NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
February 4, 2014
1:00 p.m. ET

Katie Streeter: Hi everyone, this is Katie Streeter calling here at NQF with the Endocrine project staff. Before we begin, I'd to introduce Karen Johnson who is the new senior director on this project.

Karen Johnson: Hello everybody. Reva is still alive and well, no worries there but we just – because of internal things we had to switch around. So I will be the senior on the project but if we need to call on Reva for anything, she's going to be able to us as well. So I'm excited to work on this project with you.

Katie Streeter: So thank you all for submitting your workgroup comments for the measures that are being reviewed today in this call. I just want to remind you what the role that we are discussing is. And that is to begin and lead the discussion of measure evaluation and how well the measure meets the criteria. To get a brief description of the measure, to summarize the evaluation of each criteria based on the workgroup comment that you submitted.

To highlight areas of concern or differences of opinion that may have been noted in the comments and also to review issues or questions that were noted in the staff review.

Does anyone have any questions before we begin? And actually if we want to do just a quick roll call, people from the committee who are on the call could you please state your name.

Bill Taylor: I'm Bill Taylor.

Katie Streeter: Hi Bill.

Anna McCollister-Slipp: This is Anna McCollister-Slipp.

Katie Streeter: Hello.

Vicky Ducworth: This is Vicky Ducworth.

Katie Streeter: Great. Do we have Ingrid?

Ingrid Duva: Yes. This Ingrid Duva, sorry.

Katie Streeter: Great welcome.

And just so, structure wise, we are reviewing five measures today and since we have about two hours that's going to give us about 20 minutes or so per measure. And the first measure on our agenda is measure number 0059. And we have Ingrid and Anna as lead discussing for this measure.

So what we'd like you to do is walk us through the criteria and review your comments that were submitted by the workgroup and also the comments and corrections that were noted in the staff review.

Female: And the developers are on the line, (Bob) and (Rita) do you have open lines?

(Bob): We're testing that right now. Can you hear us?

Katie Streeter: We can, great so I have (Bob) and (Rita) from NCQA on the line to answer any questions that you all might have throughout this discussion today.

(Bob): Just to clarify, we have actually a whole team of folks here including (Mary Barton) our vice president for performance measurement, myself, (Rita) and some folks here in our analysis staff, and measure development staff.

Katie Streeter: All right, great, thank you, (Bob).

So Ingrid would you want to start us off here by reviewing criteria number one importance to measure and report and specifically we'll look at 1A evidence?

Ingrid Duva: OK, so I just want to make sure I'm on the right measure. 0059, the Comprehensive, that's the one that you have on the screen, I thought in the agenda you had a different measure first.

Katie Streeter: Yes, let's start with 0059.

Ingrid Duva: OK, sorry. All right well I apologize I've not been able to get the measure up since the workgroup's comments went in. So I'm going to be a little bit rough here, presenting back to you all.

OK, so to get started, I'm going to have to use what's on your screen. This is 005 – can you scroll it up because I can't see it on my own computer, I was unable to get it to open on the site so thanks.

Bill Taylor: And I'm sorry, this is Bill I came in late, can I find out how to get into the webinar that you're showing?

Katie Streeter: I'll go ahead and send the web link to you.

Bill Taylor: Thank you.

Ingrid Duva: So my apologies. I'm just going to need to start at the beginning, I was prepared for the other measure and this specific one I cannot open this morning.

But this is measure 0059, comprehensive diabetes care and it's for poor control, so greater than 9 percent. And the description provided is it presented the patient as 18 to 75 years of age with diabetes. His most recent HbA1c, during the measurement year was greater than 9 percent.

So this is not a new measure just to get started. It's been in existence, then if we can scroll down, I can start going through the workgroup comments.

Katie Streeter: Would you like to go all the way down to the workgroup comments?

Ingrid Duva: That would be best. Since I haven't read them yet I apologize. And perhaps if my colleague wants to take over that would be fine if you're more prepared.

Anna McCollister-Slipp: Well I have it printed out in front of me. Which would be helpful.

Ingrid Duva: That would be helpful yes.

Anna McCollister-Slipp: If you like, I can actually email to you the Word document.

Ingrid Duva: That would be great.

Anna McCollister-Slipp: I don't know if I have your email. But if you send it to me, I can email that really quickly.

(Lindy): Hi, this is (Lindy). I'll go ahead and send the Word document to everyone on the call just to make sure that you have them so you can reference them.

Anna McCollister-Slipp: Great, thank you. So again this is Anna McCollister-Slipp. I'm sort of new to this process, I suppose we all are. So in terms of walking through the specific comments of the workgroup, is that the way you need the procedure should I walk through the various categories and the staff comment ahead of time.

Female: I would like to begin with evidence.

Anna McCollister-Slipp: OK, well.

(Crosstalk)

Female: So you'll just start with the workgroup comments, correct?

Female: Yes.

Anna McCollister-Slipp: Not the staff comment.

Female: Yes either one is fine.

Female: Whatever you feel more comfortable doing.

Anna McCollister-Slipp: OK. So for, just starting with the workgroup comments, they are listed here in bullets. First is, is the evidence directly applicable to intermediate outcome being measured? This person said yes like intermediate outcomes in neither short or long term outcome. So that was the question. Second bullet slash comment and A1c reading that is not performed or who's outcome is greater than nine is indicative for controls, for guidelines in the systematic review in the ADA guidelines.

Third is the intermediate outcome proximally closely related to desired outcomes, answering yes, both the desired and intermediate outcome measure, the same component control of diabetes. However why discrepancies in blood glucose are represented by smaller discrepancies in the A1c. I hope this is answering your question.

Katie Streeter: Hey Anna I'm sorry to interrupt you I think maybe and I apologize this is kind of our first time doing it in quite this labor force. So we're all learning together but I don't really – maybe because it's running through each of the comments verbatim.

Do you have just a general kind of (inaudible) of what folks felt about the measure in terms of evidence? Did it seem like it was pretty strong where there are particular questions that need to be discussed by the committee?

And then once we talked about that maybe we'll just actually do some practice of going through the algorithm and see how it would work with the algorithm. Does that …

Anna McCollister-Slipp: Sure. With the caveat that I haven't had a chance this morning. I've been in back to back meeting so I have another chance to go through this summary of it. But I can certainly walk through the comments that I made in my general impression unless other members would like to jump in and add their thoughts, that would be great if that would work.

Katie Streeter: I think that will be fine but we know that we only got these out to you yesterday so I'm not surprised if not everybody have had a chance to look at them in detail. So yes that would be fine and other folks know what they said

and the questions that they had as they were doing it so that's probably better anyway. OK.

Bill Taylor: This is Bill. Is it OK to ask – I also had a lot of difficulty accessing the technology, you know, what's coming from me. Is it OK to back up at this point and have us review some of the more basic things like what it means when it says to have a support measure and like this is a process measure that will be – is it connected to the process of diabetes control or is it connected – does the evidence connect to the outcomes we care about which is people's health or is that – should – is that just part of our homework to jump on all those?

Female: Maybe it would be helpful if we have the developers to give a quick overview. I don't know if that's something that.

Female: Well that was one of the questions that came up about this measure and one of the others because obviously the question is that are measuring, you know, what we say we're measuring so if we want the evidence to support. Are we supporting the intermediate outcome or is the evidence supporting the long term outcome?

So the evidence to point a relationship of this measure to the intermediate outcome or to the long term outcome that it's usually supporting that it gets to the structure process outcome.

Karen Johnson: So this is Karen. Let's back up just a little bit and talk about what we are expecting in terms of evidence for an outcome measure. So this is an outcome measure and if we had a chance to look at the algorithm, it walks us through and it tells us that since we're looking at an outcome measure, what we're really interested in is not so much quality, quantity inconsistency from systematic reviews about evidence, but really just a rationale and a linkage from various or selected processes of care to that outcome.

So generally, for outcome measures, the evidence discussion itself may not be that involved because again we're not asking for lots of literature about all the different interventions that could have an effect on the outcome. It's really just the rationale.

So in our review – but this is not an outcome measure. It is an outcome measure, this is a little bit – I'm sorry …

Female: Intermediate outcome measure.

Karen Johnson: It is an intermediate outcome. So – yes. So we do require quality quantity and consistency. Generally …

Female: So …

Bill Taylor: I'm sorry. This is an outcome because it's measurement – the outcome is – to control the diabetes?

Ingrid Duva: It's considered – it was presented to us and considered intermediate outcome because the outcome is what the A1c level is and so then the long term outcome is the relationship to the health outcomes, is that correct? That's the way it was presented.

Karen Johnson: Yes.

Ingrid: This is Ingrid, sorry.

Karen Johnson: So. Yes. So does having controlled blood glucose reach the desired health outcomes, for example, do people live longer or perhaps do people don't have (expectations) that sort of thing.

