

(Len): You know, I think the question about PROs is really critically important. And in some of our worlds, they're becoming more important, and systems are being put into place that do that. And this is an e-measure that doesn't capture that. And some people may rely on that PRO. There's no — I mean, it's a rhetorical question.

I think, however, I do think when we talk at the end about potential hazards or risks, whatever the measure, I think we have to acknowledge the fact that the intensity is a question that's been raised. But more importantly, that going forward, we're going to have different metrics, going into play, and as the clinician, whoever they may be, who's acknowledging that, are they going to just list the PRO, which doesn't get into the e-measure, right? Not yet.

So, anyway, I think it's just one of those hazards that has to be that we're aware of, and we ought to immortalize that again just for future consideration, because this is not, this is going to become more important as we go along. I mean, it's important now. It's going to be become even more important.

Woman 1: (Robert)?

((Crosstalk))

Woman 1: Go ahead.

(Robert): I was going to ask whether we have any more comments from the patients. Because this is not necessarily about treating pain; this is about asking the question and how critical is that, which is a little different. Which is a different issue than treatment.

Woman 1: (Stephen) or (Steve) was…

(Stephen): Yes, and I can really just think, you know, about my personal experience. I have found the process of quantifying to be a valuable assessment for myself, personally.

I'm not sure where that goes — that qualifying goes with clinical staff, but I know with me just as a comment, it's caused me to internally think about what does that mean, and what are the implications of that and what does that mean for further healing from there.

So, I tend to be analytical person, so quantifying everything in my life makes sense. And so that helps me to think about it. That's really probably the first thing that comes to my mind.

Dr. Paul Wallner This is Dr. Wallner again. Can I make a quick comment about the outcome issue?

I mean, from my perspective, the only important thing in life is the outcomes, really, in some respects. But the suggestion that this is not an outcome-related

measure, I think, is really incorrect in that the first time its measured, this is not an outcome measure; it's the benchmark, essentially.

But every subsequent metric — and this will be measured on the continuum of care — every subsequent measure is, in fact, an outcome measure of that pain management. Whether it's getting better or worse, the patient is reporting that outcome.

Woman 1: Mm-hm. Thank you.

(Ben Moffis): Hi. This is (Ben Moffis). I just wanted to add to that that in some ways this is a type of a patient-reported outcome because we are asking the patients for their level of pain intensity, and in that regard, I think that it is beneficial and helpful as a guide to kind of clinical management. We do use it all the time in our clinic.

Woman 1: Oh, we lost you.

(Ben Moffis): Oh, sorry, can you hear me now?

Woman 1: Yes, yes.

(Ben Moffis): I was just saying that in some ways pain intensity is a type of a patient-reported outcome, since we're asking it directly from the patient what their level of pain intensity is. And I think that certainly in our clinic, we do use it routinely, and it has been helpful to guide management.

Operator: Mm-hm. Thank you. And (Heidi) was going to comment.

(Heidi): I always thought it baffling that I was asked that question every single time I went in. You know, where's you pain on this scale with little happy faces, and I was asked by different people at different stages, simply going into the office, you know, at the front.

And there's so many variables. I may be in pain standing in front of you right now at this desk, but if I go back and sit and wait for my oncologist to come in, things might have settled down. So, my answer is going to be very different to him than to the nurse that had asked me five minutes before when — you know, sometimes I just need to sit. So, there's so many variables to be considered.

So, I think the outcome is crucial because literally by sitting down for five minutes, my pain might already – might have been abated. And I don't — I'm trying to speak for there are so many different variables that need to be considered, and it's just crucial.

Operator: And (Stephen), did you have a new question?

(Stephen): So, two things. One is just as it pertains to, you know, the measurement. And this is, I think, what I was getting at, and I think what (unintelligible) getting at as well is that, at the end of the day, a physician is going to want to do something about the pain. Whoever records it, they're going to have to confirm it before they enact a plan.

And so, what we want to do is make sure that at some point the physician mentally checks in and says, "Do I think there's pain?" Because it doesn't — if they put on the vitals sheet, and they don't process that information mentally, and they move on, it could say 8; it could say 2. If the physician doesn't read it, it doesn't matter, right?

And so that's where I kind of feel like this measure is a little further away from the one that we just did.

As for the PRO, I think the only one caution I have about that is as it pertains to how like reporting and quality metrics around the pain PRO, and I think in surgery we're particularly sensitive about this is that we have a number of surgeons with — during the height of the pain is a good vital sign.

Feeling very pressured to give extra pain medication to drive people's pain scores down to zero, because they were being told, "Your pain scores are higher than other surgeons' pain scores."

And being told that that made them a poor surgeon. That's something in surgery that we responded to, and in large part, generated a number of these issues that we're still dealing with today. And it's taking decades to bring back those prescribing patterns so that we inculcated them in our trainees during that time period, that you will give this much pain medication. You will try to get our pain scores as low as possible.

And that (tracks) on us as people who train people that we overcorrected it. Now we have to make sure we don't under correct on the way back. But it's just a cautionary tale about a PRO related to pain and how you report that.

Operator: There's (Denise) on the phone.

(Denise): Yes, hi. I'd like to make a comment regarding the paired measurement. I wonder if the denominator of the Plan of Care stems from typically the paired measure, the numerator of the paired measure — the pain score documented and whether those have a pain score greater than zero as written.

If that measure is removed, in either the e-measure form or not, would that change what the denominator is on the Plan of Care measures moving forward and thus make them less reliable? That's my argument for keeping them – the paired measures.

Operator: Thank you. And anyone else on the phone? Okay.

Woman 1: Developer?

(Elvia): Yes, so I was just thinking about the quantification versus the yes-or-no. And I think someone on the phone mentioned that, you know, if the patient has pain, you treat it, and if they don't, you don't.

But then for the question for the clinicians and everyone, if it's just yes, the patient has pain, then wouldn't the treatment be then yes, let's treat? And maybe then the treatment would be the same for anybody who has pain?

Rather than if you quantify it, then, you know, that if it's a low-level pain, the treatment would be very different from a higher-level pain, which is where opioids might come into play. So, that's just something that about the (possibly specific) with one of the patients.

Operator: Any comments from any of the clinicians?

Woman 6: As a clinician, I'll comment. I think that pain needs to be quantified in a (satisfier) clinician. And at least in Florida — I'm sure many of the other states have adopted this. We've had a lot of focus about narcotics prescribing mandated by the State, including improved education for patient, improved

documentation, quantification of the pain, and improved discussions about alternative plans for treating pain.

And without quantifying pain, all those measures become (check-the-box) measures. I'm going to give you a pamphlet that I'm — because I'm supposed to, rather than I'm going to really assess alternative treatments for pain beyond narcotics.

So, I just think that quantifying pain is a very important aspect of managing pain and decoupling from responsibilities for developing a Plan of Care is potentially hazardous. So.

Operator: Mm-hm?

Woman 3: The one comment I would make — I do agree quantification of the pain is important, but quantification of the pain by the person who is treating that pain to put all other dimensions in place.

If my nurse comes out of the room and says, "Your patient has pain level of 9," I'm not going to take that and go based on 9. I'm going to walk into that room and try to get to the bottom of why the pain is. I would give you an example.

Last year, about this time, I have a patient who has been my patient for three or four years. And came in with severe abdominal pain for a survivorship visit. He had a perforated appendix. Okay?

So, you know? It wasn't just the pain. It just been, he came to me, oh, he has this bad pain. He had come to me to me to do other things to get to the bottom of that.

So, I think, yes, pain should be quantified. The question is who should quantify it and how it should be done. I think the importance of the pain is pain needs to be recorded for every cancer patient. Further details, I think, should be left to the person who is going to take charge of treating that pain, whoever that is.

Woman 1: So, have we exhausted the discussion? I let this go on because it's going to be a similar discussion for the next measure that we look at, so that we feel like we've decided how to approach this.

I think it looks like everybody in the room is nodding that it's okay to move on to a vote. And any — I'm assuming, not hearing, seeing, any hands raised online it's okay to proceed with the vote. So…

Operator: Voting is open.

Woman 1: So, (Ben) just raised his hand before we vote.

(Ben Moffis): Sorry. I just lowered it.

Woman 1: Oh, okay. All right. Good.

Operator: Voting is open for evidence on Measure 0-3-8-4-e. Options are A for Moderate; B for Low; and C for Insufficient.

We're just waiting on one more vote. Voting is now closed for evidence on Measure 0-3-8-4-e. We have 11 votes for Moderate; 6 votes for Low; 1 vote for Insufficient. This measure passes on evidence.

Woman 1: So, now we need to address the gap in care. And so, (Robert), I'll go back to you.

(Robert): Let me find the updated info. Well, you did demonstrate a gap. I think what was interesting was the gap was increasing. Now, can you go over the new data? Is it a different source? Is there some explanation about why there's a change over time?

(Elvia): So, for the (P-curve) of data that we provided for the previous three years where we just provide a percentage, that was from the experience report that CMS used to publish for (PQRS), and that is across all modalities. So, they never parsed out whether it was the EHR, whether it was a registry. It was just the measure had this performance rate. So, it had decreasing performance.

And then when we did the testing, it found it even lower. So, I found the 69% is the mean performance rate, where the median is 80. So, for the EHR, it is decreasing.

And why that would be, I just based on the CMS data, I'm not exactly sure why there's a — you would hope that there's an improvement, but they're just — yes.

Woman 4: Any thoughts there's been increased reporting?

(Elvia): Pardon?

Woman 4: Any thoughts that the e is resulting in increased reporting, to where that could account for lower…

(Elvia): That could account for lower, yes. And then, the registry version as well. We always note — and I think, maybe NQF staff can attest to this — that we do notice that for all of them.

But maybe there's a certain complexity to the EHR version of the measure, in terms of reporting, but these are self-selected measures as well. So, when we find the key QRS data, sometimes we take it with a grain of salt because, you know, physicians will select — and I think, anecdotally, this has been certainly established — that they will select those that they know will do well on.

Woman 1: Go ahead. Well, any other comments, is what I was going to say.

Man 4: And there were prior comments — this was somewhat 3-8-4, and I think there was thought from the prior data that this was topping out, but I don't know that's now the case. Any other comments on gaps, performance gaps to be addressed?

Woman 1: Any comments? No?

So, I would — anybody online? I don't think so. Why don't we go ahead and move forward with voting on Is there a performance scale?

Operator: Voting is open for Performance Gap on Measure 0-3-8-4-e. Options are A for High; B for Moderate; C for Low; and D for Insufficient.

Voting is now closed for Performance Gap on Measure 0-3-8-4-e. We have 2 votes for High; 14 for Moderate; 2 for Low; and 0 for Insufficient. This measure passes on Performance Gap.

Woman 1: So, we'll move on to Reliability. And – but there were some questions about the data. So, I might — why don't you introduce it. But I do think PCPI has to walk us through the new data.

(Elvia): Heather, I'll turn it over to you to walk us through the data.

Heather Tinsley: Thank you, (Elvia). Yes, so looking at we received national data from the PQRS. And we conducted a signal-to-noise analysis for the reliability. And for providers that had at least one eligible patient, the reliability came to be 0.96.

I think the original data that you saw might have been — I think that was at, like, 10 or more eligible patients. But we did revise it in talking with NQF. And so, this is looking at pretty much all patients.

And I don't know if there are any specific questions about this since the questions previously aren't going to apply to these data.

Woman 1: So, I think that the Number 1 question was testing in multiple EHR sources. So, you need to comment on that.

Man 4: Yes, because the staff reliability and validity was insufficient.

Heather Tinsley: Yes.

Man 4: So, how has that been addressed?

(Elvia): And Heather, I'll jump in. So, the two questions when we followed up with staff on those Insufficient ratings when we first submitted the measure, we

wanted to sort of find out why. And one of them was that we had not – we had not provided data from — it was not evident that we had provided data from at least two EHRs.

And then, also, that we had tested at a patient threshold of ten-plus. So, physicians who had ten or more patients who fit into this measure. And the reason that we did that is because if a physician only has one patient, and then they score 100%, that might skew the results a little bit. So, we decided to go with a ten-plus to provide a more accurate reliability.

