NATIONAL QUALITY FORUM

Moderator: Angela Franklin
March 2, 2012
2:00 pm CT

Operator: Welcome to the conference. Please note today's call is being recorded. Please stand by.

Angela Franklin: Hello everyone. This is Angela Franklin and this is the Cancer Endorsement, Maintenance Oncology Workgroup call. And we are going to be discussing several oncology measures.

The process as we go through will be that there's a lead discussant for each measure. We will have the lead discussant tell us their highlights about the measure as we walk through the criteria. And also tell us their recommendations and then throw it open for the group for discussion.

There are measure developers on the line for each measure and if the steering committee has questions we may un-mute the lines of the measure developers to respond. So with that, I also wanted to make another note about the agenda.

Our first measure is listed as 381. And due to the expert that will be responding to that measure, we'd like to move that to the end of this discussion just before Barrett's esophagus, 1854. To give that developer time to attend and respond to any questions.

I hope I didn't confuse everyone. We'd like to go ahead and start if we could with measure number...

Male: Very briefly, it's ((inaudible)). I just joined. I'll be listening. I'm actually driving so my phone will be on mute. If you need to get me, just yell out.

Angela Franklin: Okay, thanks ((inaudible)). So we're going to start with 383 instead, taking it out of order, and that's oncology, plan of care for pain, medical oncology and radiation oncology. And that's a paired measure with 384. I have Dr. (Mallun) as the lead discussant for that measure. Dr. (Mallun), are you there?

Dr. (Jennifer Mallun): Sorry, I had my phone on mute.

Angela Franklin: Okay.

Dr. (Jennifer Mallun): I didn't know I was the lead the discussant on this, but I'm happy to go ahead. So I think there's a number of issues with this measure in terms of the evidence as well as reliability and validity. I think first of all the denominator for this measure includes essentially any pain of any severity and it's from, you know, a single time assessment when the patient is in the office.

So if someone, you know, has a mild headache and they respond 1 on a scale of 0 to 10, they would be eligible for this measure. And so in terms of the evidence about pain being an issue in people with cancer, which it surely, you know, it is. And it's a really huge problem that needs a lot of attention.

The denominator to this measure is so broad as to, I think, dilute any potential impact that this measure would have for quality improvement or public reporting. Because you have, you know, far more people in the denominator than you're really trying to target improvement on.

The second thing is in terms of the evidence matching the numerator which as this one is constructed is a G code. And there was no evidence provided as to how - whether the G code actually is valid in terms of representing what actual management happens of the patient.

I think the reliability data was provided for basically - and validity today was provided for a small sample of patients from several different hospitals and practices. But using the G code and any pain recorded and, you know, the reliability of that was very high.

But again, given the validity measure, I don't know how useful that really is. So overall, I would see the evidence for this particular measure as constructed here as low or insufficient.

Angela Franklin: All right, and - okay, low or insufficient. So that would bring us - did you want to have some discussion about the feasibility and usability?

Dr. (Jennifer Mallun): So I think the - well the feasibility using - for the current measure I think is it's feasible. It's been in use so I think from both of those standpoints as constructed it's feasible. And its usability in terms of the fact that it has been used in public reporting, you know, from that standpoint it is usable.

I think whether or not, you know, no data were provided and there's none that I'm aware of, of how the reporting, public reporting, using this information has driven any improvement or change in any practices or hospitals or management of cancer patients. I think the (cope-E) measure which as been used for quality improvement limits the denominator to patients with sever pain, which I think it was suggested was a minor difference. But I think it's actually a big difference and makes it much more usable for quality improvement.

So from my standpoint, I don't think the measure meets criteria for endorsement.

Angela Franklin: Okay. And I'm sorry; did I miss reliability and validity, the scientific acceptability?

Dr. (Jennifer Mallun): So I think reliability, as I mentioned, for the measure as constructed, it appears that having, you know, using a G code is highly reliable and reproducible. However, there's no data on whether that G code actually reflects what happened to the patient and what plan was actually instituted. So, you know, in terms of validity I think there's very little validity on the measure as constructed.

Angela Franklin: Okay, thank you. Are there comments from the rest of the workgroup? Starting with the impact of - or the evidence-based.

Female: So I guess that, you know, a lot of these measures we've looked at today have this problem where they are summarizing evidence from practice guidelines that are more general than the measure themselves. And they're not providing data on the quantity, quality or consistency of the data from the guidelines. And so I think this is a classic example of we don't know whether there is evidence that links this measure with important outcomes.

And similarly, you know, we don't have the same list of impact and performance guess as well that data that we're provided doesn't support this. And I have to agree also which is I don't understand the reliability and validity message that are presented for this in several other measures in terms of whether the reliability is just a reflection of how well people are able to extract the codes that are used to capture this.

Or whether it is an evaluation of the reliability of different people being able to transfer the plan of care in this case into the code that's being assessed and I think that's a pretty important piece.

Angela Franklin: Would we like to un-mute the lines of the measure developer to address those questions?

Operator: And which companies are we looking for?

Angela Franklin: AMA PCPI.

Operator: Okay, one moment. And Samantha, your line is open.

Samantha Tierney: Thank you. This is Sam Tierney with the American Medical Association. I also would ask if you could un-mute the line - I think we have a physician participating for us from the ASCO organization and his name is Dr. Hassett.

Operator: Okay. And that was ASCO?

Samantha Tierney: Yes. But the physician participating is Dr. Hassett

Female: Okay, we have a number of people on the line to help support the measures so I'm just not sure how they've dialed in. They haven't dialed in as representing the AMA.

Operator: We have Michael Hassett with ASCO. We also have Kristen McNiff who is - and both lines are open.

Samantha Tierney: okay.

Operator: We also have (Marjorie Rowlands).

Samantha Tierney: Okay, great. So I appreciate you opening all those lines. I'll just speak to a few of the comments, but then I'll also defer to my colleagues here and anyone on the line who would like to comment as well.

So with regards to the broad denominator, I just wanted to clarify, and I'm sure this was clear to the review although it wasn't specifically discussed. But it is for patients with any diagnosis of cancer, which is quite broad. But it is limited in some way in that it's only focused on patients who are receiving treatments or receiving either chemotherapy or radiation therapy.

And the measure does include a documented plan of care to address pain for patients with any sort of report of pain. So that is an appropriate comment that was made. However, I just want to clean up that the plan of care is to address pain is quite broad and it includes, you know, obviously the use of pharmacologic therapy and nonpharmacologic therapy. But also includes something as simple as reassessment of pain at an appropriate time interval.

And I think part of the reason the measure was constructed so broadly is in - is consistent with the NCCN guidelines that refer to management of pain according to the various intensity levels beginning with mild pain, which as a rating of 1 to 3, moderate pain from 4 to 6 and severe pain from 7 to 10. So obviously the - whatever the plan of care may be would be - should be appropriate to the level of pain that is assessed.

But I just wanted to kind of point that out that it's trying to be consistent with the guidelines and also the American Pain Society has similar guidelines that recommend for anyone with pain that something be done even if it's just a plan for reassessing pain at a later point.

I also just wanted to add, I'm sorry, I just learned that one of our other physicians is on the line, I believe. Dr. (Hammond). And if he's not on the line I think he will be in the next few minutes. But

since I know we've had difficulty accessing open lines I just wanted to mention that because - in case he would like to add anything.