Ingrid Duva: I'm sorry to interject but in the interest of summarizing at the same time, so most of the – much of the evidence presented related to the microvascular and some macro outcome which is represented by the lower HbA1c. So the evidence supports that there are HbA1c is related to the better microvascular and some macro outcomes.

Qualification with the, you know, the HbA1c for instance doesn't differentiate variability in glycemia over the time period that it measures. So I think the summary of our responses is that it is evident and then there's evidence that's substantial and it seems to be the best evidence we have at this time.

Female: Yes. And I would agree with that I mean …

Ingrid Duva: Is that I'm trying to answer the question and summarize what our comments were at the same time.

Karen Johnson: So what process going through the evidence algorithm? And make sure that we're all comfortable doing that and we realize that this is my first time doing it as we go through the various measures we'll get past right there.

So in going through the evidence algorithm, can you see it on your screen there?

Ingrid Duva: Yes.

Karen Johnson: OK. So we know the green box is particularly is not a health outcome such as locality that sort of thing. For the next box, it'll take us to box three and what we're curious about is the evidence based on systematic grading of the body of evidence.

So your first question is, was the evidence provided from a systematic review of the evidence?

Ingrid Duva: Well, it's provided from clinical guidelines and they are related systematic reviews was the summary that I can give.

Female: Yes. And the guidelines are based on March scale study.

Karen Johnson: OK.

Female: Right.

Bill Taylor: And – this is Bill, do you want the rest of us to comment or should we go – let you and through all of it and then you'll hear from us or what's the (inaudible)?

Karen Johnson: No. We would like this to be a discussion. So definitely committee members just stand by there.

Ingrid Duva: So and this Ingrid again kind of from a summary standpoint. So one of the questions again is that quality, quantity and consistency which I believe for this measure wasn't directly addressed by the developers. Although the evidence did come from a clinical guideline so I don't know if other members would like to comment on that. That's partly brought the rating down to moderate.

Karen Johnson: OK.

Ingrid Duva: Do other committee members like to comment?

Male: This is (inaudible).

Anna McCollister-Slipp: Because the guidelines were based on pretty extensive systematic review. So …

Female: Correct.

Anna McCollister-Slipp: … I don't – great, I don't know. Does that really bring the quality rating down?

Ingrid Duva: Right. Well, does that address – is that enough to address the quality, you know, the consistency and quantity if it's based on systematic reviews? Because that would be kind of the questionable area. That's the kind of feedback we need.

Karen Johnson: Right. So what we expect to see is we ask specifically for the grade. So the way the algorithm works is if a summary of the QPC was not provided, then you pretty much have to base your findings on the grades from the evidence and recommendations. So does that make sense?

Ingrid Duva: Yes.

Karen Johnson: OK. So I guess one question here, so we know that in box three the systematic review can be associated with the guideline and then box four was the summary the QPC provided yes or no. And in this one, when we did our staff review we thought that it was not, so I guess one of the questions is did anybody disagree with that finding with you? And please ask questions, we're

working through – this is the first – you are literally the first people who have worked through these algorithms with us. We think they were …

Anna McCollister-Slipp: Well I guess, my question is – I mean is it fair to say that it's not high evidence if the clinical guidelines are based on systematic reviews to consider a very high quality from large scale studies then does that – just because the developer didn't look at that, you know, firsthand and reanalyze like you UKPDS and DCCT, does that entirely mean that it's a lower quality because they're going on the advice of the experts at the American Diabetes Association?

Karen Johnson: OK. So that …

Anna McCollister-Slipp: So …

Karen Johnson: … right. No. We actually do not accept the developers to do that kind of work and as a matter of fact that's why we suggest that they will get systematic reviews, again either from guidelines or from other sources.

But the problem with some guidelines is I'm sure you all know is that some guidelines are very evidence-based and some aren't and even the ones that are very much so grade the evidence in different ways and some of them don't grade the evidence.

So again, from the NQF perspective, we would like to know about quality quantity consistency. What this algorithm does is it actually gives an (inaudible) if the quantity quality consistency was not provided by the developer.

So if it's there, then you go to 5A and think about the ratings that were given. And if not – no, sorry, you go to 5 A, B and C and think specifically about the quantity quality and consistency. And if you don't have information to be able to think about 5A and B and C, then you go to box six. And there, you our reliant on the grades of the evidence and recommendation.

Bill Taylor: And that's – this is Bill. That's what we have now because the developers said no to go to box six.

Karen Johnson: Well, let's walk – can we go ahead and just go down for the evidence.

Female: That would be great.

Karen Johnson: Yes. Let's just walk through this a little more quite, I think. And this is also the first time that we're using this (inaudible) attachment so, we've a lot of first time here.

(Crosstalk)

Female: We're all in it together.

Male: Yes.

Female: Yes. We should be able to hyperlink – there you go. OK. Now, go down, OK. So, what they're telling us is that on page 23, in 1A.3 – or sorry, 1A3.1, they're going to tell us about their clinical practice guidelines. And from that, they're supposed to fill out 1A4 and 1A7.

So, if we look at 1A4, we see the guideline through the recording. And 1A42 gives us the specific recommendation. And they have provided info and – that are evidence-grade. So you can see what the grades were from this various guidelines.

Female: Right. So, but their question is – so I think this was rated by the committee as, you know, moderate to high level of evidence because it wasn't explicitly provided the QQC but it was provided through their rating of evidence.

Female: OK. So, according to the algorithm, if they have specifically provided QQC, then the highest rating you could have is moderate.

Female: OK, well that's – but it hasn't. Are there developers on the line?

Male: Yes. We are.

Female: OK. Because most of the evidence is grade B, correct?. I think the (VABLD) the evidence is the highest level of evidence for using he HbA1c.

Male: Right.

Female: As a measure of the clinical outcomes, is that correct?

Male: I think just to provide a little bit of context of our approach on the evidence, this goes from either clinical guidelines or preventive services task force (inaudible). We generally not sure we're in the best position to be doing the quality, quantity and direction of the evidence. And are usually aren't comfortable with the moderate because of the time and effort and it's really outside of our – we think our measure development skills to be characterizing, you know, the body of evidence, but that's kind of our approach in general.

Female: Right.

Male: So …

Female: So, that would be consistent with what the workgroup – how workgroup rated it?

Male: Yes. So, that …

Female: OK.

Male: … is consistent with the thought we would have (inaudible).

Female: So, we can probably move on in to the gap and care.

Female: And just to point out in box six, by not doing the QQC what we really ask is that the guidelines have high quality evidence. And so, consider that criteria, defined right there in box six, as you saw.

Female: OK.

Female: OK. Yes. Now, let's try gaps.

Male: So before we leave it, when we get to box six, we're look – so, we're so suppose to make a judgment about what we think of the quality of the evidence?

Female: Yes.

Male: That work that we …

Female: Yes. So if – if they – if they have grades which may they have provided, then another panel have also thought about quality and quantity and consistency. So if you choose, you can just go based on those grades. That said, you guys have a lot of expertise. And many of you know that their (nature) very well so from – from that you – you would also bring in your personal expertise in order to make that rating.

Bill Taylor: Hello. This is Bill Taylor again. I'm a primary care general internist and don't have the expertise of a lot of the people of the panel but, it's look to me like the only way you can get up to a high quality of evidence and support of the contention that we're looking at here is to ignore a lot of the evidence that's out there and then cherry pick to get the DCCT. And the one study from UKPDS that, you know, showed the maculovascular outcomes might be improve with control. And then extrapolate from that and say there's good evidence with these particular A1c cut points like the nine and the seven and so on that better be included.

Female: So, Bill, you're more comfortable with lower rating than moderate?

Bill Taylor: I am, but I, you know, I don't think I know this literature as well as all the rest of you but that …

Female: I think – I think the question comes and that were we need all guidance about what – I mean, you think cherry pick from what they provided to us?

Bill Taylor: Yes and what the American Diabetes Association shows …

Female: Right

Bill Taylor: … we may meet their guidelines, you can find studies like that. But if you would ignore the court, and you would ignore (advance and you, you know, you ignore all the studies that show that, you know, tight control kills people, then you can say, you know, this moderate evidence that …

Female: So you're talking about …

Bill Taylor: … and you take type 1 diabetes from DCCT and you extrapolate, you know, people we're applying it to and so on.

Female: Oh, I'm sorry …

Female: The number is nine.

Female: Over 9 percent, yes I mean.

Female: This is a 9 percent, this is not a (tight) control group.

Bill Taylor: Right. But seven DCCT was – before, that's the basis for this largely as DCCT and UKPDS and I mean …

Female: Well, I think you bring up an interesting perspective then because we have several managers within our workgroup that are, you know, I think and ultimately were looking, will be looking at what is the best measure that we're interested in. Because we've got three different, you know, hemoglobin A1c measure submitted.