So, then, we changed that's what is included in this testing. We did the retest — we tested at the one-plus, just based on the recommendation from NQF. And then we also included the information that since it's from CMS data that we provided this analysis, the CMS does get reporting from various EHRs including Cerner, including EPIC. So, it's certainly more than two EHRs. I believe there was about seven or eight.

Man 4: Did the staff have any comments about how they would score this now?

Woman 5: We did have a chance to look again at this measure with the updated data and had agreed a reading of Moderate.

Heather Tinsley: Yes, and I'll just add, I think, you know, the clarification we were hoping to get from PCPI is how the PQRS EHR data set may differ from actual data from an active EHR system. I think that was what we were having trouble identifying there.

And then the other thing I just wanted to point out, we do have one of our EC2M experts on the line, (Chris Malecki), happy to answer any questions

about the technical review of this measure, including any questions around the data elements that we identified as having some issues with.

(Elvia): So, in terms of the first question about us providing data for a mix from PQRS, which includes a mix of EHRs, as opposed to one, it's also just the testing data there are available to us, just PQRS is really where we would go to, and we have at times difficulties finding — we would have to go, rather than go to the vendors themselves, we go to clinics.

And if they use Epic, let's say, or Cerner, or NextGen, a lot of times they don't have the resources to conduct the testing. So, we would not be able to provide then, therefore, the data on their specific EHR. So, PQRS data, again, does not parse out which e-charts that they are providing the data for. It's just an aggregate.

Woman 1: So, we thought it was important that the group opportunity to discuss this because the rating was based on data that wasn't correctly metabolized after it was submitted.

And we want to make sure that everybody understands that the original Insufficient is Moderate at this point and does not have the influence of the original assessment for today's discussion. And so, knowing that we have a short time to assess that, any other questions that might come from the group are critical at this point.

I think in the room, it looks like everyone feels comfortable with their answers. Anyone else online?

So, are we prepared to vote? Just the one part. For those online, we had a brief interruption in attendance here, so. Technology, right. That's why e-measures

are hard to assess. But sometimes it's the human element of the e-measures, right? This was a human element, right.

Reliability.

Operator: Thanks, everyone, for your patience. Opening our — voting is now open for Reliability on Measure 0-3-8-4-e. Options are A for High; B for Moderate; C for Low; and D for Insufficient.

We're waiting on just one more vote. Voting is now closed for Reliability on Measure 0-3-8-4-e. We have zero votes for High; 16 votes for Moderate; 2 votes for Low; and zero votes for insufficient. This measure passes on Reliability.

Woman 1: So, let's move on to Validity. Do you want to introduce?

Man 5: So, as a new data change, the validity — I'm wondering from both PCPI and then, I'd like the staff opinion.

Heather Tinsley: Hi, this is Heather.

Woman 1: Okay.

Heather Tinsley: This is (Heather Tinsley) from the PCPI. And so, I think the main concern previously for validity was the multiple EHR issue, which we have addressed. It is multiple EHRs within the national PQRS data. There was sufficient correlation with the Breast Cancer Hormonal Therapy Measure. And there were no threats to validity that we found.

And so, beyond the multiple EHR concern, are there any other specific questions regarding the validity?

Woman 1: I think no questions. I guess our same question for the staff.

(Elvia): Yes, just as Heather mentioned, it was our issues with originally rating validity; we were looking at incorrect data, but we were also questioning EHRs. And so, our rating has (Low-to-Moderate).

Woman 1: So, is there any other discussion among the Committee? Anything online? Why don't we proceed with the voting…

((Crosstalk))

Woman 1: Oh, go ahead.

Man 6: So, for the validity testing, I guess I couldn't see where the explanation was for why comparing it to the hormonal therapy measure was considered a good test of validity. I mean, oftentimes, you compare it to something that's a similar type of process or a similar type of measure. That seems somewhat disparate to me.

Woman 1: So, the PCPI staff, why don't you…

(Elvia): Right, okay. So, one of the things that, (it says) in the EHR measure we have to try to correlate — we usually try to correlate it — I don't think it's a requirement — to another EHR measure. And in the base and in the PQRS program, there's limited number of EHRs. So, I think that this was one of the issues — please correct me if I'm wrong, Heather —

Heather Tinsley: No, I believe that is correct.

(Elvia): And then, we also thought that we ought to choose measures that we think are going to be reported by similar types of clinicians.

Heather Tinsley: And just a note from that one. One other claim as consideration as you're thinking about validity. During at least that technical review of the ECQN, we did identify — and a few of the data elements: chemotherapy administration, procedure performed, as it was not captured by two radiation oncology test sites where chemotherapy is not administered.

I'll actually pause there and see if my colleague, (Chris) has anything to add. Just to give the Committee a little bit more context about what we considered.

(Chris Malecki): Yes. So, with that data element, we noted in the feasibility assessment, there were some issues of that element in the – particularly in the accuracy domain. So, we like to ask the developers if this feeds to that impact in any way, the validity, since this element has some accuracy issues.

(Elvia): So, yes, and I think so when we sent out the feasibility scorecard — Heather, did you want to jump in?

Heather Tinsley: Sure, yes.

((Crosstalk))

(Elvia): Go ahead.

Heather Tinsley: Oh, sorry. Yes, when we sent out our scorecards to get at the feasibility information, you know, both the test sites were radiation oncology practices,

and they don't directly manage chemotherapy administration. So, we didn't expect that they would be able to capture those elements, and so we don't think that it is a threat to validity.

(Elvia): So, they are in fact able to capture all the data elements that relate to radiation oncologists. And so, our – the assumption would be then that chemotherapy oncologists would be able to then capture the relevant or related elements related to chemotherapy, rather than radiation oncology, because it's just not something that they would capture.

Man 6: And the data – the reliability is – or validity – is from what data? What date? Is it recent?

(Elvia): The validity data, I believe, is also from the 2016 PQRS.

Woman 1: (Stephen)?

(Stephen): So, now I have a question, as far as who is reporting this because, for instance, in a surgical oncology practice, it's not at all uncommon for them to be, to be seeing me while they're getting chemotherapy or radiation therapy. And so, I'm capturing that in my note, but depending on if I'm in an integrated system or not, it may not be flagged in the EHR.

And so is the expectation — like if I refer them to, say, a neurologist who has pain, because they're having pain, and I think it's neuropathic pain, and I think a neurologist might be helpful or something, but it's during chemo and it's during radiation, who's — at every step of the way, is every single person trying to record this? Or was your intent mainly the treating physician of whoever is doing the active cancer treatment at this time?

Because I think it's a little — now, I'm a little unclear.

Woman 1: (Jamie), did you want to jump in that one? Or Heather?

(Yvette): Right. This is (Yvette) from PCPI. So, for this measure, we require chemotherapy codes and radiation treatment management codes, so only those physicians or clinicians who use those codes will be reporting on this measure.

Woman 1: The treating — so, if you're a surgical oncologist but you don't use necessarily those codes because you're not treating the with chemotherapy or with radiation, then you would not be reporting on this measure; it would not be relevant.

(Stephen): So, now I'm actually more concerned about validity if it's only on the day of those treatments, though, because that's how you're capturing at least the code.

Let's say you're on a two, three-week chemo, you know, cycle, and you come in a week after your chemo cycle because you're — after your chemo dose, because you're having persistent pain. Is that not being captured because they don't have a chemo administration code on that day? Because that would be a prime time to be having pain, if you're like on a Neupogen or something like that.

(Jamie): This is (Jamie) with the PCPI So, this measure for, like, the both kinds of it. We actually have two different populations. And so, the first population that we have that's included is patients who are currently receiving chemotherapy as we can see on the description.

And the way that we define that is that the patient has a face-to-face visit with a provider, and on either side, within 30 days before and after, there is a chemotherapy session or treatment by a CPT or (soma) code. So, it's during that face-to-face encounter that this pain is expected to be quantified, and that is its metastatic base measure.

For the second population, I'd say that it functions a little differently. Only in a sense that we are aware that the radiation treatment management codes, there are some — or at least one in particular; it's 7-7-4-2-7 — that is actually billed for a fraction of radiation treatment management sessions.

And so, we do have guidance within the measure that specifies that it's meant to be reported again and pain quantified at that face-to-face visit with a provider during the course of the treatment. So, does that help clarify at all?

(Stephen): It certainly helps me on the 30 days, because I think that's good, because it captures the, you know, the in-between cycle business. But now, it brings back the question of if it happened within the last 30 days if they got chemo, that triggers the denominator, but they're in my office.

And so I have to make sure that I'm capturing here. Which is fine, because I assess it anyway, and I think everyone who meets them probably should. But I just wanted clarity on it.

((Crosstalk))

Woman 1: You need to use your…

Woman 3: It's only for the providers who have the codes – billing codes – for chemo or radiation. If you're not doing any one of those, you're not going to be part of the measurement group.

Woman 1: And then, (Len)?

(Len): The more I hear — I have to tell you, I'm a little concerned now, because wow, that excludes a whole lot of people who need to have their pain assessed and managed.

There are patients who get oral chemotherapy. There are people who may be offered chemo and decide not to take any chemotherapy who are clearly in pain and need management. There are people who see — who may not, you know, there are situations where the physician may not be giving chemotherapy. They may be giving it in a facility; or a nurse that comes to the house.

Now, we're starting to parse in a way that a whole series of people who should be followed with this measure are not getting followed. The exclusions may be greater than the inclusions. And is that what we want to accomplish? So, I'm now a little more concerned. I'd like to have — I like to make sure that the patients – that someone sees the — asks the right questions and do the right thing.

(Jamie): This is (Jamie) again. Just, maybe to address a couple of the points that you just raised.

In terms of the chemotherapy we do allow, it's not just injections. It's also oral chemotherapy. So, there are — and based upon the codes, those can be applied for a variety of I guess (rapid) administration for that chemo.

The other piece too that we need to actually keep in mind when we talk about measures is that we are, we're trying to measure something very specific because there is evidence and guidelines to support that.

Just because there is a particular population, say, for those who just recently ended their chemotherapy treatment, doesn't mean that they shouldn't have that pain intensity quantified. It just means that — they're not going to be included in the particular measure for the purposes of performance scores.

So, you know, I don't think that anyone on our test or within the PCPI would argue that you shouldn't continue quantifying pain if the patient still has an active diagnosis of cancer or recently completed their chemotherapy or radiation treatment management. But, it's very specific to the measure, so that there are parameters around that particular denominator of population.

(Len): So, I'm going to ask for a clarification here. Patients who receive oral chemotherapy, what code are we going to use? I'm not aware that there's a separate code other than E and M code for patients who are receiving oral chemotherapy. Am I incorrect?

(Yvette): Right, this is (Yvette) at PCPI. So, for this particular, for this e-measure, it is specified using this numbered code. It is — there are two (separate) codes included in our value set. I can give…

((Crosstalk))

(Len): (Unintelligible).

Woman 1: So, if you're going to speak, always use your mic. And we have someone on the line.

(Greg): That's me. (Greg).

Woman 1: Okay.

(Greg): So, I understand the limitations with regards to the validity testing, but I mean, the comparison to the measure for the breast cancer patients just really strains my ability to draw any conclusions from the validity testing. Because, first of all, you've immediately excluded anybody who doesn't have breast cancer from the validity testing, I believe, by making that comparison.

But even if it included all types of cancer patients, I just, fundamentally — you're hypothesis that, you know, providers that quantify pain intensity for patients with a diagnosis of cancer receiving chemotherapy or radiation therapy are likely to perform the same on this measure as providers that prescribe tamoxifen or aromatase inhibitor to female patients with Stage I through Stage III ER- or PR-positive breast cancer — I don't know — I mean, that seems like a very bold assumption.

So, bold that, if you observed any correlation, I wouldn't take that as evidence of validity. I would take that as just a chance statistical observation until there is a substantial additional body of evidence to explain why there would be a positive correlation between these two very different things.

(Jamie): This is (Jamie). I know that (Elvia) has mentioned something before about how we did test it because it was based upon — I think our interpretation and understanding of the NQF requirements is that when they were going to be

testing an EHR measure, we had to also run the validity testing against another EHR measure.