That's just my general comments on the evidence. But I don't know if, again, my colleagues or anyone on the phone would like to add anything else.

Kristen McNiff: This is Kristen McNiff from Nashville. I'm not sure I have much to add as what's pointed out by the reviewer. This has been implemented differently and the (cope-E) ((inaudible)) was a more narrow denominator of how patients with moderate to severe pain - the decision was made for this particular set as Sam pointed out to include that second step for all patients, which as she noted can be as simple as a reassessment. So it's just a different specification of the measure.

Angela Franklin: Okay, thank you. Are there other questions from the workgroup for the developers?

Samantha Tierney: Actually, this Samantha from AMA PCPI. So in response to your questions about ((inaudible)) reliability testing, basically we do this in order to make sure that it's - that the information that is found in the medical record can also - it's feasible for people to find the information in the medical record. And that is - it is actually correct.

You'll notice that in the testing results the overall reliability shows 100% agreement which means that it is possible for both abstractors or people who are looking at these measures to agree that they have found the same information in this patient medical record.

In regards to your question...

Female: It is actual plan of care in the medical records?

Samantha Tierney: I'm sorry; could you ask that question again please?

Female: The actual plan of care that we're talking about. They're going to the medical record looking for a plan of care, not just documenting in a code that a plan of care was.

Samantha Tierney: That's exactly it. They're looking at the plan of care and two abstractors are going and they're looking at the patient charts and they're making sure that Abstractor 1 with Abstractor 2.

Female: Okay great, thank you.

Samantha Tierney: So your questions on the validity testing, can you be a little bit more specific? I think I may have missed some components and I want to make sure that I'm actually answering your question. Or maybe there was no question?

Female: I guess my questions had been on sort of evidence supporting that documentation of a plan of care is associated with important outcomes.

Female: I'm sorry; my question was around the use of a G code. So the G code basically is valid I think only if there's been evidence that shows that checking off of a G code is actually highly - there's high agreement with finding a documented plan of care in the medical record. And so I didn't see any evidence that that had in fact been validated.

Samantha Tierney: So the second form of reliability that we used is parallel forms reliability. And that's what you're asking about. So we're looking at this code and we're making sure that it's matching the patient chart. So that is correct.

We do look at that and we make sure that it does match - that does in fact match.

Female: And so where is the data on that?

Samantha Tierney: Section 2A - wait. Maybe we could look for that and just let you know when we have exactly which section that was in. We're double-checking the forms right now.

Female: I mean, I think what I saw reported was 100% in the rate and reliability. I didn't see any quantitative information about validity in terms of the sensitivity or specificity of the G code in terms of what was reported in the record.

Samantha Tierney: So again, we're looking at the forms again and the information for the testing project and we'll get back to you specifically on that.

Female: The other question of validity, I mean, I think you present great information on safe validity, right. And I think some of us here wouldn't argue with safe validity. But in terms of representing it as a measure of quality, I think we need - we've been asked to determine whether it's valid in terms of a representation of quality.

So that it should be linked to some intermediate or help out. And I don't see that data in there, which puts it sort of at a low level of evidence or validity.

Angela Franklin: So it sounds like we might need to have you come back to discuss those two issues. Are there additional conversations or discussions we want to have about 383?

Male: The one question I had is in terms of when you're defining the quantitation of pain and it seems that they allude to there being standardized scales for doing this. But it was unclear to me how there's really any harmonization of how they really quantitate the pain. Am I misunderstanding or is that something that was addressed?

Angela Franklin: How pain is quantified?

Male: Yes.

Angela Franklin: Is that something AMA PCPI or a representative could respond to?

Samantha Tierney: So this is Sam. I think that question speaks a little bit more to the other measure that this is paired with, the pain intensity quantified measure. And that measure specifically says if pain intensity should be quantified using a standardized instrument such a 0 to 10 numerical rating scale, the categorical scale or the pictorial scale.

And for the second measure which is the companion measure, this would apply to anyone for whom pain is present. So above a zero.

Angela Franklin: Any other questions on 383? And Sam you can let us know when you're ready to respond to the other issue on 383.

In the meantime, hearing no other questions on this one, we'll circle back to get a response on some other issues there. But let's move on to 384 which is paired and was - already referenced it and that's the pain intensity quantified. And for that one I have Dr. Pfister.

Dr. David Pfister: Yes, I think that the - a lot of the prior discussions is actually quite relevant to this and I guess that the - so I guess in the - looking to be efficient I think some of the issues that (Jennifer) brought up I don't think are worth revisiting. Because I think some of the same principles apply.

But I - but circling back to the question I just asked, that in terms of the - that they just didn't do it but rather that it was just kind of quantified. It was unclear to me about the issue of harmonization among the way that this would be quantitated.

I think that the - again, you need to have a starting point somewhere in this business so I fully am sympathetic that there needs to be a start. But I think you have - that with the focus of this measure being sort of different from the prior one having to do with the intensity that that seems to be sort of a fundamental issue that further kind of leverages into some of the comments that (Jennifer) made.

You know, I think it has, you know, pain and trying to do a better job with pain control in the cancer population, certainly has, you know, face validity to it. And I do think that as far as reporting data publicly that it would have something that would be of great interest.

Although, you know, I guess I would be less optimistic about its ability as is as a quality improvement measure. And, you know, I think those were my main thoughts.

Angela Franklin: And that, again, is that based on the - you're looking at the evidence on the reliability and validity of the measure?

Dr. David Pfister: Yes. Because I think some of the same issues with I think the reliability and validity. You know, I went through 383 and 384 since they're kind of paired. A lot of the checks look virtually identical between both submissions.

And so I think that - and these two measures are quite related and I'm not sure that the data provided for 384 is any more robust than the data for 383. You know, it's heavily guideline - kind of drive the issues with the evidence in terms of the grading and assessment is sort of in the one set of guidelines it's kind of the consensus based grading system.

I think in the pain guideline I'm not sure there was a formal range system. So there's kind of very heavy consensus that this is sort of an important thing to do, which I think permeates, you know, much of this measure is the same way you did for 383.

Angela Franklin: Other discussion from the committee?

(Brian Lloyd): This is (Brian Lloyd). And I just - as I'm listening to the dialogue developed, I'm wondering if the measure developers or the author of these section that span both of these measures, if those authors are in a position to further characterize this kind of broad body of evidence with a finer point to answer some of the questions that are being raised.

I mean, it feels like - I noticed the same thing. We've got, you know, virtually identical techs. We've got some general references, but in terms of an actual evidence assessment I'm finding myself agree with the presenters. In that I still it at best is unknown as to whether or not they really meet the criteria that we have in front of us.

Female: You know, as I think we're held to a pretty high standard in terms of what counts as evidence so it should be specific to this measure for a maintenance measure. And so when we're not presented with the evidence here it makes it very hard to evaluate.

Angela Franklin: Could someone representing the measure speak to that?

Samantha Tierney: Hi, this is Sam Tierney again. I apologize if some of you were on the hematology call because this will be repetitive. But I think it's helpful to at least share with you our overall methodology for developing performance measures.