If your – if your talking about the evidence specific to the cut point of 9 percent, not the evidence specific to poor control of glucose as represented by the A1c.

Bill Taylor: Right. But we're talking quality, quantity, consistency, those kind of things and, you know, you take a study like DCCT which was type 1 diabetics before there was an A1c measurement available. And then use that to say the cut point of nine is appropriate. It looks to me like quality, quantity, consistency is a big leap to go from that to say therefore, you know, when your hemoglobin A1c is about nine, we have, you know, strong evidence that you should bring it down.

Female: You should bring it down, OK.

Bill Taylor: I mean, I think you shouldn't but we're not asking that (being a side) question, whereas what's the quality, quantity, consistency of evidence that supports same.

Female: Does the developer have any addition? Seeing if the developer has anything in eerie responses to that in terms of the evidence that they look at.

Female: Sure.

Female: I mean, it seems like that you guys primarily just looked at A8 guidelines which references primarily as Bill said, DCCT and UKPDS, if I remember correctly.

Mary Barton: Thank you for the opportunity to speak. This is Mary Barton. What I think the challenges is to say how to make a measure that would help clinicians figure out who in the practice to focus on, help health plans figure out where there might be a quality gap in the care of their patients, their population at large. Given that – I can't point to something. I think Bill raised an excellent question and I can't raise – I can't point to a study that pinpoints the value of one threshold versus another.

You know, 10 versus 10.2 or nine for example. Any of them are, you know, to some degree or another, picking a number. But the concept, I think, is at least moderately well-supported by the evidence from those trials that keeping people's hemoglobin A1c closer to a healthy range is going to better for them over the long term. And so the question is do you want to measure or not, and so if you want to measure, you have draw a line somewhere.

Bill Taylor: But – I agree, probably. I love agreeing with Mary Baron, not disagreeing, but isn't our – the point – this point in the process to say we're not being ask what's our sentiment to put it all together, we're just asking here what's the quality of evidence that supports them. And here's our opportunity to say even though we've come to that conclusion that, you know, poor control is bad and better control is better, that the actual evidence on which that base is – is got some studies and a lot of understanding the pathophysiology. And – and, you know, sort of hope – hopeful thinking that that ties together this way in the process.

Female: So, Bill. Can I just clarify a point so that we, you know, if we make the rating at least, we're sticking – I mean, I understand what your saying, Bill, if it's about the 9 percent then we needed to say yes or no the evidence is there and if it's high level or not. But I just want to make sure that the first question about the evidence to support the measure focus is, is evidence applicable to the intermediate outcome being measured. And it sound to me that your saying no whereas in our reviews, maybe we wrongly assume (just) because the workgroup responded, "Yes the evidence is applicable to the intermediate outcome", but you're saying it's really not. It's not whether or not it's the high level of evidence you're saying is not applicable to the outcome, it's not telling us that 9 percent is the right cut point for poor control.

Bill Taylor: Well …

Female: Is that right?

Bill Taylor: Yes. I think.

Female: So then, I mean if this answers a little bit different question, it would still be as important to the overall looking at the different measures and which one are important to endorse, would you agree?

Bill Taylor: Yes, I think the specific question is is there a good evidence that 9 percent is the right cut point, we don't know.

Female: OK.

Bill Taylor: It's a broader question the way Mary framed it. Is there evidence that, you know, poor control is worse for you and better control is better for you, then there is evidence to that.

Female: Right, but I think (inaudible) clarification …

Bill Taylor: So can we (provide) those to answer this question that, you know …

Female: Right.

Bill Taylor: … the nine is the cut point. That's only one feature, so we want to say, "Yes …

Female: It's only two separate questions.

Bill Taylor: Yes, (looking) at this evidence, lower is better but, you know, it's not strong evidence, but nine is the particular good place to draw a line in the sand.

Female: So, I guess based on this (example), I guess based on that, I would ask NQF, what is our charge here? Is it to evaluate the evidence as provided or is it to evaluate the evidence in the contexts of our understanding of the goal of this particular quality measure.

Female: Well, it is to evaluate the evidence that was provided to you in light of what the measure actually is. So, in this measure potentially could be use to distinguish between providers, you know, and – and having a cut point of nine or eight or whatever. In this case, it's nine. So the question really is, is the evidence is a strong evidence that – that's the right cut point.

Female: Well, I think what Bill was saying – and again, I could be wrong in my interpretation. But I think what Bill is saying is that there's no evidence to support anyone particular cut point. It's just this one was one that (checked) based on the studies. But, you know, may or may not be bad but is there evidence that suggest that that's appropriate, I mean …

Female: And certainly been agreed upon by those who we've considered experts in the field that we acknowledge, you know.

Female: Yes

Female: Like the Diabetes Association, et cetera. So that we can assign a rating based on that as well.

Bill Taylor: Right and the process allow us to write expert opinion, get to different rating from us than.

Female: Right.

Bill Taylor: Big trial that showed …

Female: Right.

Female: Yes, you guys have that expertise so you can, you can bring that in. And we actually hope that you do bring that to the table.

Female: I mean my assessment, my hunch for what its worth, and I'm not a clinician but I'm a patient, is I'm worthy enough to be involved in something like this and work with the health research. Is that, you know, nine – I'm not a huge fan of A1c but it's the best we have at this point. It's been validated, you know, nine as you know, it's over eight, I'd be a little concerned because of (cord) in advance.

But nine as sufficiency high but I don't see we're putting people at great harm. And evidence suggests that anything above nine is the highest poor quality control and potentially care. So, if we going to draw line somewhere, it seems like reasonable place to draw it.

Bill Taylor: Yes. You probably want to move on and I have no intention of slowing you down …

Female: But I mean.

Bill Taylor: … I guess, nine would be fine with me, but that's just my opinion, right?

Female: So the question is now – we I think – so can we move on from here recognizing that it's moderate to low would basically be the evidence that we would suggest and we've given the developers a chance comment for the evidence supporting specifically 9 percent over just the idea of having poor control versus better control. Can we move on now to the next area. Are we ready to do that?

Female: We definitely need to move on and do you remember that, you know, when the full committee talks in the in-person meeting, you know, this will definitely come up again. So it'll be interesting to see the perceptions of other

folks' comments, the committee as well. But these are great question, you're actually asking the right question. So let's go into the gaps.

Believe me, it will get a little easier for you as we go through it. But I recognize completely that it's (conceiting) at the beginning.

Female: One more is we're all in it together.

Female: OK. You want to go ahead and do gap and care?

Ingrid Duva: I'm sorry, I'm – do you have it back from the screen? No. I'm just trying to get to the comment.

Female: I did e-mail it to you.

Ingrid Duva: You did, OK. Well I can only (tackle) between so many screens at one time time? OK So, this is Ingrid again. And I'm not sure who I'm co-presenting with because – it's Anna, OK.

Anna McCollister-Slipp: Yes, because it changed.

Ingrid Duva: I'm sorry. This time it change if sound disorganized, that's why.

Female: Yes. And my apologies that we did have committee members unable to keep up the commitment and so he dropped out and we didn't have his last minute shift.

Ingrid Duva: I'm getting there, bear with me. It's a gap and performance and – OK. So, the responses to these are basically that there is the large gap in performance, and that this measure had been in place and it doesn't appear that the gap has been impacted that much. So, I think when the in-person actually come in, that is a bit scary. So those were the comment, so there is a gap in performance does exist between health plans and between providers.

Now, I've did have a question about this specifically I guess. I think that the developers check and maybe I misunderstood it or maybe the wrong box is checked, but that the health plan and group between practices and between

providers. But I didn't see data and all three of these groups. Is that – am I correct or incorrect in that.

Female: So, its true that we have – there is a physician level measure in the data that we have on the physician level measure is from …

Male: From recognition program.

Ingrid Duva: OK.

Female: And, finally (PQRS) has been …

Male: But we have some data from the PQRS program but it's difficult to interpret, hard to analyze but we offer it there because it's the best (CM) outcome …

Ingrid Duva: OK. And that was the measure that came up slightly and reliable. And I guess we'll get to that later. But there was also the different plans, correct? These different (inaudible) the data was stratified by Medicare and Medicaid Commercial …

Male: Correct.

Ingrid Duva: Is it also offered between practices?

Male: Again, yes.

Ingrid Duva: Yes.

(Crosstalk)

Male: Can you hear us OK?

Ingrid Duva: Yes.

Male: OK, I'm sorry. Right below that, I mean we're on the main submission form with the performance data. And right beneath the data from PQRS is data from our diabetes recognition program.

Ingrid Duva: OK.

Male: It now recognizes 3,676 clinicians who seek recognition in the program and it shows the data over three different years from those physicians.

Ingrid Duva: And those are considered different practices or different physicians? That was my question.

Male: Different position, I'm sorry.