We have recently gotten clarification that that is not actually the case, but that was sort of the sentiment that we had picked up on in previous submission processes, and so that's why I think that the breast cancer awareness — not breast cancer awareness — the breast cancer measure seems to be a suitable option here because there are no — at the time, and even presently — I don't believe that there are any general oncological eCQMs that are part of the program.

So, this is sort of the closest match — particularly because I think we are also looking at the same type, so to speak, population in the requirements for the measure, where we need a diagnosis for both of the measures; we need a face-to-face encounter with both of the measures.

We were looking at particular types of treatment for these patients and within sort of, you know, making sure that there is a critical action performed within a particular timeframe.

So, I do think that, you know, given the understanding that we had with the testing requirements for NQF submissions as well as what was available in the program, this was, that was why we chose this particular measure. And (unintelligible) the breast cancer measure, and then also we had a colon cancer measure.

(Greg): I guess I just, from my perspective, it's like, hey, we compared two things that grew on trees, but one of them was apples and one of them was oranges. And so, any comparison that you — you know, any relationship you see between the two would be hard to qualify as evidence of validity.

Woman 1: Do we have any other — (Len), is that — an old cart? Okay. Any other comments?

I think — is there anyone else on the line? I think for the sake of discussion, I don't know that I'm hearing new comments. I understand the comments that are being made and they raise concerns, but I think we should move forward with voting on Validity at this point.

Operator: Voting is open for Validity on Measure 0-3-8-4-e. Options are A for High; B for Moderate; C for Low; and D for Insufficient.

Voting is now closed for Validity on Measure 0-3-8-4-e. We have zero votes for High; 8 votes for Moderate; 7 votes for Low; and 3 votes for Insufficient. This Measure is Consensus Not Reached.

Woman 1: Okay. So, this qualifies as, I'm told, Consensus Not Reached, which means that we'll continue the discussion on the other Measures — other aspects of evaluating the measure.

(Nicole): So, we had mentioned earlier that for a Consensus Not Reached Measure, it will still move forward for public and member comments. And the Committee will then re-vote on the measure after you've seen the comment section they've released.

And so, we'll continue to go through the other criteria.

Woman 1: (Len)?

(Len): (Nicole), when it goes forward for public comments, will there be a precis summarizing the discussion that occurred here today so people have an idea about what happened?

(Nicole): Yes.

(Len): Okay. Thank you.

Woman 1: So, the next topic is Feasibility. Turn on your mic.

…He mentioned paper. We're not all electronic. Hit your mic, too. Thank you.

(Rob): The measure is constructed using electronic medical records, which I think is valid. We had an issue of two intended items not being available, but I think that's been resolved. Is that correct?

Comments from the Committee; let's see, the EHR issue has been resolved; it's a multiple. So, that's good. Data was all routinely generated. Staff rating initially was low. Has that now changed with the additional information?

Heather Tinsley: I may ask (Chris) to chime in on this. The primary reason for our rating of Low was due to issues with the data element, but (Chris), I'll just defer to you on this.

(Chris Malecki): Yes. So, we still notice that issue with the one data element that we mentioned earlier. That issues on the feasibility assessment for not only accuracy, which we mentioned during the validity discussion, but also they're in the domains for workflow and for availability domain.

Both of those go on feasibility issues for not only just running the measure, but being able to automate it, since it's an (unintelligible). I know the developer spoke to this earlier, but maybe they might have something to add to speak to the sort of feasibility issue.

Woman 1: Does the developer…

((Crosstalk))

(Elvia): So, I'll just say — Heather, one thing that I do want to mention. There's the accuracy that you mentioned and whether all of the data elements are available.

So, when we get to the use and usability, it's the measure has been used, and we do have a sort of a feedback where we get for eCQMs through the various avenues.

And we have not received information about the problems with implementation and feasibility of the measure, and so I'm not sure what else we can say in terms of feasibility or anything else that we can add to what was already provided other than those two data elements.

Heather, did you want to add anything?

Heather Tinsley: I just, you know, reiterate that the reasons those two data elements were not captured in these two tests sites that we used, whether because they were specifically radiation oncology practices.

And so, we would expect that, should these scorecards had been given to chemotherapy oncology practices, they would be able to identify that data

element. And as (Elvia) said, we haven't been made aware of any issues with capturing that data element by those currently using the measure.

(Elvia): And we do get a lot of questions, so implementers do always provide us feedback, but we have not gotten anything on the feasibility. And I know some of the questions for the Committee are — one of them is the data collection strategy ready to be put into operation while the measure is already in use.

And then, are the required data elements routinely generated and used during care delivery. Again, the measure is being used right now, and in fact, at this point, I think — not this measure, but 0-3-8-4 — is now even in position (compare). I mean, it was just added in 2019.

Woman 1: Okay, (Len).

(Len) So, now I'm confused again. You say there's no problem with feasibility, but clearly the reason that this was changed to this radiation and chemotherapy administration is because vendors didn't know how to implement it. The statement is right in the…

(Elvia): Right. So, it was changed, and then from the change there, we addressed a lot of the issues that they had, so the current version that you've received, after we changed it, we have not received any others.

(Len): Has that been in their space, so to speak, so that they can respond to it? We haven't approved it yet, right? Now, I'm confused here, a little bit.

(Elvia): We haven't approved it, but you mean NQF?

(Len) Yes. We haven't…

(Elvia): Yes, so it's in use. So, it doesn't — NQF is for endorsement, not which CMS takes under consideration, but CMS continues to use the measure.

(Len): Okay, so by limiting it, the way it has been limited, as previously discussed — I won't go into that — that has helped resolve the problems?

(Elvia): Absolutely. Yes.

(Len): Okay, thank you.

(Elvia): Yes.

Woman 1: Any other questions from the group or on — oh, I'm sorry, first.

Woman 3: A question. When you get the electronic record, on an electronic, you know, vendors have discreet fields for pain and details of the pain assessment, things you would be scoring, or is it something that needs the development at the implementation site?

Heather Tinsley: (Yvette)?

(Jamie): This is (Janie) and (Yvette). I'm not quite sure that we understand the question.

Woman 3: So, when you get reports on Cerner, Allscripts, Epic, whatever — do all of them have pain assessments and at least scoring of the severity implemented as discreet fields, or do the sites that have those electronic records have to do something to have that recorded?

(Jamie): In order to report on the measure, and it's (hard) what's actually submitted to CMS, it would need to fall within, I think, the parameters of how the measure is actually classified electronically. We do require that there is a specific link so, that's used to capture a type of pain assessment.

And we have a value set for that, and then we do expect that there is a result that's present. So, that thereby requires that there are essentially two different fields in order to complete the requirements for the electronic specification.

Woman 3: Okay.

Woman 1: Any other questions from online?

So, I move that we vote on feasibility.

Operator: Voting is open for Feasibility on Measure 0-3-8-4-e. Options are A for High; B for Moderate; C for Low; and D for insufficient.

Just waiting on one more vote. Voting is now closed for Feasibility on Measure 0-3-8-4-e. We have zero votes for High; 8 votes for Moderate; 8 votes for Low; and 1 vote for Insufficient. This criteria is Consensus Not Reached.

Heather Tinsley: And just to clarify. At this time, Feasibility is not a must-pass criterion, but certainly should be taken into account in determining the overall recommendations or suitability for endorsement.

Woman 1: So, Use and Usability comments?

(Rob): So, usability, the question, "Indicate whether any information on improvement over time and usefulness of the measure, any trend data that didn't reach out performance is changing over time." This has been in place now for what, 12 years. And its performance over time is level or decreasing. I think that's the evidence, although the data may be from different sources.

Comments from the group were — let's see what we have — usable, you know, benefits outweigh the risks. Comments from the staff was Insufficient. This from usability, yes, usability was Insufficient because the developer did not discuss any progress on improvement. And this is an old standard.

So, we have discussion about usability.

(Nicole): This is (Nicole). So, our rating was Insufficient due to there not being information on benefits and harm, and nothing on improvement provided from the developer.

(Rob): Yes, I mean, benefits and harm; I think people are well aware of the benefits, and I think the staff and everyone at PCPI is well aware of the harms.

Woman 1: Were there any other comments from the members? I think we could proceed with the vote. Then we'll vote on Use and…

Operator: Voting is open for Use on Measure 0-3-8-4-e. Options are A for Pass; B for No-Pass.

Voting is closed for Use on Measure 0-3-8-4-e. We have 13 votes for Pass; 5 votes for No-Pass. This Measure passes on Use.

Voting is open for Usability on Measure 0-3-8-4-e. Options are A for High; B for Moderate; C for Low; and D for Insufficient.

We're waiting on two more votes. Voting is now closed for Usability on Measure 0-3-8-4-e. We have zero votes for High; 10 votes for Moderate; 7 votes for Low; and 1 vote for Insufficient. This measure is Consensus Not Reached on Usability.

Woman 1: So, I think any general comments — well, we've talked about the harmonization already, so I'll move past that. Any general comments on Usability to endorse; anybody want to…

((Crosstalk))

Woman 3: I just want to say, my feeling is this is an important measure, but there remains issues with the measurement. And the questions with regard to the validity and feasibility, I think are significant questions that everybody felt like there are challenges with the measurement.

Woman 1: So, right now?

(Elvia): Oh, we don't vote on that because today, I apologize, because we didn't achieve — we have some Not Achieved Consensus, so please explain the process.

Heather Tinsley: Sure. So, since we had Consensus Not Reached on a Must Pass criterion, it will then go or public and member comment, and we will re-vote on the measure during our post-comment call.

(Rob): Yes, I would comment the most interesting information is how this measure has — nothing has changed with the measurement over time, and it's an old measurement. I think that needs explanation, whether it's useful.

Woman 1: (Len)?

(Len): Nothing has changed over time except the definition of the measurement. Right? The definition of the measurement has changed, right? And it has become very much more narrow than it was previously. I think that's important.

Let me ask a question. I'm assuming that there can be no editorial amendments to this measure from today until it comes back. So, it will be the same issues that we've discussed today.

Heather Tinsley: Yes, the classifications of the measure cannot change.

(Len): Okay, thanks.

Woman 1: So, I'm going to vote. We're making an executive decision, and we're going to take a ten-minute break, and then we'll come back and discuss the other related measure.

Heather Tinsley: Okay. I think we should go ahead and get started, because we still have quite a bit left to do and it's 2:37. I know people have flights.

So, the next measure is 384, since we're working backwards in this group. Which would be Oncology Medical and Radiation Pain Intensity Quantified Measure.

And so, we've already talked about the CSAC discussion. So, I will turn it over to our (CC Chair) representative to give an overview. And we do have the option to apply our same voting on the evidence and discussion from the e-measure to this measure.

The others, we'll need to go through separately, but for the evidence, at least, if we would like to, we can apply that discussion in voting. Right, we would not need to vote…

((Crosstalk))

Woman 1: We would need to vote. What we're saying is we could apply the discussion around evidence for this measure, and then we vote.

Heather Tinsley: Okay. So, I'll let you start with an overview.

(Elvia): So, as Dr. Wallner has stated, in his initial presentation for this measure, it's exactly the same as 0-3-8-4-e, in terms of the denominator and the numerator. As I had mentioned, it's the differences are in the coding and the modalities for which the measures are coded.

So, this is the registry/claim diversion of the measure. And it looks up the exact same clinical action, whether it was done. I will say that the gap in care for this measure is a little bit less. It's less than for the eCQM, and that actually is something that we have found historically, that eCQMs tend to have a higher performance gap just from the PQRS data.

So, what we did for both measures is we provided additional information from the literature, which states that it's not – pain assessment is not being done, you know, optimally, being done — the assessment itself.

And therefore, as we have talked about, without knowing whether a patient has, or what level of pain a patient has, then the intent is to have the quantification of the pain in order to formulate the best treatment plan for that specific patient, rather than a more generalized approach.

And with that, I will — I think that's all I can say, just because, again, we've discussed most of this information at the beginning, so I'll turn it over the Discussant, right?

(Len): Kind of like Groundhog Day. Need to lighten things up.

(Rob): So, again, this is a measure that's been approved in the past, and it's up again. It's not just parallel to the prior discussion, but is fundamentally the same, simply different ways of extracting the data. And the same justification as we had in the past, and in relation to the linked one, 2-8-3.