We have always based our measures on clinical practice guidelines. And so in answering these questions we refer to those clinical practice guideline statements that are specific to the

measures. I can appreciate that - and as NQF added these new criteria related to the quantity, quality and consistency of the evidence, different guideline developers assess that in different ways and share that information in different ways.

And unfortunately the NCCN guidelines have quite a bit of text to add to their algorithm which are the kind of primary source of their recommendations. But generally they don't provide information related to the quantity, quality or consistency of evidence apart from their overall rating scale.

And I do know that just in previous direction we received from NQF that I know that NQF has said that they don't expect measure developers to conduct primary evidence through (you), but rather to use existing systematic reviews and grading of evidence that has been done by others.

So we are limited by what's available. I will say that I know the Cochrane Collaboration has done a number of reviews on the treatment of cancer pain, but that is specific to the different treatment modalities. And these measures are quite broad. You know, the plan of care for pain, for example, would include any number of those elements, but it's just a very broad plan of care for pain, which I think is supported by the guideline recommendation.

But something like a systematic review of a plan of care for pain is not really available and it would be specific to the actual treatment modality. So I just wanted to offer that and put that we might offer some information to put the review and information submission forms into different context. And share the struggles and challenges that we face in trying to complete those sections of the forms.

(Brian Lloyd): I'm wondering, is it -- and this is as question that I would ask with you, being new to this process -- I think one of the issues that I struggled with in reviewing both of these measures is that the document actually points out to us that the categorization arrived at by the NCCN was a

2A. Although there was uniform consensus, but they can see that this was based on lower level evidence.

And that seemed to, you know, kind of define what our level of evidence rating was without additional information. So I'm just wondering, you know, if we were asked to defend our position if there might be within this body of evidence, knowing that it gets updated periodically with new literature, I don't know if there's more recent literature that hasn't be cited in the guideline. Or in an updated iteration of the NCCN guideline that we're not seeing here.

So that's a question. And to distill that down, is there more current evidence that would take that through a place where we would no longer be - have - oblige to adhere to the category 2A recognition that NCCN has assigned to this?

Samantha Tierney: This is Sam Tierney. Sorry, but I'm not sure if that was a question for me or for ((inaudible)) staff. But I just thought I'd also mention, and I'm sure Angela and colleagues are quite familiar with this. But I know the evidence report that NQF report puts out does have some sort of explanation about what to do in instances where there isn't strong evidence of certainty of net benefit.

And that those decisions could be based on judgment, that the benefits would clearly outweigh the harm. So I know there are some kind of unique scenarios when the evidence is lacking because - and I think the report specifically says that recognizing that much of healthcare has not been subjected to research studies and therefore there wouldn't be a strong evidence base to support the measures.

Angela Franklin: That is correct, Sam, as you stated it for NQF. Is that a question, Dave, you were asking? Or were you looking for guidance in terms of how to grade the lower level evidence?

(Brian Lloyd): Yes, this is (Brian). I was the one who asked the question. I'm feeling obliged to call this grade as lower level evidence because that's really all I have to go on. It's already been stated before. I feel like I'm basically looking at a generalization of very broad body of evidence and it looks to me that the (NCCN) concurred is that this was not really an evidence based decision as much as it was a consensus based decision.

So yes, I think the guidance that was just given to me in response to that was that if there is more good then harm, or if there's less harm - I'm not sure - I'm probably ruining that. That we might be given some different guidance. But I'm not sure what that allows us to consider.

Angela Franklin: Okay. So other than - yes. So other than restating that if you feel like on the face of the measure this is an area where there isn't any additional, stronger evidence to support the measure. And it's still the benefit of the measure outweighs the harm of not having that evidence available then you could still vote in support of the measure.

Female: So when we get - so for example, I mean, it seems that based on the criteria when it's on - it's based on a clinical practice guideline that does not provide information on quantity, quality and consistency and we're sort of forced to say that there's insufficient evidence to evaluate that. And so it seems like we then have sort of stopping guidelines in our rubrics for evaluating these.

That you're saying we could move on to that and say that in the big picture despite having this evidence we could make a move forward.

Angela Franklin: Right. And then we also rely on the committee and that they're aware of this more evidence out there than what's included in the forum or that's in the guideline or if there's newer evidence that would strengthen the measure that they're aware of.

We also rely on the committee to let us know that information as well.

Female: Okay.

Angela Franklin: We'll also be able to discuss this at the in-person meeting and ((inaudible)).

Female: Great.

Angela Franklin: So any other questions about 384 or discussion? Looking at both of them, there were some questions that we had that AMA PCPI was going to address on 383. Are they ready to discuss those since this is a pair?

Samantha Tierney: You know, Sam Tierney, Angela. We're still in the process of gathering that information. Honestly, we're missing a staff person who might be able to enlighten the discussion. So I don't know if we'll be able to provide that on the phone today. Is that something we could try to get back to you in writing after the fact?

Angela Franklin: Yes, that would work.

Female: Yes, you can send it to us and we'll disseminate it to the committee members.

Samantha Tierney: Okay. Thank you.

Angela Franklin: So on these two measures, 384 and 383, do we have an initial recommendation from the committee on 383 and 384? No?

Male: I'm going to propose - it sounds like that we still have some contingencies. And I don't know if we can defer that judgment until the in-person meeting or they supply the additional information?

Angela Franklin: Yes, that's an option. You can recommend, you know, with the contingency that you'd like to see something additional or an additional response from the developer.

Male: Yes, I certainly was more sympathetic to both measures until I heard this discussion. So I think that there are things which I would have before said were adequate that now I would say is insufficient data to address at the present time. And so I think I would endorse that, you know, proceeding in that manner.

Angela Franklin: Okay. Any other comments about the recommendations on 383 and 384? Okay. All right, moving on we have measure 386, cancer stage documented. Also an AMA PCPI measure.

We had asked Dr. Pfister to talk about this measure or start us off with a discussion of this measure, if he's able to do that.

Dr. David Pfister: I found out that was going to be my role about 15 minutes ago, but I'll make the best of it under the circumstances.

You know, certainly there's no argument that knowing what a stage is, is a fundamental aspect for planning treatment. I think also there is ample data provided that stage is associated with prognosis. I think that's clear. I think - and so I think that regard those are both, you know, positives in that regard. I think in terms of the actual data provided to show that by document - so while knowing the stage allows you to predict prognosis.

Whether the connection between that we actually documented that in the notes and it's translated into improved outcome, I don't feel that there's data presented to sort of document that aspect of what the measure's intended to do in terms of quality improvement.

Also, I was a little confused because like I know I think someone made the comment about focusing on breast and colorectal is such a fundamental thing you'd expect we could expand this other thing. I was okay with the focus on these two big public health issues.

But one of the things that was a little confusing to me was that for breast it seems to be based on clinical stage. And then if pathologic information is available then you would further sort of modify the stage information to incorporate pathologic data.

But for colorectal seems to be pretty much based on sort of pathologic staging. And I guess for the breast component in terms of defining this stage that I would think that, you know, that this kind of - well, if it's available that I guess I needed more reassurance that there wouldn't be confusion in the transition in the flow of information from clinical to pathologic.

And then within the realm of the pathologic staging, I was a little concerned that, you know, well for people who have non-metastatic cancer I guess with the implication that that would make your Stage 4 and that would change the ball - kind of the playing field.