Ingrid Duva: OK. So, I just I thought I saw all three categories checked on form, I just wanted to be sure of that.

Male: (Inaudible) is right.

Ingrid Duva: But the bottom line on the performance gap is there's one that exists. Would the working group committee, do you all agree with that?

Female: Yes.

Ingrid Duva: OK.

Female: Absolutely. And I do think it's frightening.

Ingrid Duva: That was a comment.

Female: That was me.

Female: OK. I think we can probably 1C, high priorities.

So, diabetes is obviously a high impact problem in the workgroup agreed that is was a significant health outcome to be concerned with. There is question about stratifying I think because there were such a performance gap that the stratification seem like less of an issue. Am I correct in summarizing that?

Female: Yes. I mean, just if I remember correctly, there was compelling data that certain ethnicities were worst than others. I believe it was Hispanics and maybe African-American …

Female: I think it's African-Americans compared to non-Hispanic whites.

(Crosstalk)

Female: Yes. From my perspective and what I can see is that that doesn't necessarily warrant a specific risk stratification by ethnicity. It's just that certain ethnicities are, you know.

Female: OK. So …

Female: … variations in care.

Female: So that was the comment from the workgroup. Does we have any other questions or ready to move to reliability.

Bill Taylor: I don't want to (inaudible). There are different rates of compliance by ethnicity. How do we include that it's not worth stratifying by ethnic classification? Isn't that a situation where you would want to look by ethnicity or am I not understanding what we're saying here.

Female: No, I think you're understanding it correctly. Go ahead.

Female: Or maybe I'm misunderstanding, but – the risk or the number doesn't necessarily change regard – bases on your ethnicity. It's just that quality of care may vary between, you know, people who are predominantly white versus people who live in, you know, or African-American or Hispanic.

But there's a lot of risk associated with A1c, and then the measure doesn't change. It's just performance in various population is different but maybe on misinterpreting what we're doing.

Karen Johnson: So let me give you just a little bit – sorry, this Karen from NQF. Let me give you just a little bit of background in disparities. Sometimes, if submeasures aren't – it's not always obvious from the gap in care information that there may actually be a gap in care. So, you know, performance level maybe high or there may not a lot of variations.

So another way the look at potentially having a gap in care is to see if there are the disparities among certain population subgroups. So that disparity

question can also help you answer the gap question. And then, further, and as we've already – you don't really need that information to know that there is a opportunity for improvement. But then also, just in terms of how the measure might be reported, this question is just to get your opinion on whether this is the kind of measure that you would want to report separately out for disparity.

You know, other than how it relates to gap, if it does. It's not really something that would catch or fill a measure. So, it's in there because we feel that thinking about disparities is important, but really it's the mostly therefore gap and to get your expertise about whether you think in reporting at some point, it might should be reported out separately.

Does that help at all?

Female: Reporting it out separately, do you mean main including information on race as part of the reporting or?

Karen Johnson: Well, in this report, it gets a little tricky because, you know, what were interested in is really the construction of the measure itself. And how anybody might want to later on report out is kind of their business. So, this would just be a nod from you guys if you decided to telling folks who are looking at these numbers, hey you may want to slice and dice by some of this (associate) (inaudible) group. It wouldn't be binding or any way affecting how the measure would be calculated.

Bill Taylor: So – this is Bill again. Help me with this. I mean if we are told that there's disparities in a rate at which these goals are achieved, disparities by ethnicity. Wouldn't we sort of automatically in an important issue like (inaudible) control one and say, this is an issue you might want to slice and dice by ethnicity.

Female: Yes.

Female: I think that would be incredibly helpful information to have for a variety of reasons. I just don't think the number of nine should be different based on race which is the way I'm interpreting the question.

Female: Right.

Female: … risk stratification.

Female: Yes. No, this is not actually getting to risk stratification.

Female: OK.

Female: It's really more about reporting or again on perhaps eliminating a gap for certain sub-populations.

Female: OK, OK. I think we can probably move on to reliability.

Female: Can you scroll the screen up? This is where the physician reliability should have – having a lot of noise. But as far as differentiating between the plans and with higher reliability, greater than 0.7, I believe. And I'm not sure if there's anything else that you want to add about anybody else in the workgroup on the comment.

I think we've already heard from the developed about it?

Female: So, are you comfortable? Let's pull up the algorithm for reliability let's just see this exercise. Just so everybody knows how this evidence work?

Male: Appreciate it, thank you.

Female: Sure.

OK, so reliability. So the first question is, "Are the specifications precise and ambiguous and complete?" So, one of the things that you want to think about is you're thinking about this measures as just about the specification. So, how was the numerator calculated or folks kind of sorted into the denominator, any exclusions? So you think about that and that we actually ask about first, and then we get into reliability testing. And I'm kind of going back and forth myself between the comment. I don't remember particular comments about the specification.

So, let's assume that, for now, that (measure) is (specifying). Then the next question is, "What empirical reliability testing conducted using statistical tests with the measure specified?" So in this case, the answer is yes. So they didn't just give descriptive statistics and that sort of thing. So – let's just stay under box four, "Was reliability testing conducted with the computed performance measure score?" And the answer to that is yes because they did the beta-binomial reliability testing and they described to you what that methodology is and that is an accepted methodology for testing. And then they show you the results.

So box five is, "What's the methodology described and appropriate for assessing the proportion of the variability due to rural differences?", such as the signal (inaudible) and that's actually exactly what they did, the signal (inaudible) analysis kind of popularized by (John (inaudible).

So then that takes us to box six, and box six said based on the reliability statistic and sort of testing, how do you feel the results are? So basically you have to take into account the sampling that was used and then the results of the testing. So they hold you for the plans, you have a lot of plans or you know, a lot that's subjective that's decided that seems like enough to you, and how many patients per plan. They give you that information. They give you the medium number, and then the results.

Female: As you said, the results for the different product lines for plans are pretty good. They all seem to be on the very high (0.79) and higher. And for this particular test, 0.7 is generally considered a pretty good threshold for reliability. But then, the difficulty is at the physician level where you see that the reliability is actually very variable.

So then you have to ask yourself, where in box six do you come down. There's a certainty that the performance scores are reliable. And then you would rate either high or moderate, based on basically how you feel about the results that were presented.

So, for this one, let's think about it a little separately. Is there any concern about the reliability at this plan level?

Female: No.

Female: OK.

Female: It would be from our comments.

Female: OK. And then would you have any concern about the reliability at the physician level?

Female: Based on what the developer provided us, it appears that we don't have a great reliability measure because they are saying that there is a lot of noise or a lot of reasons why it wouldn't be that kind of data. So it doesn't seem like we can judge that.

Vicky Ducworth: Right. This is Vicky Ducworth. I felt like reliability here is difficult just because it doesn't really account for patient factors and we're evaluating, I think kind of provide our performance and quality of care. But you all might have more experience and knowledge there than I would.

But I know when I work with several providers, they would always, like you said, there was just always a lot of noise that they were saying, you know, I can only do so much. I can practice evidence-based guidelines. But then – the data would have difficulty with various patient factors again like socio-economic issues, et cetera.

Female: My impression was that this data was unreliable because of the manner in which the data was collected. It's not necessarily accounting for patient factors, is that correct from the developers?

Female: That's the sense that I had as well. I mean …

Vicky Ducworth: OK, thank you.

Female: I mean, if patient factors may be an issue and always can be, and it is something we have to consider when we look at these measures and how we account for that. But I think in this case, the data that was provided was the best the developers had but it wasn't from their standpoint It had a lot of noise

because it wasn't I guess ideal for circumstances and was it self report, I don't recall. But the developers can comment again if needed.

Male: Sure, and I'll ask for assistance by our Vice President of Analysis. The recognition program is distinct and you know, when Helen Burstin did the diabetic project two years ago, she did kind of do a really nice job of describing the recognition program. So this is voluntary program where clinicians who have – do take care of large (panel) of patients with diabetes, seek recognition and the recognition is achieved through meeting a variety of standards and then also performance on hosted measures, many of which are staying here today.

So in a funny way, it is a clearly a self-select group of people who are choosing to do this from their range of practices, urban, rural, small, big, large. And you know, and I don't know reliability is really driven by the nature of the programs itself, what these are doing I believe a sample of, I think – Robert, is it 30 patients? And so, you know, this could be – it's a random sample so they don't know which patients they are picking. And then they look at the charts and basically do their quality assessment essentially on their own and then we validate that later on.

But that could be driving what appears to be noise by the nature of it. So this isn't like in the PQRS program. A lot is similar because you choose which measures you're going to report on. So, Robert – I don't know if Robert wants to add anything to that.

(Crosstalk)

Male: Go ahead.