And this came back to us after scientific review committee, again, and summary from this group. Do we see a need to vote again? Anybody want to add any information to the prior discussions, or shall we, can we accept the prior vote?

Woman 1: We do need to vote, so we can apply the discussion from the previous measure to this one and not re-discuss it, and just go straight to voting, if people…

((Crosstalk))

(Rob): Who would like to add any other information, then?

Woman 1: Okay, on Evidence. And remember how you voted last time.

Operator: Voting is open for Evidence on Measure 0-3-8-4. Options are A for Moderate; B for Low; and C for Insufficient.

Voting is now closed for Evidence on Measure 0-3-8-4. We have 10 votes for Moderate; 8 votes for Low; and zero votes for Insufficient. This Measure is Consensus Not Reached on Evidence.

Woman 1: The previous was 11 Moderate…

((Crosstalk))

Woman 1: So, can we vote for – on the Evidence exception that we did on the previous one?

Heather Tinsley: Yes, if we were to go the exception route, we would have had to pass a majority Insufficient reading.

Woman 1: Three and eighteen. So, (unintelligible).

Who is here on the phone?

Heather Tinsley: So, for folks on the phone, we've had the Committee voted on Evidence. It has come to Consensus Not Reached because about 56% — we had a 56% of the group vote for Moderate and 44% vote for Low. So, this Measure – we will continue to discuss each criterion and also vote on those criteria. We will not go through an overall Suitable for Endorsement vote, and the Committee will revote on this measure following the comment period.

Woman 2: Can I just ask a process question? So, how will we — when it goes out for public comment — how is it going to be couched, in terms of the difference where the evidence for the EHR and the difference for the evidence for the — it's the same clinical action, and so…

((Crosstalk))

Woman 2: Yes.

(Jamie): My initial recommendation would be to have the Committee actually discuss more, since the votes did change. And I'll run that by my staffing Co-Chairs. But we will need to give some sort of rationale along with the vote.

Woman 1: Because we applied the same discussion and we got a different vote.

Right. So, so much for the easy (fights).

((Crosstalk))

Woman 1: If there is someone who has changed their vote from the e-measure, versus just the regular 0-3-8-4 measure, if you're willing to speak up and share why you may have changed your vote on evidence.

(Greg): This is (Greg). I'm not saying that I changed my vote, but I'm just saying that I think the explanation for why people may have changed their vote is after extensively discussing it, at length, over a very long period of time, then once again reflecting on the evidence that had been submitted, I could see how somebody may have come to a different conclusion than their initial appraisal.

So, if you're looking for an explanation for that, that seems like the most likely rationale, as I see it.

Woman 1: Anyone else have anything to add?

Woman 5: When we reviewed this last year, we said Insufficient, and half said Insufficient with Exception. And which is the same thing that we voted on for the Plan of Care Measure.

Woman 1: I guess I'd like to open it up for discussion again, recognizing we won't get through all of our measures, but it's really hard for me to imagine that we have the exact same evidence and we voted to apply the discussion, so I want to hear from the group.

(Len): I'm going to agree with what (Greg) just said. What we talked about this time around is much more informed than what we talked about the last time around at this point of the discussion.

So, I, you know — it's just this a lot of information that came out along the way that could influence how one might interpret this. You know, sometimes you want to say — well, I won't make assumptions.

Woman 1: But also keeping in mind that we're just focused on evidence. I recognize your comment and the comment made on the phone, but just don't want there to be any confusion about thinking about the measure totality versus (unintelligible) just yet.

(Len): Well, the other, if I can make a suggestion that you wanted them to make a motion to revote and see if people change their mind because of the concerns have been raised. I mean, there's nothing in Parliamentary procedure, you

want to take a vote to see if you want to re-vote. Or else just make an executive decision to re-vote and see if it changes.

Woman 1: Agreed.

(Len): I mean, the concern is that this time it didn't get passed, and that last time it was Insufficient.

Woman 1: Well, last time…

((Crosstalk))

Heather Tinsley: And we're talking about last time, I mean when we voted on 3-8-4-e, where it passed. And that was just a couple of hours ago. I'm not comparing it to last cycle, right.

(Len): I'm talking about this morning.

Heather Tinsley: I think a lot of the concerns that came out in the meantime were about how it's being used more than about the evidence. So…

Woman 3: Some of the discussion was actually about the evidence. So, like the discussion about the window of measurability, the providers that are being measured are relevant how they're not fully discussed. And we discussed the pure evidence.

So, you can make a case that that better information or further discussion has brought up issues with regard to the relevance of the evidence. Although I do agree that if rules allow…

Heather Tinsley: Well, I still think a lot of those concerns are related to some of the other issues as opposed to the evidence of quantifying pain. I mean, I think that we relayed other issues when — and (Len), you had raised your hand for something.

Man 8: Yes, I mean realistically, it's a difference of one vote, because last time it was just one vote different from this as far as the number of Moderates. I think there's a little bit of change between the bottom two categories, because (what) if they can't remember which one they voted for last time.

But I think to Dr. (Vargas)'s point, I think, as we were discussing it more, I think we did keep going back to well do we think there's a strong link between what's being measured and what we hope the outcome would be eventually.

So, I think you can say that the evidence was re-discussed a few different times. But when you're talking about the linkage, that's a key component of the evidence, right, is there a linkage between what you measure and what you care about.

Woman 1: What did you…

(Elvia): If we, we do not have the option to revote. So, we'll take the votes as (they can).

Heather Tinsley: So, our process is that we give you — so just to kind of clarify how this all works. So, we had a conversation, right, you asked that you apply the criteria. We asked that you apply the criteria. You applied the previous discussion, and then you voted. And so, unless someone can tell me that they voted incorrectly or they think there were some problem with the vote we will need to continue with what you voted.

Man 1: I don't think there was any problem…

((Crosstalk))

Woman 1: Okay. I don't want to be in any place. I have a process that I need to follow but I want to, you know, you can have…

((Crosstalk))

Man 1: … problem.

Woman 1: Okay. We can have that discussion and if we think that there was an incorrect application of the criteria then I think we have an option to think about re-voting but otherwise I will use it then we can…

Man 1: I would agree with…

((Crosstalk))

Man 2: Green usually means go rather than stop. You know, I would move that - there have been other considerations during the course of the vote and I think there may have been misunderstandings and therefore I will request to revote if that is appropriate.

Woman 1: Yes. I guess I would also like to just hear from other members from the committee if they also feel that way or they agree with that.

(Greg): This is (Greg). I agree with that. I mean I had raised a rationale for why someone might have changed their vote but I would say that that would be somewhat unusual.

In as much as all of that information was regarding a different measure or be at kind of the same measure but a different measure nonetheless and so just like some of the, you know, testing was done differently with this measure. I think it has to be considered in its own line.

Like I mean there is no point in discussing this measure if we really think it's just a carbon copy of the old one.

So I think it's both plausible that somebody would change their vote on any one of the criterion for this measure. But on the evidence one that seems like potentially the least likely one.

Woman 1: Okay. Other comments, thoughts?

Woman 2: Just that if that is the case then we will get the same vote. I agree that we should repeat the vote.

Woman 1: And could you kind of vocalize why you think we should repeat the vote?

Woman 2: Because there was a 2½ hour delay and I think I didn't vote the same but not because of any change in opinions, simply because I forgot what I voted.

Woman 1: Okay. That's fair. Okay. Other comments?

Man 3: Yes. And I don't know if this is justification for re-voting but it just seems like that since we carried over that discussion and there wasn't any more

discussion, nothing really changed I'm just surprised that the vote didn't come out the same. I just wonder if someone may have made a mistake when they voted but I don't know.

It seems like it should be the same since we made a conscious statement saying we weren't going to discuss it anymore. We were just going to take the same discussion.

Woman 1: Okay. Other comments?

Man 1: I mean, I don't see harm in re-voting. The reason why they may have been different again could be this error or people rethought the evidence polling for the discussion.

So I think the fun - the interesting question would be whether we should re-vote in E because we now have more discussion. But I don't know that that's an option.

Woman 1: No. That's not an option. But I appreciate the question. So I think what I'm hearing and I can concur to co-chairs is that we think that there was maybe some confusion about the vote and that people may be taking more - it sounds like maybe people were taking more conversation into than just the evidence or that people forgot their vote.

So it sounds like we - in my mind that is grounds for a re-vote.

My suggestion actually though would be for us to have a discussion about the evidence just if there is additional discussion here so in case we actually do have a different vote we can document why that would be. So I am - does that sound good to the co-chairs? Okay.

So I'm going to open it up to the co-chairs I think if you would have a little bit of discussion about evidence, I think that would be helpful both for just a refresher of people's memories as well as also documentation. Good. Is there comments on the evidence that they would like to hear?

Man 2: I guess we're just repeating ourselves to get on the record. I can get back and I can start there. I think my biggest concern about this is just that it doesn't have as close of a link to that line especially on the context of having to read. Three being that you have to have a plan for people who are in pain.

So this tells you whether they are in pain and somewhat how much pain they are in but ultimately your plan incorporates how much pain they are in. And so from my perspective whether it's an E measure or a chart measure ultimately this is just twice as much work to get to the same outcome.

Woman 1: Anyone else? Anyone on the phone has comments on dividends?

Woman 2: Since we are going to remind ourselves why we vote on evidence before. Its again it's a guidelines driven piece of evidence where we had previously discussed that there wouldn't necessarily be a direct link between the evidence and the published evidence in the guidelines recommendation.

And it was a 2A evidence from NCCN suggesting expert consensus with small studies. So I just wanted to remind everybody that it was a guideline rather than measure.

Man 2: Yes. And just to remind the committee exactly what we are voting on what's the relationship between documenting pain intensity and the following

outcomes for patients with a diagnose of cancer receive chemo and or radiation. So is there a link?

Woman 1: I think we did get our comments. Yes, please.

Woman 3: So I just wanted to make and maybe it doesn't have bearing on the information - on this discussion. But I don't recall your name but in terms of the denominator statement for 033 it is all visits for patients regardless of age with a diagnosis of cancer receiving chemo or radiation therapy who report having pain.

So while, you know, you have a preference for all three - A3 just it is based on the reporting of pain from 0384. So I just wanted to clarify that.

Woman 1: Okay. So we will need to vote on that. So the lines are open for evidence on measure 0384. Options are A for moderate, B for low and C for insufficient.

Woman 3: Is NQF still on the line? We haven't heard anything for a while.

Woman 4: Yes. I haven't heard anything either.

Man 2: Well, they sent a message saying they were having technical glitches…

((Crosstalk))

(Heather): This is (Heather). They sent in a message. The message is on the chat saying that they were having technical problems and they'll come back on.

Woman 1: All right.

Woman 2: You can go ahead.

Woman 1: Okay. All right. Hi everyone. Before we got disconnected, we were voting and so bring results for evidence for measure 0384. We have 9 votes for moderate. 5 votes for low and 4 votes insufficient. This measure's consensus is not reached on evidence.

Woman 2: Just to see if we continue going through and then we will I guess have the opportunity to revote after we hear public comment. Okay. So performance gap.

Man 2: All right. So again the data is slightly different than the E version but PQDS testing demonstrated with 251 physicians a minimal performance gap median was 098, the mid however was 080.

And compared to 2013, 2014 and 2015 data there is improvement. There is not disparities data. Well, so there is limited improvement of the staff approach and I think the comments from the committee almost were identical pretty much to importance to measure and performance gap.

There is a gap - actually there are disparities and the literature demonstrates disparities. Where I would assume that there is disparities based on - there is always disparities in patient treatment based on ethnicity. So that's kind of a given whether it shows up in pain or not.

So is there comments on performance gap and disparities?

Woman 1: Anyone have any comments? Okay. (Len)

(Len): The generic comments have probably caused at least some mad perceptions almost everything we've discussed - in fact everything we discussed today because we don't have data on disparities.

And I think it's critical to note that that is a problem and I just want to make that comment for emphasize purposes. We really need to address that issue because somehow its important.

Woman 1: Okay.

Man 2: PQRS do you have any comments? Do you have any comments?

Woman 4: Yes…

((Crosstalk))

Woman 1: Can you switch your mic on.

Wolman 4: Sorry. But they just don't make the disparities on data available. Yes.