But there are people that, let's say, might have non-metastatic cancer, but that would not be surgical candidates that we would never - they won't be able to go to the operating room. So you would have incomplete pathologic staging. But when you look at sort of their description that denominator exclusions that look like there were not enough.

And so I thought that that scenario would be something that would not be addressed. I thought that, you know, as I said I think that as far as public reporting goes and public interest, I'm not sure how high that would be. I think that the - you know, I think that the data for the harms of including that in the note I think are basically the data to show that there's no harm is kind of insufficient.

Because let's say that you have inclusion of a stage which I was actually a little surprised to see that it seemed that the data provided for reliability or, you know, inner observer rating seemed to be as high as it is. Because my own personal experience is actually the ((inaudible)) staging can vary quite a bit among observers.

In that something placed in the chart that may have a certain kind of ((inaudible)) status that may be perpetuated even though it happens to be the wrong stage. So I think that - again, I think that the strength is that I can clear that it's important for treatment planning, I think it's a low bar for that. I can clearly - stage is related to prognosis.

But some of the other ((inaudible)) criteria is that are sort of listed are the measure had more tough ((inaudible)) with.

Angela Franklin: Thank you. Are there other comments from the other workgroup members?

(Brian Lloyd): Yes, this is (Brian Lloyd). I echo the presenter's comments on perpetuating this staging information. I believe that there may be an issue about some of those guidelines or stagings that are in controversy or in a refinement period or perhaps still in development.

My other comment might be as I was reviewing the measure and I noticed the data source being administrative claims, it wasn't clear to me how anyone could ascertain much of anything about staging from administrative claims. Because I have yet to see an ICD-9 code that informs stage. So that was my concern.

Angela Franklin: So just to...

Male: Yeah, I...

Angela Franklin: Go ahead.

Male: Yeah, I mean, (Steve) is maybe muting his phone or driving or something. But, you know, he brought up that same concern about, you know, at least in terms of the electronic capture of this information and the limitations, you know, involved there.

Female: So the clarification I guess I wanted to point out is it seems like, you know, the ultimate question is whether this measure captures or represents quality. And it seems like - so what we're - what's happening in this case is it doesn't matter what stage is documented as long as some stages has been documented, this then represents a measure of quality.

And I don't think we've been presented with evidence that suggests that this code is accurately associated with quality or that there's evidence that it represents quality. And maybe part of that is, you know, it shows also that there is very little variability in performance on this. So the average was 96.9 or something like that with very low variability.

And that's probably because the bar is pretty low, it's just that the stage is documented. Not that the correct stage is documented.

Angela Franklin: Do we have a response from the ((inaudible)) on that?

Samantha Tierney: This is Sam Tierney. I'll just make a few comments. But I also know that our clinical expert, Dr. (Hammond), has joined us. So then maybe after I make our comments you can open up the line for him in case he has anything he'd like to add.

I just wanted to clarify I think there may be some confusion between the information that we've provided and the guidelines, which is anything that's relevant to documented phase and the actual requirement of the measure, which is something you could find in the numerator statement.

So the measure requires that patients would have a baseline AJCC cancer stage or documentation that the cancer is metastatic in the medical record. I know one of the physicians -- I'm sorry, I don't recall your name -- mentioned a concern about no exceptions. But I think that this - the case you described was for a patient whose cancer was metastatic. So that would count as meeting this measure.

So I think that's why there has been no exceptions noted.

Male: No, but I think for pathologic stage that you need sort of pathologic material to stage them in that manner. Like a clinical stage you could do without the pathologic material. But it seems, at least with the colorectal, they're quite explicit about it being a pathologic stage.

So I gave you the benefit of the doubt that for metastatic disease you just say metastatic disease. But there are reasons people don't get complete pathologic staging information beyond having just a metastasis. Like let's say you have a medical comorbidity that you don't - that they feel they can't really, you know, do surgery on you, for example.

Samantha Tierney: Okay, well, I think - thank you for that additional feedback. Again, I think, and you know, we could look back and see NCCN guidelines. I think you might be focusing on this section where we're quoting verbatim the guidelines. And I'm just - I'm not sure if that - if looking at the numerator of the measure might alleviate your concerns at all, as in 2A 1.1.

Just to clarify, that's the requirement of the measure. And the guidelines are just things that are supposed to help substantiate the reason for the measure. If I could also just comment on - someone commented on the performance gap data.

And so we have data from a number of programs on this measure. We have data from the (cope-E) measure, which isn't an adaptation of this measure. And while the average performance rate is, you know, fairly high at 83%, the range is from 35 to 100% and we have data from ASTRO's parrot program, which has an average performance at 87% and a range of 10 to 100%.

There was also a study that was done fairly recently looking at documentation of colorectal cancer care provided by academic and community sites in the United States, which might be even considered to be more representative than those other programs, which are voluntary and for folks that are focused on quality improvement.

And that actually saw that only 38% of the eligible medical records reported TNM stage. So I think there is quite a bit of variability provided from the gap information. But I'll stop there and see if Dr. (Hammond) or anyone else on the phone would like to add any other comments from our measure developer group.

Operator: And Dr. (Hammond), your line is open.

Dr. (Jim Hammond): Thanks. This is Dr. (Jim Hammond). I think that the intent, you know, to be honest the document is quite lengthy. So I don't want to misspeak here. But I think the intent is to report, you know, the best available stage as, you know, would be the case in, for say, a tumor registry.

So sometimes that's going to be the clinical stage, sometimes that's going to be the pathologic stage. But I think all of the practicing oncologists who are on the phone would probably agree that the lack of reporting of the patient stage is a sign of poor quality.

Now, you know, you can argue whether that's poor quality of documentation or, you know, that's poor quality, you know, lead support quality treatment. But, you know, I - in my opinion that represents a poor quality. And, you know, I think that there is a variability out there.

I think that, you know, you - I'm fairly familiar with the (cope-E) program. You know, there is self selection of, you know, practices that participate in that program. So to look at that program and say, oh, they're at 96% and that's representative of what's going on in routine clinical practice.

You know, I don't really think that that's accurate. And so I think that some of the data that Sam just shared regarding the variability is probably, you know, shouldn't be overlooked. Thanks.

Angela Franklin: Thank you.

Dr. Michael Hassett: This is Mike Hassett, can you hear me?

Angela Franklin: Yes, we can hear you.

Dr. Michael Hassett: Okay, I didn't know if I was on or not. I would also support Dr. (Hammond)'s comments. Universally, you know, I think there is always going to be change in the fundamental aspects of staging. But having the best available data is what I think this measure is trying to target.

And in some cases that may be clinical, in some cases that may be pathologic. But I don't think the measure is - I think it's relatively agnostic on the finer points. And I think increasingly we're finding surrounding oncology perspectives that stage is as fundamental to decision-making as the ((inaudible)) cancer diagnosis.

I think it would be hard to hold as a standard for this measure a link between stage documentation and improved outcomes, just given the inherent challenges of cancer medicine de novo. And I guess the comment that I would make from a quality improvement perspective as we move more from a paper to an electronic based chart, I think the importance of stage

documentation is only going to increase to the extent that it will contribute to clinical physician support in a population based analytics.