Bill Taylor: I was just going to very basic questions, it's Bill again. Reliability is how reliably this reported A1 result that is used to figure this out. How well does that reflect on the ability or the capacity of the physician and the practice to control the A1cs of the patients in practice. Is that right? Health plans, as well as A1cs, how reliably this A1c report was like the degree to which A1cs are controlled in allowing the patients (inaudible). Is that what reliability means here?

Female: I think we agree with you Bill.

Female: The last thing that you said, that last thing. I can't repeat it.

(Crosstalk)

Female: Yes, it's about the reporting whether and not they are able to report that A1cs that reflect the actual care in the office or A1cs that are collected or actually it's for the patient, but you said it better.

Bill Taylor: So the idea is we have good evidence with the plans, reports reliably reflect the A1c control with the plans. The physician reports are problematic because there's people who volunteer to participate and that they bias how we draw our conclusions on physicians.

Female: Well, on the way in which the data is collected whether through the electronic account record, the paper record, administrative record, it reflects that, correct?

Female: That's my understanding. What I'd love to hear – affirms by the developer. It's more an issue of is that the best way to measure quality from, you know, it's the way they did it for the recognition program and effective way in measuring quality, or should you do the way that health plans does.

Female: Well, I think that's a different question. That's a validity question. The reliability is how it's being – how reliably or how consistently you can collect the information and the same information. So for instance if you're collecting it and the developers can comment here. But through many different sources, it can become unreliable. Or if it's a different design in each entity, then it becomes unreliable. But the developers may want to comment.

Robert Saunders: Hi. So this is the developer here. This is Robert Saunders with the analysis group here. So you can see the reliability is measuring the proportion of the variation that is accounted for by the physician as a proportion of the total variability. So we've heard a sort of few different sources – other sources of variability besides the providers. So, if you have differences in your – who is

serving your patient population, and who can contribute to that. There can be different methodological issues in the collection.

If you don't have any concerns about the methodological – methodology of collection of the data from any of the practices that are participating on this. We also don't have particular concerns about differences and the underlying patient populations for this measure.

So you can see – if you look at the history on the physician level of control in the histogram, where you start at (0.64) and the right half of this, you see what is basically a normal distribution. But as (John Adams') work has demonstrated and I think NQF put together a panel that – (HHS) (inaudible) as a tutorial about reliability, that reliability is a metric that is specific to if not – reliability is an attribute or the function of the system itself.

So there are going to be individual cases, in this case a handful of practices that are particularly well-performing on the measure. So there will be a reliability for, you know, Mary as a physician and Bob as a physician and Robert as the physician. And it means if the measure performs poorly for that particular person, that small proportion of the variances explained by that particular person.

We have some practices that are participating in this. This is a voluntary program for people to report in and people who want to be recognized. So they are – this is not a random sample of physicians across the country that are participating, and this is people who want to get recognized by NCQA. And it can be – their willingness to participate in this program may perplex their – showing off how good they are. But if a person who's (inaudible) at 1.0 and being really good at it, but it could also be people that have other incentives to join a program but who may not be quite good at this particular measure.

They're recognized by NCQA across a variety of measures and they don't have to pass on every single measure. So on this particular measure, they may do kind of poorly on 9.0 control but that reflects the mix of measures that they (systemized) and recognized by NCQA.

(Crosstalk)

Karen Johnson: I'm sorry. I'm sorry to interrupt you. This is Karen. I'm sorry. Because of time, we really do need to go on. I think, just in terms of reliability at the physician level, and you take the statement that the developer is actually said, the vast majority of physicians did not need or exceed the minimally accepted threshold of 0.7. And just think about maybe why that might be. And one of the reasons is sample size.

And I think one of the things that's a little confusing about the submissions right now is that we don't really have a good feeling we have with the histogram – but we don't really have a good feeling at how many physicians are in that histogram and how many patients each physician had. But …

Robert Saunders: I think it's – sorry, this is the developer. I think you have exactly that (inaudible). You have the total number of physicians that are recognized and that (were in the form). And the bars represent the percentage of the practices. And so, we can line up the bars to say this is what percentage if we – Mary, if you can pull back up – that the distribution of most of the practices are centered around the points of sides which is very close to 0.7.

There are a handful of sites at 0.3 and at 0.4 about – but those are again, we're explaining why you would have observe some practices may not do as well as others on reliability.

(Crosstalk)

Karen Johnson: I'm sorry. Just so that we can go on with our call, what we'll think about more in the in-person meeting is you know, the actual data. You'll have a chance, you know, you'll have more time to look at this and think about what these numbers are saying and what this (inaudible) are saying to you.

But we'll also have a chance to think about, you know, do you like this measure for health plans but maybe not so much for clinicians. And we could talk about potentially what we might do in that situation.

So with that, I think we …

(Crosstalk)

Female: I guess my only question is could we say that maybe it's a good measure for physicians but just not in the way that the data is collected for this particular recognition program. But if you did (inaudible) from EHRs or claims data, then maybe it would be an OK measure for physicians. I don't know that because I don't know that we have any of that, I mean …

Female: Right.

(Crosstalk)

Karen Johnson: … what you're looking at in evaluating is the measure specified. So, however they have said that things are going to be collected, that's how we see that it's being collected. And one of the things about reliability, one of the basic questions is, is it being collected consistently in such a way that it is fair to pay our provider who is for example pulling their data and using in EHR versus somebody who is doing a medical record.

If the answer to that is no, then you really have to think about whether it is a measure that should be used in a national consensus as a national consensus standard.

Female: And also, I have one more question for the developer. Now, when the providers are pulling the data, (inaudible) performance periods ranged. All they have to do is demonstrate that they met the criteria for 30 consecutive patients, so they can keep pulling back over the course of a performance year until they hit that mark.

Male: They don't get to cherry pick when they start, like they can go fully sample of 30 and that how we found that was (inaudible) still find another time. There's strict procedures about the selection of the starting point within the year for the collection. But it is a 30 consecutive members process.

Karen Johnson: OK. I'd like us to move on to validities if that's OK with everyone.

Female: OK. So for validity in general, the workgroup agreed that high validity was demonstrated. There was three I believe different types of validity submitted by the developers. Space validity amongst experts, (person) correlation was (ran) and then the validity of this measure against other quality measure – diabetes quality measures from the same developers, so whether or not they were correlated with quality.

And so from the data that presented, they were all high. I think there's one moderate measure and that again was related to physician – the physician's action. But I think that the workgroup has a question for whether or not, you know, this clinical outcome can truly represents quality based on, you know, some patient factors that are not included. Or that can't be accounted for, excuse me. So, that was kind of the summary of the comments and what is represented by the developers for validity.

And I don't believe there are any questions about exclusion. I think it was – the policies to go (inaudible) the age range, yes. There was a question of other than 18 to 75 years, and I think that question comes up because it's like the denominators with diagnoses of diabetes within the last year, and then it's only between 18 and 75 and what about the pediatric patients with the question from the workgroup.

Does anybody have any comments, question, concerns?

Shall we move on?

Risk adjustments stratifications, not applicable. Identification is specifically significant, meaningful differences and performance. It seems that – the workgroup agreed that it seems that it does have some measure differences in performance. The question about the different populations within the plans, population of the patients between commercial Medicaid and Medicare?

Let's see – I'm trying to find if there's anything else you need to comment. There's a comment on the multiple data sources? No major validity (threats) and again, anymore comments from the workgroup? Shall I move to feasibility ability?

Female: Now.

Female: Sure.

Female: If no one has – yes. If no one has any comments from that section.

Female: It appears that these data elements are routinely generated through care delivery and they're located in electronic (foreman) at our sources, so that – that demonstrates high feasibility and these are already being collected and reported. So at least from the health perspective seems to be feasible.

And these (usability) are using at for the diabetes recognition here in PQRS. So from the developer perspective, it seems to be useable and then the workgroup comments were questioning how you (solve) the measure in terms of high quality and efficient healthcare in general responses for, – yes, it is high, but the (FDA) wants the poor control would be a good thing to look at to associate (inaudible).

They are already widely adopted by participating organizations. Does anybody have comments about that that I missed?

Female: No. I think you're ….

Female: All right. From …

Male: So, how useful is the measure for developing high quality efficient healthcare? One of the – I guess – from one of our colleague says the benefits substantially are (weighted) consequences. So is that what we're supposed to be answering at this point?

Male: Yes.

(Crosstalk)

Ingrid Duva: That's my understanding. This is Ingrid. Did you want to add any clarification or concern?

Male: I keep speaking, I'm going to slow us down. So tell me if we should do this another time but – the rationale for this that higher, you know, bad controls, (inaudible) so we should encourage people to be good consultants. But we haven't spoken at all about are there any adverse consequences when we embark on the process of trying to help people have those control and how much (inaudible), whether or not, you know, things that aren't being done because this is being done.

Is there any – and it only appears in the analysis by, you know, sort of what was it called? Unexpected, you know, consequences as oppose to what are the benefits and what are the harms.