Man 3: So…

((Crosstalk))

Woman 4: Oh. Didn't take it personal of course.

Woman 1: Okay. Are we ready to vote on this performance gap then? Voting is open for performance gap on measure 0384. Options are A for high, B for moderate, C for low and C for insufficient.

Voting is now closed for performance gap on measure 0384. We have zero votes for high, 15 votes for moderate, 3 votes for low and zero votes for insufficient.

This measure passes on performance gap.

Okay. Next let's switch to reliability.

((Crosstalk))

Man 1: … staff sits under reliability and validity with moderate - I don't have comments. My initial comments in the document most people thought there was adequate reliability and no concerns.

Reliability testing from CMS was excellent. And in terms of validity testing everybody thought it was no major concern. Expert panel of 31 experts thought that there was 84% was valid for (unintelligible). Any more discussion on validity?

Woman 1: Yes. We are under - well…

Man 1: Reliability. I'm sorry.

Woman 1: And there was some comments from committee members in the survey that there are concerns about it being - the measure being implemented on a consistent basis. Hard to compare. So I don't know if any of the folks who expressed concerns in their responses survey wanted to talk about those.

Man 1: Yes. There was questions by individuals is this group reliability that was raised.

Woman 1: Do you have a comment? Yes. Sure.

Woman 2: Yes. So for the - when we receive the PQRS data we receive both - they don't pass out whether its individuals or whether its groups however we know that there is both.

But without the ability of being able to tell which data are just for the individuals and which data are just for group we had to I think the recommendation last time was that we would have to then establish that this would be for group because then we can't determine what our reliability is for individuals.

Woman 1: Thanks.

Man 1: Anyone else?

Woman 1: (Kate), can we go ahead and vote on this?

(Kate): Yes.

Woman 1: Voting is open for reliability on measure 0384. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for reliability on measure 0384. We have zero votes for high, 16 votes moderate, 2 votes for low and zero for insufficient. This measure passes on reliability.

(Kate): Okay. Validity. Next.

Man 1: The developer did a good correlation of this measure PQRS 144 and reliability was judged moderate and again reliability testing from CMS was good. Any other comments?

((Crosstalk))

Man 2: I'm going to reiterate what we've discussed before this for the record that on the validity and especially to validity the evidence that was presented and the guidelines presented dealt with patients with cancer on a much larger unrestricted population.

And then it became evident from our last discussion and I just don't want to repeat everything. And it's relevant to this discussion that in fact a substantial number of people are excluded from this measure. So a substantial number of patients are excluded from the measure and that represents a concern from my perspective.

Man 1: So it may not be generally valid - reliable for all patients strictly the subset that was tested.

(Len): Exactly. There may be an assumption that we are talking about all patients and the evidence was for all patients but the measure has been restricted in order to - a very select group of patients for reasons that have already been discussed here today.

Man 1: And the select group would be female more than male which is also a potential.

(Jenny): This is (Jenny) with the PCPI. I think what you are referring to is the validity testing that was done for the EHR measure which was the breast cancer

measure. But I believe this validity testing was completed with the 0383 version that's in - that was in PQRS, PBP now 144 which is again it's the same all patient visits for those with age with cancer. With that…

(Len): And I agree with - I'm sorry. But I agree with you. The measure has been changed with what was measured in PQRS. Now we are limiting the measure in those small part to meet the concerns of the vendors who had difficulty measuring them. We had that discussion.

So this measure that we are talking about today is not what was measured in PQRS. This is a different set of patients with all sorts of different issues and the common set to validity are their exclusions out of concern and my comment is yes. I mean others may not agree but I am concerned about that.

(Kate): Respectfully, I don't think that there is a change. If somewhat they measured in PQRS the data that's there that I think has the same description as this patient population. So it seems they are being measured consistently in the program.

The only difference I think in the way that its currently implemented is that in order to kind of tease out because we heard that we cannot as I think for the conversation that we had for the last measure that radiation oncologist don't report on chemo therapy.

It was we split it into two different populations that would then be combined into a single rate so that we could determine if there was concerns for chemo therapy or radiation or so that we could actually measure those who report on the radiation side if there was maybe more of a gap there than there might be (unintelligible) chemo therapy.

But overall the same performance rate and there isn't a change.

Man 2: (Staph) could you clarify the validity which patient population was being tested?

(Staph): So I'll take a stab at it. So (Len) your question is were the patients since we evaluated this measure the first time at the last session - the spring session of 2019 that the validity testing was different on this measure or that the validity testing is on two different populations between the E measure and this measure.

(Len): So let me try to parse the questions.

(Staph): Okay.

(Len): This measure was initially approved when? Help me out. I can go on the calendar looking for…

(Kate): 2008 I believe.

(Len): 2008. When this measure was approved in 2008 did it - was it limited solely to patients with chemotherapy and radiation therapy?

(Kate): Yes.

(Len): This measure was limited to patients with chemotherapy and radiation therapy in 2008?

(Kate): Yes.

(Len): Okay. Then I have to withdraw that concern. However, the next question is the validity are there exclusions of the patient populations that may be relevant that's listed here as a validity test? And I'm concerned that there is a substantial number of patient populations who are being excluded by these limitations.

So yes, that's - I stand corrected. I regret. I'm sorry, I don't regret. I apologize for misunderstanding. But I still have a concern that all the evidence documents here I think I look at them quickly here I suspect all deal with the population of patients with cancer. That's the evidence basis.

The measure is exclusive to a very limited number of patients. It excludes all patients not getting active chemotherapy and radiation therapy and therefore it's an accurate measure of care.

Now I'm more than happy to hear what other people have to say obviously but that is my concern that this is not measuring what the evidence says we should be measuring.

(Staph): So can I take one stab at it and then we can move on? So I guess my only comment is I think that whenever you're developing a measure by the nature of how you have to develop the measure to find your patient population and focus it you're going to get down to a smaller subset of patients so that you can get reproducible data. So you on any measure development you are going to have some decrease in your generalizability of the data.

And also then over time as the measure is applied the users - I think we have too many of these on. The users are going to give you feedback that helps you continue to modify the measures. So I think that's just a general problem with measure development but that's my opinion on it and I value others opinions.

Woman 3: So I think the issue with the validity of the E measure was the way the validity was assessed. So the way they measured the validity. So they went to two radiation oncology practices, correct me if I'm wrong, right?

((Crosstalk))

Woman 3: But validity was using the same data sets because that's how the confession to the hormonal therapy came up in the validity assessment in the signal to noise ratio in the validity assessment.

Woman 1: Right. So for the validity and again (Heather) jump in if you would like to but for the validity and for the reliability it was PQRS data was used and those data are, you know, they're for just - it was not just for radiation oncology.

Now the feasibility when we needed to find out how to get feasibility as a measure and we sent out a scorecard the ones who were willing to provide us information were from radiation oncology groups. So maybe if I have some oncology medical groups, I don't know whether maybe they were not available. They didn't have the resources to fill this out. They didn't have the time et cetera.

So for feasibility it was the data elements that were missing again because we went to radiation oncologists but for the validity and for the feasibility, we used PQRS data.

(Greg), do you want to comment on something?

(Greg): Yes. So well, I guess I have been trying to avoid comparing this with the previous measures to give it its own due. But because we keep referring back

to the other one, you know, at least for me the main problem with the validity testing for the E measure was that they compared it to another E measure that had no conceptual relationship with what was being measured in this case in my opinion.

But here for the measure at hand that they are dealing - that we're currently discussing they compared it to two measures back the other pain plan of care for managing pain measure. So comparing this measure to another measure, you know, addressing pain in the oncology patients and seeing that there is a correlation I think provides some evidence of validity that was lacking in the previous measure.

So compared to the previous measure I think that the validity testing here has, you know, provides more support than what we had seen previously.

Woman 4: Thank you. I brought up that document from the previous discussion. So in clinical validity co-relation testing was conducted using breast cancer hormonal therapy for stage 1C, 2C, 3 and surgery receptors. So the committee was concerned about how validity was tested not - so this one is actually the validity is exactly.

This one the validity is actually assessed in a different phase so the issue of the validity was how that was tested and nobody could kind of understand the relevance to the comparison to the estrogen therapy.

So I think the difference between tracking this one then I reviewed the documents I did not have any issues with the validity. I think the difference between these two is how the validity of these two documents tested. And I'm reading because I don't remember all of the details (unintelligible).

Woman 1: Does anyone else have any comments on this before we vote? Are we ready to vote?

Woman 2: Yes.

Woman 1: Okay.

Woman 2: Voting is open for validity on measure 0384. Options are A for high, B for moderate, C for low and D for insufficient. We are just waiting on one more vote.

Voting is now closed for validity on measure 0384. We have zero votes for high, 15 votes for moderate, 2 votes for low and zero votes for insufficient. This measure passes on validity.

I think the next here is feasibility.

Man 1: Right. Feasibility and we are looking at the data source for the measure and are there any feasibility concerns. For the prior comments there were no concerns, feasible, limited data collection issues available in most electronic medical records. Staff feasibility was moderate so no major concerns with feasibility in this. Discussion?

Woman 3: You are not asking, are you?

Woman 2: Okay. Any comments on the phone on feasibility? All right. We'll move through.

Voting is open for feasibility on measure 0384. The options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for feasibility on measure 0384. We have zero votes for high, 17 votes for moderate, 1 vote for low and zero votes for insufficient. This measure passes on feasibility. Thank you.

So the next category is use and usability.

Man 1: So this is publicly reported. It's part of accountability program. It's been used in MITS, in PQRS. Preliminary rating by the committee was pass. Most committee comments were positive, it's using multiple reporting programs (unintelligible) using performance results I think were the negative ones. Other comments?

Woman 1: Go for is (Len).

(Len): Can I just clarify? We've talked about some of these measures before and MITS came up in the past and was not publicly reported. This one says publicly reported.

Woman 2: It is now in physician compare.

(Len): Okay. All right. Thank you.

((Crosstalk))

Woman 1: … insufficient for usability the committee comments were…

Man 1: Usability - are we doing use or usability now?

Woman 1: Yes, we're actually talking about both so yes. It seems like the committee comments on the survey were generally positive but staff had rated usability as (unintelligible) can you elaborate on that at all?

Woman 3: Sure. So the insufficient rating is just due to the lack of information that was in the submission form regarding some of the improvement results progress on improvement. Yes.

Woman 2: All right. Let's go ahead and vote on these then. Voting is open for use on measure 0384. Options are A for pass, B for no pass.

Voting is now closed for use on measure 0384. We have 18 votes for pass, zero votes for no pass. This measure passes on use.

Man 1: Usability. One of the issues discussed obviously was risk benefit. There is great benefit but there is however significant risk but risk benefit was rated valid by the committee. And I think there was the comment about insufficient use was just addressed. Any further comments on use? Excuse me, usability.

Woman 1: I think we can go ahead and vote on it.

Woman 2: Voting is open for usability on measure 0384. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for usability on measure 0384. We have zero votes for high, 14 votes for moderate, 3 votes for low and 1 vote for insufficient. This measure passes on usability.

All right. So we are not voting on the overall suitability and I feel like we've discussed relate (unintelligible) quite a bit today. I think we can conclude this

one, yes? Okay. Good. Thank you. I know you had to be the - we discussed it twice on (unintelligible). You can take a break now.

Man 1: We are not going to do this again, are we?

Woman 1: Fortunately no.

Man 1: I may call in sick.

Woman 1: All right. So next…

((Crosstalk))

Woman 2: Moving onto - okay. Okay. Okay. Great. All right. Now we are done with pain for today. No more pain. The next measure we are going to turn to is 0220 other than hormonal therapy recommended or administered within one year for women with stage 1B to 3 hormone receptive positive breast cancer. And the measure developer is Commission on Cancer and I believe Commission on Cancer staff are on the phone to be able to provide an overview for us.

So I will turn it over to - I don t k now the names for the folk from Commission on Cancer but I want to let you all give us the overview of this measure.

(Bryan Taylor): Hi real quick. I'm (Bryan Taylor). I'm joined today with (Ryan McCabe) both of the Cancer Program for the American College of Surgeons. We are both part of the staff measure development team and we thank you for having us.

Measure 0220 do you want me to give a quick overview of this?

Woman 1: Yes, please.