Angela Franklin: Okay. Any more questions from the workgroup regarding this measure?

(Brian Lloyd): Yes. I'm hearing - again, (Brain Lloyd), new to this process, but clearly hearing what the commenters are saying. What I'm hearing is that there is a two-piece argument here. One that quality of documentation is desirable and from where I sit that is - that's non-debatable. I would agree.

However, when I'm looking at, you know, what our charge is in terms of trying to evaluate within the quality of the evidence that's available and somehow link that to quality of care, I think that held to that standard I think we're hearing the committee members -- and I won't speak for every committee member -- but myself, I'm hearing you say - hearing us say that's a higher standard.

I don't know how you get to that higher standard until you start to catalogue the information. So in terms of desirability I would have to say yes. but in terms of meeting the criteria that are put in front of us, linking the quality of document - or the documentation of cancer staging to quality of care, and is there a supporting body of evidence, if that's the criteria then I think, you know, or at least my recommendation would still stand where it is.

So I'm...

Female: I think you've done an eloquent job of sort of raising the issue that we're facing here. Thank you. That's a great way to put it.

(Brian Lloyd): So having said that I don't know if there is an option for us to make a recommendation that is desirable in the pursuit of being able to get after the greater good. I don't know how you get to

one without the other. I don't know how you'd get to that linkage without first obtaining that information, as the commenters are suggesting.

Angela Franklin: So yes, it's definitely plausible for you to make a recommendation for endorsement. If there's something that the measured developer can provide to strengthen the measure in your opinion, you could let them know at this point.

Male: Well, I think that, you know, the document provided is lengthy. But, you know, we have to evaluate the measure as what's provided in the documents. And if, you know, if someone who was developing the document must have felt it was important enough to go into some of these nuances of the staging and the different types of staging and so forth it went into creating the numerator, the denominator, the metric.

So I think if - while we're really not looking ((inaudible)) in fact granular, we're really looking to get some documentation. Then I think the document needs to be more concrete. That that is actually what you're going to measure.

But then my follow-up question to you would be is ideally here, you're right, we need to start somewhere. And secondly, we obviously want to try to leverage behavior in a direction which is going to be positive in terms of the ultimate improvement in quality goal.

And I guess the question is to what extent do you have a measure that it becomes more important to fill in the stage box. Even though it may be a mish-mash what goes into the stage box as opposed to sort of filling with something that actually is going to be accurate and used in the way that you expect it to be used.

So I guess that's a little bit of the issue I'm struggling here. If you're like a busy clinician, okay, I know I'm going to get evaluated, I need to put something in that box. It's kind of like I'm not sure

I'm going to get back to this later, I'm just going to fill in a, you know, my best guess at this point is that will still the fill the metric, but is that the type of documentation you're trying to leverage?

Dr. (Jim Hammond): So this is (Jim Hammond) again. You know, listen, I don't want someone just to fill in a box, you know, for the sake of filling it in. I mean, that's - if they fill it in incorrectly that's an error, right?

But I think by making people think, wow, you know, this is an important piece of information. You know, I better try to stage the patient as accurately as I can. And, you know, I might need to actually, you know, look up something in the staging manual and fill it out properly.

You know, and that information, like you said, you know, is going to, you know, be carried along with the patient over the whole course of their illness. You know, if we're not trying to, you know, send a message to people, what we think might potentially be important, you know, this seems like an important part of developing a care plan for a patient.

And you know, I don't know, you know, I've been involved in this process before. I'm sure it's been changed over time. You know, from my opinion, and that's all this is, is that, you know, to rigidly adhere, you know, and not sort of, you know, what the requirements are for endorsement of a measure. I mean, it seems to me like it's a bit of a relative process, not an absolute process.

Male: Well, I know the criteria that NQF lays out to sort of evaluate the measures, clearly there's a judgment involved, but it seems to be quite like explicit in the - and, you know, as much as I think the opinion of an experienced clinician is certainly - no one would say that's not important, but again, the criteria used to evaluate the measures that opinion is going to be viewed as kind of a lower level evidence to sort of support a direction.

I think that in as far as the - your best guess at the stage, I think that the, you know, again, I keep going back and saying you've got a measure that has no exclusions to it. And that - so there's going to be - as described, again, we're responding to what's written down in terms of how you're going to construct a numerator and denominator with no exclusions.

One might argue that I'm seeing somewhat for the first time there's incomplete staging data. And then in reality what would be the type of behavior you'd want to leverage is that - is an option to say, you know, staging information is incomplete for something like that. So it would reflect the fact that, you know, someone else is going to fall from this person. I evaluated them.

Other things are going to be done which, you know, are not really part of this initial evaluation, but that will not leverage the person to sort of just, you know, say well, I think it's - I kind of think it's Stage 3 disease.

Dr. (Jim Hammond): So this is (Jim Hammond) again. So, you know, all of the AJCC staging includes a TX, NX, MX, which means, you know, that it can't yet be assessed. So I mean, that would be, you know, an appropriate stage, you know, to report if that's, you know, where the patient was in their valuation.

Dr. Dr. Michael Hassett: This is Michael Hassett. Also, I think the stage, at least the measure is formulated focuses on documentation within the first year to give the provider time to gather the information that would be needed to come up with their determination about stage.

Samantha Tierney: This is Sam Tierney with the AMA. If I could just make another comment. I wonder if it's possible for NQF staff, if they haven't already, to share with the steering committee that exception that is noted in the evidence report.

I know from colleagues who attended the recent care coordination steering committee were obviously a lot of randomized control trials haven't been conducted that they seem to conduct the review in two phases, looking at whether or not the evidence met the requirements for quantity, quality and consistency. And if not they look to see whether they felt that an exception applied.

And I think for some of these measures, and even the one you're going to review next, you're not necessarily going to have clinical trial data that, you know, assesses patients with stage and patients who don't have a stage to try to determine their ultimate outcomes.

I think that that's just going to be a situation where evidence is lacking. And so I guess I would say that it seems like the exception to NQF policy and requirements would apply here. And perhaps the steering committee could benefit from that full information that's included in the report about when that exception is appropriate.

Female: We'll be happy to share that with the committee. We had sent it out, actually, in the beginning but we can send it out again.

Angela Franklin: Right, right. And the committee should have it in front of you. It's for you want to make sure that in the case where you believe that there would not be evidence of - in the - as the high quality randomized controlled trials that the measure is important enough and reliable enough to continue to review and still find it to be recommended for endorsement.

Male: Could you explain what you mean by continued to review?

Angela Franklin: Well, we do have two must-pass criteria. And those are the evidence important section and the second section which is reliability and validity in the form of the measure. And if those two criterion are not met, then discussing feasibility and usability is not engaged in.

Now we're in an initial discussion right now and at the in-person meeting this will be subject to a final vote. So it will be more critical as we get to the in-person meeting to make your decisions about the measure based on those two must-pass criteria.

And then if the measure passes that criteria then moving on to feasibility and usability.

Male: Okay, again, new to this process. But can I make a proposed recommendation?

Angela Franklin: Sure.

Male: And could we say that it does not pass the criteria but we could support the cataloguing of this information? Is that...

Angela Franklin: Yes. That would be a possibility.