Yes. I think the answer is yes to this but we – but our process doesn't gives us any opportunity or comment on something, you know, unintended consequences. That was one the one that was asked earlier as oppose to what are the benefits and what are the …

Ingrid Duva: I think some of these concerns may come up with the tight control measure, the less than 8 percent.

Male: Oh good let's just wait …

Ingrid Duva: … you know because we have that measure also.

Male: Yes. Good.

Ingrid Duva: And then we have the general HbA1c measure and hopefully – this I Ingrid. I'm just a workgroup work member but hopefully at some point we'll be discussing the benefit of which one …

Male: Beautiful.

Ingrid Duva: … and what – that's what I assume, NQF, am I wrong?

Female: No, you're correct. And just to make sure everybody understands. When you did things like reliability, you look differently, kind of, with a microscope specs and testing, that sort of thing, and validity. All of those kind of (role) up to talk about the scientific acceptability. For usability, we want you to think

about whether it's in use, whether there's venue for event and is there any evidence of unintended consequences. So you kind of have to think about all three of those together and then make your determination of how you want to rate usability in this. So, hopefully that helps a little bit.

Male: When we hit these topics for tighter control we can – we haven't loss our chance to talked about it because we don't like it here.

Female: Absolutely.

Male: Good.

Female: From the harmonization. So – or harmonized – are they harmonized in their workgroup comment for -- yes, to their distinctly different performance.

Karen Johnson: And this is Karen again. Let's not worry in the workgroup call too much about the harmonization issues. That is – it's a little bit complex and as this come up, we'll walk you through that in the in-person meeting. So, more to come on that in in-person, but we won't worry about it too much in the workgroup.

Female: So, for this measure, I mean, we basically have just discussed all of our work group comments and then got in some input from the developers. Are we finished with this measure for now and comments noted for follow up in our in-person meeting, Karen?

Karen Johnson: Yes. That's what we – will be – we'll summarize our discussion on this measure. And also just so you know, in the in-person meeting, you'll be rating individually all these things, the major sub-criterion. And then at the end, you'll have a chance to just give an overall recommendation for endorsement. So, you have that kind of weight of kind of weigh all the pros and cons and what you personally feel. And then make a recommendation. You spoke about whatever it should or shouldn't be endorsed.

Female: So, I'm going to take Bill's role again of asking another question. So, can you just get this in the (mix) before we move in. Then if we're going to implement, will there be at some point where we discuss whether or not these

measures that are in existence have made an impact already? You know, I'm only looking at the data to see if there's been improvement. That's the question that came up from the staff group and I think from the workgroup too. But I didn't know for this captured in our comments today and it may come up with other measures. (inaudible).

Karen Johnson: Yes. And, just to be clear, that is something. That's actually one of the things that we have for in the usability in use.

Female: OK.

Karen Johnson: Under that criteria and generally what the developers would do is when they are telling us about their gap information and the use and for this measures they did, give three years of data which is great. You know, what would be really (inaudible) if it's at really old measure, it might be interesting to know, right from way back then. But I'm not sure, I don't think most of those will go further back than two, three years generally.

Female: Right. OK.

Karen Johnson: That's definitely a valid question that you can definitely ask with developers. And since they're on the call they know that you're interested in that.

Female: Well, I just think as a committee, we have to consider what to continue and to move performance for quality. But I am ready to move on to the next measure. I didn't mean to really slow us down.

Karen Johnson: That is a great question.

Ingrid Duva: OK. Great. Vicky would you like to walk us through measure 0575? We have – this is Ingrid again as (reference).

Vicky Ducworth: Sure. OK. So I'm just going to start at 1A as an (inaudible) support measure focus. The groups, for the most part, there is consensus that there was sufficient evidence and there were I think a few comments when we try to go back to (inaudible) again.

One comment was the data reference does not demonstrate correlation between A1c levels and other measures as significant. They have (provision) of secondary complications such as quality of life, productivity and ability to function, but I think again for the most part, everyone felt that because this is a process outcomes that – process outcomes more and more (inaudible) demonstrates general quality of care at a provider level. So, does anyone have any comments there?

Ingrid Duva: I have question, this is Ingrid. Are we calling this a process measure versus the other one we called an immediate outcome? Is it right that we're labeling this differently? One is poor control and this is just less than 8 percent. It does seem like there's an important distinctions slightly?

Vicky Ducworth: OK. I'm sorry, yes this is actually the measures (inaudible) comprehensive diabetes care for HbA1c testing for the (percentage) of patients, 18 to 75 years of age with diabetes who did actually received that test. There is a measurement (inaudible). So we're not actually looking at clinical outcomes values. So I thought like that was more of a process measure but …

Ingrid Duva: OK. Thanks.

Vicky Ducworth: Sure, and you're welcome. OK. So let me just go back (inaudible).

Female: And just to make sure everyone – we're looking at measure 0575.

Female: Yes, I think we're getting a little confused on which measure we're looking at here. You should be looking at the hemoglobin control less than 8 percent.

Female: Which is an immediate outcome measure?

Vicky Ducworth: Oh, I'm on – I thought I was suppose to do 0057, my apologies.

Female: Sorry, I didn't mean to throw …

Vicky Ducworth: I was confused, too. (inaudible) you asked, because, yes, we might have gone down the wrong road there. OK. So do you – do you want me to go through the other item then?

(Crosstalk)

Male: I'm sorry. (Looking) at 0075, less than eight, A1c?

Vicky Ducworth: Yes, right. OK. I'll just do that one. I wasn't prepared for that one but …

(Crosstalk)

Male: … intermediate clinical outcome, we're agreed that it's a process measure?

Female: No, we're agreeing that we were on two different measures. So the less than 8 percent is considered an intermediate outcome. But 0057 the comprehensive care would be considered process measure.

So, but since Vicky was prepared to discuss the 0057, should we move forward with that one, Anna do you have (inaudible). Just tell us with where we're going with next?

Female: Do we have someone who can discuss the less than eight because I think it might go rather quickly because it's going to be very similar in many ways to the one that we just discussed, the greater than nine.

Vicky Ducworth: (inaudible) Ingrid, again. You guys keep putting me on the spot with the spot with the one I'm least prepared for us. I was ready for the one you gave me which was (eyes). But I can do that, I just with my – can you e-mail me the comments? It's easier if I can see the comments.

Female: Yes, I sent an e-mail. It came from the (msn@qualityforum.org) mailbox, and all the documents.

Vicky Ducworth: So while I'm doing that, do you guys want to move forward to 0057? Bill and Anna, are you ready with that one? Would that be better?

Anna McCollister-Slipp: We can do that.

Bill Taylor: I'm – this is Bill. I'm quite unprepared. So …

Vicky Ducworth: Oh, you're unprepared oh, sorry.

(Crosstalk)

Vicky Ducworth: I didn't mean to throw you under the bus, I just have to …

Bill Taylor: No, no it's OK. I mean we can talk about it. I think that, you know, we can (inaudible) together but I will have trouble being anymore of a leader than any of the rest of us.

Female: Let me see if I can get this one out.

Vicky Ducworth: Well, I have that – I mean I'm the alleged primary discussor for that one. I'm happy to turn it (inaudible) …

(Crosstalk)

Ingrid Duva: I'm sorry. This is Ingrid. So, Vicky we can go through less than 8 percent. My documents are pulled out. I haven't been able to get two documents up one time yet but it worked. So we'll move on. We've decided that's an intermediate outcome, you know that.

Female: Yes.

Bill Taylor: And everybody's OK with that? It looks like process to me but I don't want to slow us down anymore …

Female: The less than 8 percent, I think that's a differentiator …

Female: … intermediate outcomes, yes. That's like …

Bill Taylor: The outcome being that the diabetes is better controlled? I mean it's really …

Female: Yes, this is the tight control, Bill. The one you have more concerns about I think because it's looking for, let's read the denominator and the numerator. Patients whose recent HbA1c level is less than 8 percent during the measurement year. So the outcome is the result of the HbA1c test, getting desirable control of diabetes.

Bill Taylor: So that's calling that an outcome, desirable control as opposed to …

Female: As an intermediate outcome …

Bill Taylor: … or better vision or, you know, fewer myocardial infarctions or longer life which is usually what I think of as an outcome …

Female: It's intermediate, intermediate outcome. So the idea is that not only are we testing the HbA1c but you're looking for the outcome of OK we're – these many patients have less than 8 percent, you know, are good control. So it's an intermediate quality outcome, the quality being that you're, you know, treating the patient differently based on their A1Cs.

Bill Taylor: OK.

Female: Changing (inaudible), et cetera. OK. I speak out of term but OK so that's – so we're all clear in what we're talking about.

Bill Taylor: Yes.