(Bryan Taylor): Okay. So this measure was implemented in the RQF and NCCP reporting systems in 2008. At the time of development approximately 2005 this measure had a baseline compliance rate of 46% which has risen to approximately 92% in 2016 so the measure impacts about 80,000 cases annually from about 15 - 1400 excuse me hospitals across the United States.

All data points are transmitted to the National Cancer Database using the Standardized North American Association of Centre of Cancer Registries Manual. And this measure is a maintenance measure with no changes in evidence since the last evaluation.

Woman 1: Great. (Len)?

(Len): Yes. I have to declare another conflict. I'm not involved with the Commission on Cancer. The American Cancer Society provides extensive financial and logistical support to the Commission on Cancer and the National Cancer Database.

I have had no involvement in any discussion on this measure. I don't perceive that I actually had a conflict but the view of the fact that people should weigh that for themselves I want to make that clear.

Woman 4: Okay. Thank you. And I'm on the Commission on Cancer Advocacy Committee but I have no involvement with their quality measures. So I didn't think to mention that…

Woman 1: Yes. I know. I don't think those are actual conflicts but appreciate the disclosure.

Woman 2: Okay. So starting with evidence (Jeda) would you like to take us through?

(Jeda): I might. And just to reinforce this measure captures and they have talked about the numbers but its 80% of all newly diagnosed so a very broad, you know, cross-section.

One might notice that in terms of the description we are talking about primary tumors and we are talking about either progesterone or estrogen and what therapy is recommended or administered within one year.

So it is common in terms of variety dividends issues there is some lag when you are doing a one-year measure, you know, but to the developer's point considerable, you know, change over time in terms of capturing.

The denominator and numerator data is, you know, succinct. If we look in terms of the evidence it's been indicated that developers provides evidence that did systematic review, talks about no quality, quantity and consistency of evidence provided. I wouldn't mind some comment on that. But evidence indeed is graded.

And if you look at issues over time certainly the evidence has remained important but what we've seen is an expansion in terms of incorporation of, you know, predictive things in terms of R2 positive and negative and how that correlates with ER, PR. So currently the rating for evidence is moderate.

Woman 1: Any discussion or questions? (Len)?

(Len): Point of clarification. I'm looking through the exclusions and one of the exclusions is no surgical treatment of the primary tumor. I'm assuming that

someone had a radiologic procedure with biopsy that that would be - that is intended to be excluded.

So for example, you can have a biopsy, you can determine that there may be a small tumor and for whatever reason you may have surgery and they would like to - since its hormonal sensitive than usually situations. So I want to make sure that that particular type of patient is going to be excluded from this measure and that was given some thought so the word surgical there is intentional.

(Jeda): Excellent question. And it would be great if the developer kind of addressed that because I mean sometimes you can give decisions that, you know, for pathology that potentially (unintelligible).

Man 2: Yes. And similarly Cryoablation now a viable option for small cancers and I don't know how that's being addressed.

(Bryan Taylor): So this is (Bryan) again. So if I understand the question correctly what we're doing here is excluding any case that had a previous cancer diagnosis so it's presumed to be the first or only cancer diagnosis. On the example you gave if that was being done for cancer it would not be included in this measure.

(Jeda): So that quite answers the question you were asking.

Woman 1: Yes. To the developers.

Woman 1: I think the question was if you had - if you didn't receive primary surgical or curative intent surgery and or radiation were you excluded from the study?

(Len): Any therapy - so what this says is that any therapy other than surgery excludes a woman from this measure which is quite uncommon. I mean, you know better than I. But I just want to make sure that's intentional and not something that nobody thought about when they did these exclusions.

Woman 1: So does the developer understand the question, do you exclude patients that might be completely receptive with RB or some other kind of procedure rather…

Man 2: Correct.

Woman 2: Okay. I think that's the concern.

Woman 1: I couldn't identify where that was addressed in here so I think that's a good question which is a good message. And certainly I have noticed they talk about meta-analysis demonstrating a 25% reduction in risk with some evidence going as high as 50% in terms of hormonal and so that's I think a low estimate potentially.

Any discussion in terms of evidence?

Woman 2: Anyone on the phone? Okay. Can we go ahead and vote on (unintelligible)? Voting is open for measure 0220. On evidence options are A for moderate, B for low and C for insufficient.

We are- okay. There you go. Voting is now closed for evidence on measure 0220. We have 18 votes for moderate, zero votes for low and zero votes for insufficient. This measure passes on evidence.

(Jeda): Just a quick clarification. Why wasn't high one of the choices?

((Crosstalk))

Woman 2: Yes. That's right. Before we move on, Dr. (Rosenberg) we had you listed as a conflict on these measures and I know that that our error but just for transparency can you verbally state that you do not have any conflict…

Dr. (Rosenberg): Yes. I do not believe I have any conflict. I do work with oncologist and surgery but not with COC and not on any of this specific measure. Thank you.

(Jeda): I just will say that the evidence is high and just because they didn't update it doesn't mean that it got less high or something.

Woman 1: You want to be on the record saying it's really high.

(Jeda): I did four years of this therapy so I just think it's pretty high.

Woman 2: Okay. Great. So let's move onto performance gap.

Woman 4: In terms of gap in care I want to just thank the staff for making sense graphical as compared to the other where it was a run on of data. But here we can see that from 2008 to 2015 it went from 78.8% to 92.7%.

Then I'm going to throw in this side light when we look at the data for five years and then adding even AIs on top of people who are pre-menopausal and went to post-menopausal to remind us that this is a measure that is not telling us whether people continue to take it, how long they take it, what the issues are because we look at young women and so on.

So I think there is some real room to look at this measure in a more robust and quite honestly meaningful manner then did we recommend or was it administered when it does not indicate the length of time rather than it occurred within one year. So I personally will spy us on that one.

And if you look at disparities in terms of gaps for the most part the positive increase in terms of capturing race and ethnicity, age and broadness in terms of insurance. And so ultimately that was rated as moderate with no comment in terms of that aspect of it.

Woman 1: Okay. Thank you. Is there any discussion on performance gap increment that people have?

(Len): Just so everybody is aware I just have to say this, please note that they have insurance data and disparities data and that's because of the National Cancer Database. It would be great if other databases had some more information.

Woman 1: Thank you. Good comment.

Woman 3: I was just going to also comment the sites that are to add to each other's comment, the people that are reporting the data into the registries are more likely to be motivated to respond to getting these quality measures and so I think you look at a select group off reporting entities.

Woman 1: All right. Some of you have read about it on the podcast. Yes. Okay. Thank you.

Woman 2: Voting is now open for performance gap on measure 0220. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for performance gap on measure 0220. We have 3 votes for high, 14 votes for moderate, 1 vote for low and zero votes for insufficient. This measure passes on performance gap. Thank you.

Okay. So next is reliability.

((Crosstalk))

Woman 4: In terms of reliability we've ultimately rated it as moderate. It's a process to registry base it certainly has a long history of usage of the same measure score. Data source was utilized. No risk adjustment for, you know, current reasons. And nothing exciting from my perspective or someone else.

Woman 1: Any comments on reliability?

((Crosstalk))

Woman 1: Okay. Are we ready to vote for reliability? Okay.

Woman 2: Voting is now open for reliability on measure 0220. Options ate A for high, B for moderate, C for low and D for insufficient. Well, we need one more vote.

Voting is now closed on reliability for measure 0220. We have 2 votes for high, 15 votes for moderate, zero votes for low and zero votes for insufficient. This measure passes on reliability.

Woman 1: Thank you. So now risk gap validity.

Woman 3: And it should be noted within the context of the review that validity and reliability kind of married up together and just a as back drop I did fail to

indicate under reliability that over the years it went from a coefficient of 0.74 to 0.83 so along with that increased reporting the reliability considerably increased with does have some impact on validity obviously.

And so validity often was identified as moderate. Again it's a process, was not able to indicate high related to lack of it being a score level testing had not been conducted. In other words there was very little comment other than that related to validity.

(Len): Do the committee comments become part of the long-term records? I mean they are public - I mean these comments that we see in the committee do they become part of the record going forward. I guess so let me be more specific so you know why I'm asking.

There is a comment in here that says - one of the committee members made a comment, there is no exclusions to this measure. In fact there are exclusions to this measure illustrated directly or indirectly as we talked about before.

So for example men are not included, people who have treatment other than primary surgery are not included. So I guess, you know, I don't know if we have to make a comment on that or not. You know, it's there but I don't know if it's correct and that's why I'm bringing it up.

Woman 3: Thank you (Len). I missed that because there is 10 good different exclusions.

(Len): Anything you want to talk about here just so I'm not misstating what I'm seeing. It says there are no exclusions. The fourth column under 22 - 2B47 clips.

Woman 1: The orange boxes down here are not (unintelligible).

(Len): I don't want to get tied up on this. It's not a major issue. I just made an observation and that's why I'm just bringing it up for consideration - for your information.

Woman 1: Okay. So to answer your question, the whole measure worksheet that has that orange box that says comments that is shared with the documents related to this project.

The other thing on the Web site is that when someone searches for a measure on the (unintelligible) Web site it just measure so that wouldn't show the comments. But if someone looked up the actual measure form it would also show the committee comment.

(Len): Then I just think somewhere we need to say that there are exclusions and, you know, I don't know how you deal with it but it's one of those things that it's not consistent with what the measure says.

Woman 1: And the excluding will be included with the measure so someone would see that.

Man 1: (So (Len) should we as the committee say that we reviewed it and viewed that there were exclusions that were appropriate?

(Len): They are appropriate exclusions. We had that conversation so.

Woman 1: Okay. Are there other comments on validity? Anyone on the phone? Okay. Can we go ahead and vote on validity then?

Woman 2: Voting is now open for validity on measure 0220. Options are A for moderate, B for low and C for insufficient. We need to make a correction on this voting slide. It just has the option of high, moderate and low. Thank you everyone for your patience.

Voting is now open for validity on measure 0220. Options are A for high, B for moderate, C for low and D for insufficient. We are just waiting on one more vote.

Voting is now closed on validity for measure 0220. We have 5 votes for high, 12 votes for moderate, zero votes for low and zero votes for insufficient. This measure passes on validity.

Woman 1: Great. Thank you. Let's move on to feasibility.

Woman 3: Feasibility is surely a short check as I mentioned before we are capturing 80% of those newly diagnosed breast cancer across the country. It's identified as high in terms of feasibility related to Web based audit feedback, you can go to new and update.

You know, the breath of that is also involved in rapid quality reporting system and getting rapid feedback using hospital registry data that's utilized for a number of other reporting as well.

So there were no pre-evaluation comments and was identified as high feasibility.

Woman 4: So this it's a very feasible measure for the centers that are COC centers and to monetary contribution to the COC center. It perhaps is not feasible if we look at the broader population and that's something important to keep in mind.

Woman 3: Excellent point. But yes, any data that's a year is difficult data regardless. But to your point Dr. (unintelligible) level of commitment for breast cancer may be higher than.

Woman 2: Any other comments on this on feasibility? Anyone on the phone. Okay. Let's go ahead and vote on feasibility.

Voting is now open for feasibility on measure 0220. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for feasibility on measure 0220. We have 9 votes for high, 7 votes for moderate, zero votes for low and zero votes for insufficient. This measure passes on feasibility.

Woman 1: Thank you. And now use and usability.

Woman 4: In terms of usability and use this is (unintelligible) imported, it's currently used in accountability programs in terms of CP3R and (CQUIP) phone calls and emails. It repeats in here that the measure in terms of compliance has an increase from 78.8% to 92.7% and hopefully that's reflective of 20% of people or women who are not followed in terms of that.

Hospitals are sent alerts under the rapid quality response system so it's more timely. As a clinician I get notes from insurance companies, "We know that your ERPR positive person is not filling their prescription." And it's like, "Yeah, I know. Tell me something I don't know." But it was said.

So it should certainly have an impact in terms of - and so if we look at preliminary rating and usability high with no pre-evaluation comments.

Woman 1: Okay. Any other comments with that?

Woman 3: I did want to mention potential for under reporting just and I think that aligns with the virtue of, you know, a long line (unintelligible) potentially.

Woman 2: Voting is now open for use on measure 0220. Options are A for pass, B for no pass.