Male: Well, I'm going to be so bold as to make that recommendation and get my fellow committee members to react to it.

Female: So I do think this distinction between a new measure and a maintenance measure is important, though. Because I think the message also is that, you know, there may not - for a lot of these measures we wouldn't expect randomized clinical trials. But we would begin to see evidence that's documenting that these measures are associated with important patient outcomes.

And as these measures are used I think there's more - there should be a higher standard for what kind of evidence is presented for validity of these measures, as measures of quality. So these are maintenance measures that we're talking about. So yes, I need some guidance in how much expectation I should have as a reviewer that we should start seeing something beyond face validity.

Which most of these measures have. Nobody can argue with the importance of these things. But the question really is, is whether or not they're associated with important outcomes. And ask these measures have been out there for awhile it seems like we should start seeing some of that data in these reports or in these documents.

And that's what I'm not seeing for almost - for really all of these measures that we're reviewing today.

Female: Or I think - I mean, I would agree and I would think that at a minimum how they, you know, how they actually have contributed, say, to quality improvement.

Female: That's right. Absolutely. And that's what we're not seeing. And so if this was a new measure and we're going to talk about a new measure coming up my standards are much lower because we don't think there's data. But in this case these have been sued in large sample quality improvement or, you know, performance initiatives for three years now.

And it seems like we should start seeing some of that data. Or at least in the requirements that I'm looking at that's been given us by NQF, I'm expected to see some actual data from the use of these measures, which I'm not really seeing other than this very broad swipe at reliability and the face validity.

Angela Franklin: Okay. Well, in the interest of time if there's - is there a quick summary of how we - how the workgroup feels, the sense of the measure on this one?

Male: Can I make one last question, which I think is follow-up to the very relevant comment we've just made. I mean, what is, you know, what is the expectation regarding these maintenance measures

in terms of, you know, in terms of what the developer needs to deliver on and the timeframe to do it.

Angela Franklin: In terms of - in the interim between the first endorsement and the ((inaudible)) endorsement?

Male: Yes.

Angela Franklin: Okay.

Male: Yes, in terms of the - if theoretically you got something, well, we kind of have face validity. We'll give it a go, we'll see how it goes, and it's with the implication that you're going to see some additional data at some point to say, oh yeah, this is making, you know, this is leading to improved outcome, for example.

And I guess that when you have these things which have been out there and at least based on what, you know, we've been given our view it seems that the main metric of validity is that a bunch of experts think it's important. But it's not like, you know, on the first roll-out that's efficient in terms of validity.

Once you've been working with a measure for awhile you'd think at some point you would start to say okay, we'll look to see how, you know, the people that did and didn't do it this way and then, you know, what impact and outcome there is. And I guess what's unclear is, well, is that the idea that that's going to take ten years and three years is too short of a timeline to do that.

Like what is the expectations of the developer in terms of providing that additional validity data, you know, as synchronized with these views.

Female: We do as steering committees expect to see some interim data that supports the measure in the - between the first endorsement and the maintenance review. So those - that is the expectation. And developers can indicate that on their forms or indicate, you know, why there is maybe not that - the expectation is not realistic for us.

So there is an expectation for data, but that can also be explained by the developer if it is not apparent in the measure.

Dr. Michael Hassett: This is Michael Hassett. The request for outcomes linkage to measurement over time I think is certainly highly desirable. But I think it's a particular challenge for cancer related measures given that the most common outcome in cancer survival can be far off and difficult to assess. And there are really no great interim validated outcomes measures to connect with these process based measures that we're talking about.

So it's a challenge that I think almost every cancer ((inaudible)) measurable phase whereas if one was looking at a measure of diabetes there are better surrogate outcome measures such as hemoglobin A1c. The same could be said for vascular disease and surgery and (30-day) mortality.

But with cancer related measures, it's quite difficult to come up with measures that are not very far off or confounded by decisions ((inaudible)) and other factors.

Angela Franklin: Thanks. We still have two measures to review. Can we have specific questions for the developer on this measure to address some of our concerns at this time?

Operator: And who did you need open?

Angela Franklin: I think the line for the AMA PCPI folks are open.

Operator: Okay. And all those lines are still open.

Angela Franklin: Okay.

Samantha Tierney: Angela, this is Sam Tierney again. I guess I just wanted to comment on the discussion that just took place. I'm not aware of any NQF guidance that the endorsement criteria for new measures versus maintenance measures is different. So I guess if you...

Angela Franklin: There is no difference, in answer to your question. There's no difference, however there's an expectation in terms of evidence provided.

Samantha Tierney: I also...

(Lindsey): This is (Lindsey) from NQF. I think also you'll see on the measure submission form we ask for evidence from current use on the performance gap and on disparities. I'm not sure that we specifically asked for ((inaudible)) liability to your point, Sam.

Samantha Tierney: Yes, and we do have information, just to clarify, related to the current use of the measures in terms of performance rates. I'm not aware again of any guidance related to need for us to document as part of our testing projects or a part of our reliability projects, or even the testing guidance that NQF put out that the measure actually results in improved outcomes.

As Dr. Hassett said, obviously that would be ideal. But can ((inaudible)) tests of that nature I think is extremely cost intensive. And actually to my knowledge in the medical literature they've only been done in a few existing measures out there. And many of them are publicly reported at the hospital level.

So I guess I just wanted to share that. It seems like there may be some - again, it seems like folks are maybe holding the maintenance measures to a higher standard. I'm not necessarily sure that that's provided for in the NQF criteria or guidance. And I think we do have data regarding the use of the measures which certainly you would expect from measures that have been out there for some time. So I think we've addressed that.

Angela Franklin: Okay, thank you. We still have on our discussion 381, oncology treatment summary communication. And (Jennifer Mallun) I have as the lead discussant for that. Is that correct? Are you all on?

Dr. (Jennifer Mallun): Sorry, I'm on mute again.

Angela Franklin: Okay, sorry about that.

Dr. (Jennifer Mallun): Let's see, I'm trying to see if I can find - I didn't know I was discussing that one.

Angela Franklin: (Heidi Donovan), I'm sorry. (Heidi Donovan) is...

Dr. (Jennifer Mallun): Oh good. I was going to say, that was news to me.

Angela Franklin: Sorry about that. So if you could take us through it.

(Heidi Donovan): Sure. I think some of the issues apply, you know, that we've been having these conversations about. I appreciate all the input from the sponsors of the measures and I'll be rethinking some of my comments and rereading through things again.

So I think the two things that stand out for me on this measure and this is the measure of treatment summary communication for radiation oncology, is first I'm a little bit concerned about the - I was concerned about the ambiguity, perhaps, of some of the aspects of this measure.

Now at the time that I was reviewing this I was not clear on what (inter-rater) reliability represented for these sponsors. And I just want to confirm again that (inter-rater) reliability for this measure does mean two independent raters go in, extract data from the chart that address the three different components of assessment for the numerator -- the dose delivered, the roles and assessment of tolerance, to progress towards treatment goals and subsequent care plan.

They do report that they have good reliability, but I just wanted to confirm that that is the process that was used. So that's one of my primary questions. And then the other one just like in all these other measures is that there really is not evidence that's provided or the importance of impact of this measure other than from a clinical practice guideline that does not seem to have a systematic review of the literature.