Female: All right. Moving down through the – the pre-workgroup comments for the evidence is that is an outcome measures so this would be one. It's an intermediate outcome, excuse me, it's one that we desire the quality, quantity consistency and the – so the group reports that the consistency of the body of evidence is a (four). So, it's directly applicable to the immediate clinical outcome measure of the evidence (inaudible) no change and all cause of mortality was intensive glycemic control.

And it did show a significant reductions in the risk of microvascular complications, retinopathy and MI. And I just read that specific comment because I know this is – Bill is concerned based on the accord study.

Bill Taylor: Yes.

Female: I think, I don't mean to …

Male: Yes.

Female: So, OK. So, is the measure focus closely related to the outcomes? Our workgroup says yes and evidence of the measure is strong. It's kind of the

summary, the connection between A1c less than eight and increase – I'm sorry greater than eight, an increase rate of microvascular complications is well-documented. I think that the final – oh this one, I'm sorry, was the QQC addressed in this, developers? You can jump in but it looks like it was the quality, quantity, and consistency was addressed in this measure, the less than eight.

Developers are you still there?

(Crosstalk)

Female: I know. In a (rocky), right? But we can slow down or we can just kind to go to the next in interest of time, assuming that the workgroup got it right and then it is concluded there is strong evidence for this in terms of the outcome.

Bill, do you have concerns here?

Bill Taylor: Yes. One issue that quality, quantity consistency. If you have a very small improvement in something like renal function or the number of people who would go on and develop blindness or whatever, which is what they found in BCCT. But the key is much less than 0.1 and then you see it somewhere else like in UKPDS, is that strong evidence, you know, quality, quantity – I mean the quantity doesn't mean the amount of benefit you achieve when you do this, right? You just need for there a lot of studies that show us that it helps to have a lower in A1c. I mean what if you had to treat 1,000 people there or 10,000 people to …

Female: I think the high it was given based on the fact that it comes from the clinical practice guidelines which pulls an evidence from the systematic review and it can – it can – it can receive a high rating because quality, quantity and consistency is addressed. It doesn't necessarily mean that quality, consistency and quantity are all strong but it's a high level of evidence. I think that's where the rating comes from that maybe right or wrong but – from our group comments.

Bill Taylor: And slow me down or tell me to stop but this problem of – there is a benefit which is you have less retinopathy and you have less (nephropathy), but is this

a point in which we worry about currently what we might be doing by and as you as if you (inaudible) to some people are going to less than seven and some people going to be hospitalized for hypoglycemia and …

Female: I think is this a place where we discuss it this there in validity because I mean, you know, are we not to saying it is a valid measure of – of better care because we're – it maybe not necessarily, the problem with the accord study is the patient who were tightly controlled were having hypoglycemic episodes that weren't being caught because of method of, you know, monitoring the patients is through the HbA1c and not to say, glucose monitoring, right?

Anna McCollister-Slipp: Yes. Where they saw that – where the saw the increase in cardiovascular (vents) was on the group that was – that was shooting for ballistics 0.5, right? (Inaudible) relatively big difference between 6.5 and 8. So, I think we're just again, this is a highly imperfect process and I am not a huge fan of A1C at call. But you've got to start drawing a line somewhere in terms of starting to define what is the (bounds) quality that we're at least shooting for. So, I mean …

Female: Is that Anna? Thank you.

Anna McCollister-Slipp: Yes. This is Anna.

Female: For the summary.

Anna McCollister-Slipp: Again this is – this is opinion, it's not, you know I'm not …

Female: (inaudible) about the 6.5 but I think Bill's comment wants further discussion.

Anna McCollister-Slipp: Right, yes.

Female: In this group meeting because it is – it is evidence that out there, there a study that was gone and …

Anna McCollister-Slipp: And I'm usually the one that raising those issue. So thank you, Bill. I'm just putting double that to (inaudible).

Bill Taylor: Thank you. Well talk to the group.

Female: And I mean, the question maybe is the same question that (Darryl) asked the first is, is that eight an arbitrary number? It's hat cut point specifically what the evidence is saying that eight is good yes, no I mean that, you know. So that maybe the question were asking just like you said last time for the poor control. But in general we've rated it as a high level of evidence to support tight control or, you know, good control of less than eight.

Bill Taylor: I think this is the last time. I don't want to slow us down anymore …

Female: OK, so …

Bill Taylor: What – what about period (density) what about, you know, you check it once a year, you check it twice a year, you check it everyday or every 10 years, that's part of …

Female: I think that's part of your validity. You say a literature?

Bill Taylor: I just think we don't know much about it, you know, how often to do it but we are going to recommend it'd be done with a certain (inaudible). And I do – we don't have chance to say that I would think that.

Female: Right I think it's important because the measure specifically during the measurement year, you know, so that's what were asking to them report. And people to report and so that's – that's important if you think that how often is an issue and then something …

(Crosstalk)

Female: So you're saying that maybe it – it shouldn't be just one measure per year or it should be patient who have a measure at least three a one sees per year?

Female: One is the adding supporting the once per year I mean that – that part of evidence of whether or not isolated to that somewhere looking at.

Female: Correct.

Female: I think it's (inaudible) to put out there. So good question about evidence. I think some of that comes up in the validity also where measuring what we say we're measuring.

Female: Right.

Female: Yes, you can. You guys are definitely asking the right question so under evidence you need to think about how the measure was specified and it's a good evidence saying that how it should be done. And that's exactly what you're evaluating when you do the evidence. If it gets a little trickier because you're right, we also look at evidence under validity.

So even though we split them out into these things, there is a little bit of overlap. So, you're going to be thinking about evidence throughout the process.

Ingrid Duva: All right so, in the interest of the time I'll move – since we got exact that captured, I'll move on the performance gap and there is large gap in performance by the developer and let me see if anybody has anything else to add. High priority obviously the diabetes is high impact, yes, this are important worry for the implications for morbidity and mortality cause of care.

The next is reliability testing. This data – let me look real quick but I believe it looks the same as for the poor control. There was high reliability, particularly for the health plans. It's a little bit of reliability difficulties for the physicians was all over the place but we've already discuss that.

Exclusions, lets see. So, we move on to specifications. Are the – are all the data elements clearly defined? I think that specifications kind of gets at the discussion we just had a little bit about, is it once a year, you know (inaudible) – I mean, the measure itself is clear that it's once a year but is that the best measure.

Female: Yes and I guess one of my questions for the NQF people is and I kind asked this o the last introductory call that we had is, is there any – do we have to just vote up or down because we say yes, this is the right number. But instead of

once a year maybe it's only for patients who have anyone takes three times a year because it's – you've got one patient who only comes in once a year, than those probably – that's going to the (inaudible) selection of somebody who may not be as interesting in or able for one reason or another to be able comply with or adhere to their regimens.

Whereas, if they come in three times a year at least, then there's at least a greater degree of either capability or commitment to taking care of their disease. And, you know, I'm not sure it's fair to judge our physician on quality if the patient is only coming in once a year for whatever reason, whether or not those are issues that the patient could control or not.

Female: You could probably argue that

Female: (Inaudible).

Female: And you could probably argue that both ways, of course with the access issues, I mean, from, you know …

Female: Right.

Female: So …

Female: You maybe very good physician (inaudible) a very good quality of care in a very poor neighborhood where people can't buy their medications on a regular basis. So, is that fair? I mean, I don't know, maybe. And again, it's binary nature of the up and down kind of bothers me.

I think it would be – I mean that maybe this is just not the right forum for this, but maybe – I think it would be great if we could say, "Yes, is the right code, but let's changes the methodology a little bit," and you know, maybe that's not the way that's work but it would seems to be a way of getting the advice of the committee members.

Karen Johnson: So, this is Karen. And to answer to that question, in the in-person meeting, we will me as evaluating the measure as specified. So, you pretty much are going on what – how the developers have decided to construct the measure,

you can certainly offer opinions and the developers, they hear you now. They'll hear you on – in the in-person meeting and they can take that under advisement. So, that's kind of how that works.

Ingrid Duva: And so, in terms of what is specified in this measure, it is clearly defined as once a year. So, it rates high in specificity but one committee members states why does the numerator included missing (past). So, that would be a question I have for everybody? So, if you look back the numerator – let's see.

Where did they take that – I can't see it, I'm not playing it up right now but ..

Female: It's up on the …

Female: The numerator was people who've had like one HbA1c within the past year, and had a diagnosis of diabetes into that area the previous year or something like that, wasn't it?

Female: Right.

Female: There it is.

Ingrid Duva: Yes. And so, in the denominator is eight – patient 18 to 75 years of age, by the end of the measurement years they had the diagnosis. So, that was the – and then the numerator is – was recent was eight percent and I don't see them missing. (Inaudible). I know that is in one of the measures. So, I don't see it here immediately. So, I have to look at that more closely.

Karen Johnson: This is Karen. You are right it is in the numerator details. So …

Female: Yes.