Voting is now closed for use on measure 0220. We have 16 votes for pass, zero votes for no pass. This measure passes on use.

Voting is now open for usability on measure 0220. Options are A for high, B for moderate, C for low and D insufficient.

Voting is now closed for usability on measure 0220. We have 10 votes for high, 6 votes for moderate, zero votes for low and zero votes for insufficient. This measure passes on usability.

Woman 1: Okay. Great. Thank you. There were other some related measures, there are competing measures. I don't know if there is - pardon.

((Crosstalk))

Woman 1: Yes. I was going to say I don't know if we need to discuss them unless anybody had any specific comments on them so we can talk about - if you do have any comments then we can discuss them otherwise we talk about the overall suitability for endorsement.

Woman 3: I just had a quick question related in competing. So for 0385 it talks about prescribed within 12 months as opposed to this measure talks about administered and recommended. I mean it would also be a prescribed as opposed to it ministered.

Woman 1: So (Jane), you're saying that it would be prescribed - you're - because there is a difference between the two measures. There always has been a difference.

Woman 3: Yes. Absolutely.

Woman 1: So - but one is solely prescribed and the other is administered and recommended but not received based on a (unintelligible) potential expenses. I'll be honest I haven't read 0387 so I'm not sure how prescribe is different from administered because I'm not administering them. Ort hormone I'm writing out a prescription and the following up on.

Woman 3: I'm giving the opportunity.

Woman 1: So not me but probably anything other than my own self-gratification and learning.

Woman 2: So were there any comments on over - any other comments on overall suitability?

Woman 5: For you that's familiar with the system - HS system, there is a delay of two to three years to get the aggregate data. I know they have RQRS for the individual provider. That's a very long time. I mean, a lot of things could happen in two or three years. Is there any way that that can feed back and they can then be made a little bit faster? I'm just asking…

Woman 3: Can I comment on that? So the data that they are giving us is delayed data because they wait for any appropriateness of follow up. But I think that COC should be congratulated on enforcing the RQRS system which gives us much faster turnarounds and more immediate and actionable data.

So I think that gap got addressed by that adding the RQRS rapid reporting system especially on some these kinds of measure where there is a timetable.

So I think there is a difference between their ability to analyze and report data for this kind of purpose and create a useable piece of information for a clinician at that site. And after January 2017 one had to participate in RQRS in order to remain accredited to COC so certainly the direction is pushing in the right direction.

Woman 1: (Len)?

(Len): I gave you somewhat a flip response I didn't mean to do I hope it wasn't (unintelligible). The RQRS comments I think ate absolutely valid. But one of the issues we face in cancer care - in fact in medical care in general is a substantial delay in the event to reporting.

So for example, mortality statistics (unintelligible) statistics are delayed by the people who measure those things for anywhere from two to three years because of validation issues and delayed reporting.

So your point is extremely well taken. There is an incredible concern about making sure we can do things more quickly and that is a focus of a lot of effort. So it's a very important question across the board and we reported mortality statistics for last year for which we had as 2017 but that's because of quality assurance issues. But yes RQRS is intended to try to…

Woman 5: So the reason I bring it up is this one has the RQRS which means its more timely acquisition of the data. And so the time to reporting the population aggregate should be shorter than, you know, other elements that doesn't have that.

(Len): And I don't know if that - and I don't know the answer to this but I would say there may be a difference between individual institutional reporting and return of information versus validated population basis. I'm not sure about that but there are - as you know there are data audits that are done on this it was commented on say about 10% of the patients.

So that may be a different - so yes, one is okay. Not perfect. The other one they weren't trying to make it as well and that may be the explanation. I'm not 100% sure about that.

Woman 1: All right. I think we ready to vote on overall suitability for endorsement.

Woman 2: Voting is now open for overall suitability for endorsement for measure 0220. Options are A for yes, B for no.

Voting is now closed for overall suitability for endorsement on measure 0220. We have 16 votes for yes, zero votes for no. This measure passes.

((Crosstalk))

Woman 1: Yes. So it - we can probably fit in one more measure however we are in risk of losing quorum. So we mostly need to know from the people online will all of you be able to stay on the meeting and we will maintain quorum just so you don't feel (unintelligible).

Woman 6: This is Dr. (Unintelligible). I can stay.

(Danielle): This is (Danielle Zonic). I can stay.

(Denise Morris): (Denise Morris). I can stay.

(Shawn Miller): (Shawn Miller) and I can stay.

Woman 1: Okay.

((Crosstalk))

Man 3: …as well.

Woman 1: Okay. So that means we are going to have a quorum so let's just move onto 223 other than…

Woman 3: So sorry. Can I just absolutely but we will still have one measure left to review, correct?

Woman 1: Right.

Woman 3: So we will probably get through one and then we will have to do the other one for the post coming call.

Woman 1: Yes. So just before folks leave, we do have another call scheduled for next week. We really think it's really important for folks to be there or to listen in so that we have quorum. We find the best conversations are when we have the most amount of people.

So I would just sort of beg you before you leave to make sure that that is on your calendar and please confirm. I think we have maybe a smaller number of people who get accepted. So when you are on - when you're heading home tonight and you're flipping through your phone just hit the accept button there.

Woman 4: So what's the date?

Woman 2: Wednesday.

Woman 3: Wednesday March 4…

Woman 2: It's March 4, 11:00 am Eastern Time so 1:00 pm Eastern Time.

Woman 1: Well, so we have fewer to review at that time let's go through 223. (Unintelligible) chemotherapy is recommended or administered within 4 months, 120 days post diagnosis for patients under age 80 with AJCC 3 lymph node positive colon cancer. And we have a question.

((Crosstalk))

Woman 1: I'm sorry. I'm sorry. So then I've wasted time reading that title. And so the next one is adjuvant radiation therapy in breast cancer so administered within one year. And the reviewer is (Susan Chen) but we have the developers on to give us a very brief introduction.

(Len): Sure. This is measure 0219 radiation within a year of diagnosis for women under the age of 70 undergoing breast conservation surgery. This particular

measure has been in our NCCB reporting systems, RQRS and (CPRs) since 2008.

The original baseline compliance was 76% in 2003 reaching 92% in 2015. This measure impacts approximately 60,000 cases annually from about 1400 hospitals across the United States. And this too is a maintenance measure with no changes in evidence since the last evaluation.

Man 1: So starting with evidence as you heard this is for radiation therapy within a year of diagnosis done adjuvantly so as you know these are landmark papers. The evidence hasn't changed due to the way the structure of the rubric works it ends up being recommended as moderate because there is not a new systematic review. I don't think any of us have any concerns about.

Woman 3: So is there any discussion about the evidence? We have the option I guess not to vote on evidence but you recommend that we vote on the evidence.

Woman 2: Voting is open for evidence on measure 0291. Options are A for moderate, B for low and D for insufficient.

Voting is now closed for evidence on measure 0219. We have 15 votes for moderate, zero votes for low and zero votes for insufficient. This measure passes on evidence.

Man 1: All right. Moving onto the next is performance gap. As the developer presented, they have made significant strides over the years however you can see from the disparities data that there are some disparities as they pertain to race and insurance status and with almost two acts different but they are closing.

The preliminary staff recommendation was moderate and I don't think there was much in the way. There are some people who are concerned that the gap is getting smaller and not as striking but I would suggest that (unintelligible) some groups.

Woman 1: Are there any further questions or comments (unintelligible) that there is disparity data that was provided? So why don't we proceed with the vote on this.

Woman 2: Voting is now open for performance gap on measure 0219. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for performance gap on measure 0219. We have 2 votes for high, 12 votes for moderate, 1 vote for low and zero votes for insufficient. This measure passes on performance gap.

Man 1: The next element is reliability and validity. Taking reliability first this was by the staff was moderate. They did do physical testing and they end up with coefficient of - now I can't find it again but I want to say it was like 0.6 roughly - oh, there it is. 0.58, 0.62, 0.71 for each of the three periods tested. 0.71 is for the two years combined that they tested.

So overall that's about to be moderate to very good and so moderate is the recommendation.

Do you want me to go over validity now too or do you want to do the…

Woman 1: No. We'll vote separately. We'll do for the (unintelligible). So is there any question regarding the reliability there? Okay. Why don't we move to the vote?

Woman 2: Voting is open for reliability on measure 0219. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for reliability on measure 0219. We have 3 votes for high, 12 votes for moderate, zero votes for low and zero votes for insufficient. This measure passes on reliability.

Man 1: Going onto validity. So validity is assessed by onsite audit basically that's done by random sampling across variable stem charting. There isn't risk adjustment being done here. The staff had preliminary marked as low and had marked this as not addressing stress to validity. The online pre-comments there was somebody who said, yes they have a concern but didn't elaborate.

Otherwise most people didn't have any issues. Can the staff elaborate a little bit on what they were concerned about what threat to validity weren't being assessed well?

And I think it's because all critical data elements were not assessed and I think it's because they don't check every single element. They take a sampling of elements and they check those and I suspect that that's what generates the staff's recommendation.

Woman 1: Yes. That was it.

Man 1: From my perspective having been through a number of COC audits they are generally pretty thorough but certainly not every element is checked. These elements I think for radiation are pretty clear cut and unlikely to incorrectly entered I would think but that is certainly subject to that. My personal

recommendation would be that it would probably be moderate but in terms of the staff recommendations low.

Woman 1: Any discussion from the group? Or just comment…

((Crosstalk))

Woman 1: Yes. Please. Go ahead.

(Glen): So because they didn't do score level validity testing that maximizes the rating for validity at moderate. Is that correct?

Woman 2: Yes. That's correct.

(Glen): Okay. Thank you.

Woman 3: I was just going to comment that it's basically the same process for the last measure so I'm surprised in the disparity in the staff rating but I actually have moderate as the (unintelligible) I think we did have a disconnect between (unintelligible) for the low, I didn't see the low. (Unintelligible). 0219. Okay.

Woman 1: So is everyone ready to proceed?

Woman 2: Voting is now open for validity on measure 0219. Options are A for moderate, B for low and C for insufficient.

Voting is now closed for validity on measure 0219. We have 14 votes for moderate, 1 vote for low and zero for insufficient. This measure passes on validity.,

Man 1: The next element is going to be feasibility. This is data that's constantly generated by anyone who has a cancer registry including anyone who is in the COC. And so it's highly feasible for anyone who has one, I'm just concerned there are some smaller hospitals that don't but the staff rating is high I would concur.

Woman 1: Any other commentary from the remaining members? Yes (Len).

(Len): Is this measure - so I'm looking for the information here about the NCDB. Is this measure limited to NCDB hospitals? You know, it talks about how the facilities available NCDB and all that. But not a lot of major large institutions don't belong to the COC. So I'm assuming this is not limited to NCD - to hospitals that have National Cancer Database but anybody can report this.

Woman 1: I direct that to the developer's instructions for the measure (unintelligible).

(Bryan Taylor): Yes. This is (Bryan) again. The NCDB so as a benefit of being part of the Commission on Cancer we can only report back to Commission on Cancer programs. The measure specifications are fairly well known. They can be applied to any registry data if we regardless of whether they are in a COC reporting hospital or not but would only report back to CVOC hospitals.

(Len): Yes, it answers it. I'm just not sure it's the answer I wanted to hear because it's used in other reporting systems we haven't gotten there yet. I mean, we just - we'll hold off any further (unintelligible) and just wait to see where we end up, you know,

Woman 1: Okay. Any other comments before we move to a vote?

Woman 2: Voting is now open for feasibility on measure 0219. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for feasibility on measure 0219. We have 9 votes for high, 6 votes for moderate, zero for low and zero for insufficient. This measure passes on feasibility.

Man 1: Next element is use. It is publicly reported. It's used a cross a number of accountability programs. And to answer your question (Len), it's used in the Pennsylvania program so I'm guessing that not all of them are COC hospitals so you can do anything possible but it may be difficult.

So the recommendation is for pass and I would agree with that.

Woman 1: Any other discussion? Or discussion on the phone? So let's move to a vote.

Woman 2: Voting is open for use on measure 0219. Options are A for pass, B for no pass.

Voting is now closed for use on measure 0219. We have 15 votes for pass. Zero votes for no pass. This measure passes on use.