And so that was - those were my two primary concerns for this one.

Angela Franklin: Thank you. Did you have an initial recommendation?

(Heidi Donovan): Well, I guess based on those two criteria I felt like it didn't meet the criteria that we were asked to hold measures to. So I had a no.

Angela Franklin: Comments from the rest of the workgroup? Any comments on this from the rest of the workgroup?

Female: I think I - I mean, my main question with this one was -- and there wasn't a lot of evidence provided about this -- but I think this has been another standard expectation for a long time. And I'm not sure there's any variability on this measure.

Angela Franklin: Other workgroup comments?

Male: No. I thought it was largely a documentation measure that is - I thought it was, you know, certainly potentially important as a component of kind of care coordination communication which would be a priority. But it may be a necessary thing but not a sufficient thing to indicate that.

And I, you know, and I thought some of the evidence provided, I think was kind of lowish or insufficient. And so I similarly thought the vote is a no for endorsement.

Angela Franklin: All right. Additional comments or do we have a response from the developer on variability and evidence?

Dr. (Jim Hammond): Sam, this is (Jim Hammond). Can you speak to the question about the (inter-rater) reliability?

Samantha Tierney: Yes.

Dr. (Jim Hammond): And then I can address the other - some of the other issues.

Samantha Tierney: Okay, sure. So I can't remember the reviewer's name and I apologize. But the way you described it is absolutely correct. However, I just wanted to mention that these chart reviewers, these chart abstractors, excuse me, are going to in and they're making sure that there is agreement between what they found in these charts.

So you're absolutely correct. One abstractor goes in and then the second abstractor goes in. And they basically compare their results and then come up with a rate of agreement. So that's your answer to your (inter-rater) reliability testing question.

Female: Okay. So they would look for those three factors in the chart.

Samantha Tierney: Yes. They're looking for the data elements in the chart.

Female: Right, okay. Thank you very much.

Dr. (Jim Hammond): So this is (Jim Hammond) again. I maybe can speak to some of the other issues that were raised. While it may be true that there's a perception that this is done commonly in radiation oncology practices, in fact that some of the data from -- I'm not remembering the project that was done with (Rand) and ASCO -- found that rate - I want to say it was around 65%. So it's not as high as you might think it is.

The other thing that's worth looking at is what we're looking for in this treatment summary. So just to speak, I know there was a question about the doses is the doses. And that should be specified. Subsequent care plan really refers to follow-up.

You know, were there any plans made for follow-up. And then how the patient tolerated the treatment and was the treatment completed, you know, gets the second point.

The measure has specified - also asked that this communication be timely within one month of completing treatment. And also that it be sent - not only that it just be, you know, put into the patient's chart in the radiation oncology department, but that it be sent to the physicians who are, you know, involved in the care of the patient and also to the patient themselves.

And I think that, you know, if we really think care coordination is important, I don't know, you know, what's more important than providing this sort of, you know, communication to the patient, to the providers in a timely fashion. So I wouldn't say that it's potentially important. I would say that it's critically important.

And that it's absolutely necessary to provide at least this minimum information so that, you know, people are aware of what happened while the patient was, you know, undergoing their treatment. I'm not sure that there are a lot of care coordination measures out there that, you know, necessarily have, again, high level evidence to support them. But I think this is critically important.

Male: (Jim), how often do you make a decision and follow-up on a patient that you're trying to make the (treatment) decision on - just based on the treatment summary?

Dr. (Jim Hammond): So are you asking me if I need to retreat someone?

Male: Yes. Like if you're looking, you're going to make a treatment decision ((inaudible))...

Dr. (Jim Hammond): Once you get the place to start...

Male: ...piece of information or would you always request additional information?

Dr. (Jim Hammond): Well, I think it's a place to start. I mean, I don't think that the patient is going to want a copy of their treatment plan. But, you know, as a radiation oncologist, if I was going to retreat someone, you know, I would, you know, need that information.

But, you know, certainly if I have a treatment summary I know who to contact. I know at least some basic outline. I know what total dose they received and, you know, it's a starting point.

But, you know, I think that, you know, most - other oncologists who aren't radiation oncologists would be, you know, very - I would think that they would want this information, know this information about their patients who are treated. I don't, Michael, if you wanted to comment on that.

Dr. (Jennifer Mallun): Could I interject? This is (Jen Mallun). So I just wanted to clarify in the (Nick) study, actually the indicator, the denominator was actually the treatment summaries themselves. And the numerator was the percent of treatment summaries that included all three components, the dose, the dose per fraction and this site.

So the fact that it was 50 or 60%, depending on the region of treatment summaries didn't have all three pieces of information in them. So I think that was a separate issue.

I think that having treatment summaries is really important. I guess my question just was I was under the impression, and perhaps this is my misunderstanding, that in order to get reimbursement for delivering the radiation therapy that these summaries were required.

Dr. (Jim Hammond): I'm not aware of that being the case.

Dr. (Jennifer Mallun): Okay.

Dr. (Jim Hammond): So it's not a requirement.

Dr. (Jennifer Mallun): Because I think that's sort of similar to an operative report, right? That's kind of a basic requirement.

Dr. (Jim Hammond): I mean, the way that, you know, this is handled on a reimbursement basis is that typically, you know, we submit a charge for what's called a weekly management. And those, that

code is the same for the first week and the last week. So there's no distinction made at the end of treatment.

Dr. (Jennifer Mallun): Okay.

Dr. (Jim Hammond): If you want to start reimbursing us for that.

Angela Franklin: Questions? Additional questions from the workgroup on this one? All right, so we'll move on to Barrett's esophagus. This is our new measure that's being reviewed by committee.

(Brian Lloyd): Yes, so this is (Brian Lloyd). I think it's important for me to disclose before we get started that I am a member of the College of American Pathologists and have been so for many years. Also a member of the American Society for Clinical Pathology. So unless that presents a problem, I'll proceed.

Angela Franklin: No problem.

(Brian Lloyd): Okay, in general I'll keep my comments brief on the front end. I think the things that we've heard on some of the other measures apply here. A, it's a new measure. B, my overall impression is that this would be desirable to catalogue. I think it's an important step. So it would be much more desirable to have the Barrett's esophagus documented versus not having it documented.

However, I would say in holding to the exercise that we're asked to evaluate, I think three of the four committee members said no, it is does not meet the criteria necessary for endorsement. So that's the bottom line. Having said that I think a lot of that is driven by the fact that it's new and many of the criterions that we were asked to evaluate are still yet to be determined. But again, I would emphasize I think it's desirable to catalogue.

As I walked through it some of my general impressions and knowing that the college is on the phone that I might ask them to respond to would be here's some of the comments that might help prepare the larger group as we think about meeting together in person. I was curious as to a couple of issues.

One, the esophageal biopsy report under Section 2A 1.1 says we need to know whether it's present, absent or indefinite, but it really didn't get into grading. And I was curious as to why the college though that the low versus high grade dysplasia would not be important.

Number two, I was hoping as I was walking through this document to better understand perhaps some of the confounders that I might understand. I was trying to link Barrett's esophagus to a quality outcome. Meaning document it ultimately either results in an endpoint of getting treatment or prevention of adenocarcinoma (aided).