Man 1: Next element is usability. And so here we can see that we are seeing improvement as the measure is having effect and I think it's important to note that the groups isn't kind of a disparity research that we are doing our folks are improving the most so that's heartening to see.

Did you know that critically some rural hospitals and things like that maybe having more trouble complying which makes sense so they have further travel time to radiation oncology centers often time.

However overall the staff their rating is high and I agree with that.

Woman 1: Any comments on the phone or in the room? So let's move to a vote.

Woman 2: Voting is now open for usability on measure 0219. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for usability on measure 0219. We have 12 votes for high, 3 votes for moderate, zero votes for low and zero for insufficient. This measure passes on usability.

Woman 1: So there is no related or competing measures so the final job we have is to vote on acceptability overall suitability. Any comments from the room or online? So I think we are ready to move to the (unintelligible) feasibility - suitability.

Woman 2: Voting is now open for overall suitability for measure 0219. Options ate A for yes, B for no.

Voting has now closed for overall suitability for endorsement for measure 0219. We have 15 votes for yes, zero votes for no. This measure passes.

Woman 1: Is everyone okay with looking at the next one and we should have (Steve) do it again since he - no. I'm just kidding. First we should complement the developers for the data that they presented to us. This is very helpful.

So the next measure is 0223 and I have already read it to you adjuvant therapy and that's (unintelligible) stage 3 node positive. The developers can - if you could please give us a brief introduction?

Man 4: Sure. This measure has been in our reporting system too since 2008. The original baseline compliance was 66% 2003 data and its now approaching 89% in 2016. This measure impacts about 20,000 cases annually from, 1300 hospitals across the U.S.

Woman 1: So (Jennifer) - Okay. Great. So (Jen Maylan) is the lead discussion. If you could open us up with the discussion of evidence.

(Jennifer Maylan): Yes. So I think this can go probably really quick so there is lots of great evidence for this measure. There have been multiple randomized trials to support the link between giving adjuvant chemotherapy and then proving survival.

On the preliminary rating too that was reported of moderate and I think that may be just the developers may have not provided kind of, you know, as comprehensive a summary as the evidence of success.

Or maybe because there isn't necessarily guess the evidence that this measure has improved the outcome. But overall I think there is multiple randomized trials, meta-analysis and a strong evidence base for this measure.

Woman 1: Any discussion among the group? So I'm going to move that we - again we don't have to renew this vote. We can accept the old evidence but we've been voting on all of the evidence so I recommend we proceed with the vote. So we're ready.

Woman 2: Voting is now open for evidence on measure 0223. Options are A for moderate, B for low and C for insufficient. We are waiting on one more vote.

Voting is now closed for evidence on measure 0223. We have 15 votes for moderate, zero votes for low and zero votes for insufficient. This measure passes on evidence.

Woman 1: (Jennifer), if you want to discuss disparities and performance again.

(Jennifer Maylan): Yes. So as the developers mentioned the main performance has improved since the measure was first introduced in 2008 from 83% up to just about 89% in 2015. But that's still a gap there in performance.

And then when you look across racial, ethnic groups there is evidence of disparity with (unintelligible) 82.5% of Hispanic receiving adjuvant therapy compared with a high of 90.2% for non-Hispanic whites.

I would say the one limitation of this data is that, you know, its 2020 and we are looking at 2015 data so, you know, its theoretically possible or maybe even likely that its improved further over the last five years. But at least based on the data we have there is still room for improvement and disparity.

Woman 1: Thank you. Any comments from our committee members? So I would move that we vote.

Woman 2: Voting is now open for performance gap on measure 0223. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for performance gap on measure 0223. We have 5 votes for high, 10 votes for moderate, zero votes for low and zero votes for insufficient. This measure passes on performance gap.

Woman 1: So we'll move to reliability discussion. (Jennifer).

(Jennifer Maylan): Yes. So in terms of reliability so as with the last two measures the data for this measure is used from data extracted by the Cancer Registry and so I'm trying to kind of find the - when it is here but - so there is a substantial number of patients including end of liability assessment.

I think the one challenge here if I can find it so the reliability measure is in a moderate range. It was 0.5 in the most recent period 2014 to 2015 when it was unadjusted. So the reliability was quite a bit lower in hospitals that have less than five cases a year potentially not surprisingly.

So I think, you know, the reliability give that it's extracted from data that's in the chart at the hospitals and they may not always have all the data with moderate and not as good other hospitals essentially.

Woman 1: Any discussion regarding that - these comments? I guess I would ask the developers if they have any speculations on why that reliability is less for those groups?

(Bryan Taylor): So this is (Bryan) in CDB. We know that case volume is what's driving these models. It's actually three different factors - case level is one of them, hospital level performance. But because we did a hierarchical model the variability and performance across hospitals is also factored in.

I can't specifically tell you why we're seeing lower reliability especially across the board depending on, you know, no matter what year you are looking at it's still lower. But we know that it's driven chiefly by volume and you can see that when we bundle the two diagnosis years together 2014 and '15 reliability both table 1 and table 2 bump up.

Woman 1: Okay.

(Jennifer Maylan): Yes. Although just kind of pushing back a little bit I mean you have over 10,000 cases each year. I mean that seems a little, you know, it doesn't seem to be like overall an issue.

(Bryan Taylor): It is and you're absolutely correct. Its 10,000 cases but its spread across - how many hospitals do we have? Approximately - I'm sorry, the number is in there. Its 1300 - I'm sorry, 1406 hospitals.

Woman 1: Any other discussion from the group? So are ready to vote?

Woman 2: Voting is open for reliability on measure 0223. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for reliability on measure 0223. We have 1 vote for high, 13 votes for moderate, 1 vote for low and zero for insufficient. This measure passes on reliability.

Woman 1: (Jennifer).

(Jennifer Maylan): So moving onto validity the developers did not really provide any data regarding validity. They provided a description of what I would actually call really more a measure of reliability kind of a re-description of how they re-extract 10% of annual caseloads for the registry. So that one could almost think of as more of like (unintelligible) some sense to reliability more than validity, I think.

So they really didn't provide any data on validity.

Woman 1: Any discussion from the group?

Woman 4: I mean I definitely expect the measure developers - I mean this is a measure that's been around for a long time and we cleared passed it before. I would assume that previously validity data was submitted but just wasn't submitted this time.

(Bryan Taylor): Actually I want to say that was true but we don't do any type of case re-re re-abstraction outside of what's documented in our standards. So let me just take one step back and say that if we're defining validity as being a comparison between what's in the patient chart and what actually gets abstracted to the extent that we do that either takes place during the survey once every three years.

Or it's assessed by standard 1.6 which is required to maintain accreditation and that's some percentage 10% of the annual analytic case load is reviewed for factors such as staging and treatment - first course treatment.

Woman 1: (Len)?

(Len): I guess I know the answer but first I would like to comment on the insufficient because then what are the implications if we have an in sufficient rating?

Woman 2: So I think this came out with another measure but essentially, we believe that that's just a call trying to assess data quality wasn't performed and additional validity on testing of the critical data elements also was not performed. So that will be the rationale for our rating of insufficient. And I'm in the process of looking through the algorithm to make sure that we determine what an option would be.

So we would not have - if the question is whether or not we would have a similar option if this were rated insufficient to sort of override that as we did with evidence in an earlier measure. That is not possible. So if the rate is insufficient this is a must cost criterion and the measure would not move forward.

(Len): When I'm reading this under validity testing it says for maintenance measures less emphasis if no new testing provided. Will this qualify into that?

Woman 4: So in the meantime I will make a comment. We should accept that the same validity testing for the two proceeding measures so - because I think that the idea of re-auditing audited charts seems like a reasonable idea. But that's just an observation.

Woman 2: To (Robert's) comment. The measure still has to meet the current criteria and we understand that you had mentioned it just in principle but it still has to.

Woman 1: But we can as a group vote that the data is or the data provided is sufficient.

Woman 4: Okay.

Woman 1: So just so everyone knows. So we want to be consistently audited too, right? Anyway - but any other comments? And (Jennifer) you raised the concern so any comments?

(Jennifer Maylan): So I mean I guess it's more just a process comment. I mean, personally I'm comfortable with validity of this measure. You know, I think that there are other measures out there like it I mean I think it has face validity, you know, I think the I guess the (unintelligible) sometimes looks registry data that its

chemotherapy not given in the hospital and there is not a consistent way of accessing the external records it could be enough.

But, you know, I think, you know, the measure has been in use so I don't really have concerns it's more just a - it's a little odd that when no information on validity is provided to rate it anything other than insufficient.

Woman 1: Okay. Is the group ready to vote? Okay.

Woman 2: Voting is now open for validity on measure 0223. Options are B for moderate, B for low and C for insufficient. We are just waiting on one more vote.

Voting is now closed for validity on measure 0223.

Woman 1: So we'll move to discuss…

((Crosstalk))

Woman 1: So it needs to be 6 greater than or equal to?

Woman 2: Voting is now closed for validity for measure 0223. We have 9 votes for moderate, 4 votes for low and 2 votes for insufficient. This measure is consensus not reached on validity.

Woman 1: So we should finish the discussion and we will have to come back to this and vote after public comment. So (Jennifer), can we move to feasibility?

(Jennifer Maylan): Yes. So I mean I think in terms of feasibility, you know, the - again this measure has been in use for many years. The elements are routinely collected during care delivery, they are available on the chart EHR, you know, can be

obtained not necessarily the save data but the treatment data from claims data. So I think, you know, the perceivable measure and the registry captures all these data elements.

Woman 1: Any other comments…

((Crosstalk))

(Jennifer Maylan): The preliminary rating was moderate and I wanted to know (unintelligible) rate it as moderate or high personally.

Woman 1: Okay. Any other comments. So let's proceed to vote.

Woman 2: Voting is now open for feasibility on measure 0223. Options are A for high, B for moderate, C for low and D for insufficient. We are waiting on just one more vote.

Voting is now closed for feasibility on measure 0223. We have 3 votes for high, 12 votes for moderate, zero votes for low and zero votes for insufficient. This measure passes on feasibility.

Woman 1: So onto use and usability.

(Jennifer Maylan): So I think as it was previously described the measure is in use. It has, you know, I think the use of the measure has led to increase in rate of adjuvant chemotherapy from 2008 to 2015 but still, you know, opportunity for improvement and, you know, through the rapid - the COC Rapid Response System, you know, was used in QI efforts as well. I think the evidence you provided are the feasibility.

Woman 1: So any other discussion from the group? So let's do bundle the two votes.

Woman 2: Voting is open for use on measure 0223. Options are A for pass, B for no pass.

Voting is now closed for use on measure 0223. We have 15 votes for pass, zero votes for no pass. This measure passes on use.

Voting is now open for usability on measure 0223. Options are A for high, B for moderate, C for low and D for insufficient.

Voting is now closed for usability on measure 0223. We have 2 votes for high, 13 votes for moderate, zero votes for low and zero for insufficient. This measure passes on usability.

Woman 1: So we don't need to vote on suitability at this time. We'll bring that back to the group later.

And I think that closes our discussion. I know the staff had some other plans for the last half an hour. But any other comments from the staff…

((Crosstalk))

Woman 1: Yes. Public comment. For the people that stayed on the line the entire time we appreciate you very much. We are open for public comment.

All right. We have none at this point. So…

((Crosstalk))

Woman 2: Thank you. I'm just going to review a couple of next steps before we adjourn. So this first line here with the first meeting wide meeting we will not be having that next Wednesday because we just had that discussion…

((Crosstalk))

Woman 2: …the measure today. Yes, I will take it off everyone's calendar. We will have - we will open the public commenting period where we will post the draft report with a summary of the measure discussion and voting outcomes.

Once the period closes on April 19 following that we will have a post (unintelligible) wide meeting where we will discuss any comments received during the public commenting period and we will discuss and revote the measures that consensus was not reached today.

Is there anything else?

Woman 1: I just want to say on behalf of the team here thank you to the committee, those on the phone who stayed through an entire day on the phone we greatly appreciate your volunteering, your participation in the meeting and certainly to everyone who flew in thank you so much.

Woman 2: And I would just like to say thank you to our staff who did help facilitate the meeting.

((Crosstalk))

END