So I know there's a lot that goes into that in terms of frequency of biopsy as well as the sampling technique. And just as a finer point on the sampling, I was hoping that in here somewhere that I would have seen a mention of their ((inaudible)) of the biopsy side as being an important component of this. So all of that being said, I think that in general there as some disagreement, but the high impact fell into the moderate category.

I believe that the quantity of the evidence, there was two articles cited. One from 2001 which talked about the quality of - it was a mention of the quality of care in Barrett's esophagus. But again, I was having a problem taking just that topic line and trying to say yes that really nails it down in terms of how putting this down on paper and cataloguing that information led to an improved outcome.

And another - number two, the Netherlands study from 2010 which talked about the standardized review of endoscopy and pathology reports, I wasn't clear on the measure developers - the intent

of providing that article to the review committee as I was asking myself the question of what was in that article that really elaborated on the performance gap.

So I very much appreciate the colleges - or the developers, measure developers' impressions on that as well. So there were two articles which kind of forced me to put the quantity at moderate. And if I was missing some information, please let me know. And I believe that I've covered my summary points there.

Angela Franklin: (Danielle), can you open the lines for - actually all the lines, please.

Operator: Absolutely, one moment.

Angela Franklin: Thank you. So do we have a response from the - do we have other workgroup comments on this measure?

Operator: All lines are open.

Angela Franklin: Any response from the developer of this measure?

Female: This is ((inaudible)) with the College of American Pathologists. I'm trying to figure out where to start. I think that - so we have a lot of trouble finding evidence in gaps of care and we did have some evidence in this case.

The Netherlands study though, it was published after we developed the measure, provided some additional evidence of a gap in care. They did have - they did find the information in there and I'd have to go back into the article to find exactly what - where it is. Of...

Female: I'm having a very difficult time hearing you. Can you speak up?

Female: Sure. Can you hear me now?

Female: A little bit.

Female: Is this any better?

Female: Much better, thank you.

Angela Franklin: Yes.

Female: So the Netherlands study had some information about the absence of this statement. And having that statement, again, is pretty critical to deciding next steps in care in terms of frequency biopsies as was mentioned.

We didn't - some of the other elements that were talked about including in the measure, it was hard to do that in the context of the denominator and coding choices we have. It just made it - it was extremely - it would have been a very complicated measure to vote for. So we were somewhat limited in what we could do.

(Brian Lloyd): If I could just pause right there. This is (Brian Lloyd). And I understand - I think I understand, are you saying that the ICD-9 codes or the proposed ICD-10 codes do not make the distinction between high grade and low grade dysplasia.

Female: The ICD-9 ones do not. The ICD-10 is going to complicate this and I actually don't want to get into that right now because it's - the ICD-10 coding for Barrett's is insufficient. We'll just put it that way.

(Brian Lloyd): Okay.

Female: But the current coding for Barrett's in ICD-9 does not distinguish between the grades. So, you know, asking to include that grade in the measure would only account for those cases where it was - dysplasia was present. Wouldn't affect the ((inaudible)).

So we - and we weren't really allowed to use secondary codes to separate them. I don't know how to better explain it than that. But ((inaudible))...

(Brian Lloyd): No, I think ((inaudible))...

Female: It was difficult to code in the first place and then to code for all these variables would have been hard to do if not impossible to do.

Angela Franklin: Okay. Any other questions for the developer? All right, thank you. And as Dr. (Lloyd) noted, this is a branded measure.

(Brian Lloyd): Yes.

Angela Franklin: Go ahead.

(Brian Lloyd): I'm sorry. Did you all have evidence - did the college have evidence that said documentation of dysplasia resulted in earlier intervention and then additional information, which I think exists, that says early intervention potentially resulted in a better treatment outcome?

Female: I thin that information is in the actual guideline in the ((inaudible)) article. And I will go back and check on that.

(Brian Lloyd): That might be useful to articulate. Is it in the (Ochman) or the (Curbers) article?

Female: I think it's in the guideline itself which is (Samplen). I'm not sure if I'm pronouncing it correctly.

(Brian Lloyd): Is the guideline a - are you talking about the American College of Gastroenterology guideline?

Female: That was the guideline we used.

(Brian Lloyd): Okay, and is that guideline grounded in evidence basis or is that a consensus opinion?

Female: I think in evidence.

(Brian Lloyd): That would be useful I think for the broader groups to have in making a determination I would submit to you.

Female: Okay.

Angela Franklin: So you can provide that. And I also want to remind everyone it's a new measure. It is untested but we did allow it in for consideration because it is in the federal program, the (PTRS) program. And it is only eligible for time limited endorsement. So I just wanted to remind everyone of that.

(Brian Lloyd): So if it's my job to make a recommendation it would be it does not meet criterion but is useful to catalogue.

Angela Franklin: Okay. Other comments from the workgroup?

Female: I thought since this was a time limited measure and it would have the time to provide additional information it seemed like it would kind of meet the bar of being able to be included at this point and give them the time to develop the additional information on validity and reliability, et cetera.

Angela Franklin: That is correct. Other comments from the committee?

(Brian Lloyd): And I'm sorry. This is (Brian Lloyd) again. For completion sake I would say I think it is important to include the contingencies that we described. Meaning determining whether the guidelines that were being proposed as evidence are either evidence based or consensus based I think is an important feature to know.

Angela Franklin: Okay. And say, is that something you can provide to us offline by email, rather?

Female: Sure.

Angela Franklin: Okay. Any other considerations? Okay, so that brings us to the end of our discussion of all the measures and at this time we have a - some time for public comment. If there are members of the public on the call that would like to provide additional comments, you may do so at this time.

Operator: And all lines are still open.

Angela Franklin: Any other? Okay, hearing then - (Adele), do you want to tell us about next step for the workgroup?

(Adele): Sure. In-person meeting is in two weeks. If you haven't made any travel arrangements or anything like that then please contact us and we'll get that set up as soon as we can. We'll also be providing the preliminary results right before the in-person meeting and we'll be combining all

four workgroups preliminary results so that you will have it before the meeting in order to help you do your final voting.

And (Lindsey), if you're there, do you have anything to add?

(Lindsey): This is (Lindsey). I don't have anything to add. Just to thank you all for working overtime and we really appreciate your efforts here.

Angela Franklin: Is there an opportunity for us to modify our preliminary reviews or just leave them as is with...

(Lindsey): Absolutely...

Female: Oh, go ahead, (Lindsey).

(Lindsey): Absolutely. If you want to modify your votes as a result of this conversation you can go ahead and submit them. If you'll just make a comment on there so that we'll know which of your last version of voting, we'd really appreciate that.

Angela Franklin: Okay, so going in to the same SurveyMonkey and...

(Lindsey): Exactly.

Angela Franklin: Great. And when do you need those last?

(Lindsey): If we could get them by Tuesday of next week that would be really helpful. But if you need extra time if you could just send us an email and let us know.

Angela Franklin: Okay. Any more questions about the next step? Okay, hearing none that will be the end of our call and please feel free to email us as well if you have any additional concerns or questions. With that we'll go ahead and conclude the call and thank you again for working overtime on this one.

Operator: And once again, ladies and gentlemen, that does conclude today's call. Thank you again for your participation. You may now disconnect.

Angela Franklin: Thank you.

END