Karen Johnson: Yes. And but those are still (Inaudible). I imagine that they are thinking that tips of provider, isn't able to get their folks in for the test and they probably need a (ding) on, on the quality. But I'm sure they can explain their thinking.

Male: Karen, (inaudible). Thank you.

Ingrid Duva: OK, got it. So, reliability testing, basically sufficient reliability was (inaudible). One question could be attributed to patient factors versus medication control. I'm not sure what that means, maybe the intensity of the care. Differences in performance by individual physicians to provide is less reliably distinguished. So that's attributed to the noise for the physicians.

OK, so validity testing, basically receives – mostly yes, it is valid, that there's correlation – because, somewhat, the developer's presented there were correlations between this measure and other diabetes measures from them. There is an – all five of these are from the same developer.

So high (inaudible) and (space) validity. High correlation in case validities, excuse me.

Then there's a question from NQF staff about exclusion for the appropriate. For instance, the (youth) and doesn't really look like the workgroup addressed that.

Anna McCollister-Slipp: And the exclusion in this case were what the (inaudible)

(Crosstalk)

Female: Right.

Female: Sorry. Say that again Anna, please?

Anna McCollister-Slipp: Oh, it's just sort of reiterating the conclusion. I mean, you know, I don't completely understand why it doesn't include children or (inaudible) but I know that there are different concerns for each of those populations and maybe is appropriate.

Female: Yes. I know there different (inaudible) but I can't – and maybe it has to do with the evidence we went back to look at what the evidence is. So do you have any input on that, Vicky?

Bill Taylor: My guess is that those are the people who are included in those studies that we're using to justify doing this but I want to go back and let them move before I take anymore of your time. (Speculation).

Female: Vicky. Did you have any comment about that?

Vicky Ducworth: No. I kind of (inaudible) to go back and what (inaudible). Thank you.

Ingrid Duva: And then small threat to validity, what is the variation in measure reporting where the data comes from? EHRs claims, lab data, electronic form, newspaper.

No missing data, feasibility, high. These data is already being collected, routinely generated through care delivery and most of EHRs and claims data make collection analysis of A1c relatively straightforward.

Usability, there's five current uses that were described by the developer including public reporting which is the higher rating and usability. Questions that the committee were posed – let's see if they're addressed. There are the patient factors that are as a concern again. Yes. Generally, it's yes. It's probably the most useful measure for high quality, efficient healthcare. Well-established with basically the comments from the group in terms of the evidence we have. Yes, it's easy to use.

And that's it.

Female: You're doing a great job, Ingrid.

Ingrid Duva: And now – I'm number two. I think Vicky has the next one.

Karen Johnson: So, this is Karen. And we're just looking in our time. And unfortunately, we're not going to have time to go through each of the other measures and the detail that we've done for these.

So I think as we talk about the next three, if there's anything (burning) that you particularly found problematic or something that you particularly would like to ask the developer, you know, since we have them on the call and you can get your questions answered, why don't we do it that way and see how far we can get? We have about a little over 15 minutes on the call.

So will that work for you to go on to number 57 and just kind of do a high level (inaudible)?

Anna McCollister-Slipp: Sure. And I think I'm (inaudible), although, Vicky if you want to jump in since you've prepared for it, that's fine by me.

Vicky Ducworth: Sure.

Anna McCollister-Slipp: I just need to find that on my (screen). So in general – this is Anna. In general, I mean my perspective on this is that, you know, there's lots of evidence to support A1c as a process measure. You know, my personal opinion is that – it has many issues associated with it.

It's a very, you know, there are lots of problems that I have especially with A1c in terms of its inability to capture glucose excursion and (hypos) and that kind of stuff. It's just sort of an average over a period of time rather than some of the things that have been linked to complications. But, you know, it's probably the most validated measure that we have and is certainly is useful to collect. But in terms of quality, I see it lessens the measure of high quality than the (inaudible) of it was being a measure of low quality, if that makes any sense.

And I'm looking to see committee comments. Let's see, sorry. The evidence seems to be very strong according to the committee. I mean, first (looking) the high priority, and it's (addressing) significant health problems, high impact, you know.

There seems to be agreement there. Let's see, so reliability, it's a relatively straightforward measure to collect, it's very specific. The specifications seem to be consistent with the evidence.

Again, I think of – I mean, just being conscious of time I think this is a relatively straightforward process measure, you know.

Bill Taylor: May I?

Anna McCollister-Slipp: Please.

Bill Taylor: Only one thing which is – it says unintended consequence is none. Then there would seem to be somewhere we need to think about when, you know, when recommend – getting A1c testing has already confirm that occurs as a consequence of it.

Somewhere, we ought to bring that up maybe in the, you know, in-person meeting.

Female: Sure. But we'd be talking about the process of doing it. So it's doing that causing the process of getting (inaudible)

Anna McCollister-Slipp: Infection or …

Female: Infection or ignoring other important things for your patient. But this is – I just throwing it out there, I don't think this is where the hypoglycemia comes in.

Bill Taylor: The reason we think it's a good thing to do is because of the benefit that would come of measuring and hope that we're going to prevent nephropathy and blindness and myocardial infarction and so on. But do we do that without then asking what was – might we do with hypoglycemia or (inaudible) to this and failing to talk to people about their smoking or – I don't – just tell me where does that fit in? Can we only talk about the benefit and we can't talk about the harm?

Female: No, no. No, I just was saying related to the process itself.

Bill Taylor: Yes.

Female: Not necessarily to the outcome.

Bill Taylor: But the – if we limit it to, you know, if you get a hematoma or you get infection because the needle wasn't clean, because that's the process. But …

Female: Oh, I think you can go beyond that because you can go beyond, OK, we're getting the information now. Are we doing anything with it?

Bill Taylor: Yes.

Female: And that is kind of the harm that presents itself because you're taking the (stab) of getting out information, especially if it's changing the care you give your patients. Anyway, I just threw it out there, I didn't mean to slow us down.

Bill Taylor: OK.

Anna McCollister-Slipp: All right. I mean, I think you (inaudible) the questions are the right ones to ask in some form. I guess, from the NQF, I want to if that's what you want from us because I can give you lots of stuff and it sounds like Bill can as well.

But in terms of looking at the harm caused by the actual processes, getting a hemoglobin A1c, I think that it's pretty minimal in most cases.

Karen Johnson: And this is Karen from NQF. And those are definitely the kind of things that you have to think about. But also, I haven't been able to see out who's talking in. I don't know your voices well enough.

But also, when you think about measures and you think about the infrastructure and the resources that are used to collect data for each kind of things, it is fair to ask, is that a good use of resources or it's the opportunity cost? So that also is a fair thing to consider.

But what we, you know, theoretically, there could be all kinds of horrible things go wrong with any measure. So we want to limit the discussion to things that there's pretty good evidence that there is something going on.

So we don't want to just try to theorize about all the possibilities but really, limit that conversation to these (contenders) if you will, if there are any. And also, we do that under the usability and use criterion.

Ingrid Duva: OK. So this is Ingrid again. Did you guys – did you – does the (comprehensive) have questions for the developers?

Bill Taylor: I guess my only question is this. Is there any way to get at some kind of measure with adverse effects, unintended consequences? I don't know where

you would find it because the studies are all done, you know, telling you efficacy, knowing the fact that in study conditions, we know what happens. If what happens in real life, you have to capture another way and I don't know where that's recorded.

But if it's out there, it would be great to know about it.

Female: I completely agree. And I don't have to develop (inaudible) comments on the data sources that they looked at or decided not to include. But I mean, this is a very well-established measure in the diabetes world. We can talk about frustrations or, you know, lack of the diabetes community in developing better measures perhaps.

But in terms of the current thinking with the (new) diabetes communities, that's a very well-established measure and the process measure I think, you know, fits appropriate as a very minimal baseline, and that not doing it is probably a sign of poor quality of care.

I think it's adequate, I mean, to come up with measures but that's a different question.

Ingrid Duva: So I'm sorry to keep interrupting. This is Ingrid. I do have specific question of the developers on the 0055? So that's why I was trying to get to the next measure if you guys didn't have something specific for the developers.

Female: Well, I think we can go ahead and move?

Female: Go ahead.

Ingrid Duva: Are you sure? I didn't mean to cut you off. I just …

Female: No, no, no, that's OK.

Ingrid Duva: I have five more minutes. So I was asked to present the comprehensive diabetes care, the eye exam. And kind of the last minute but I did have a chance to go through the work group comments and I'm not going to go through them one by one. They are very similar to the ones we've already gone over in terms of the evidence, the performance gap priority.

But I did have one question because it was a question I had about because of the specifications. And then I noticed another workgroup member asked about – it was an appropriate measure – so my question was, if we're measuring the eye exam, the implementation of the eye exam for the patients with diabetes is – and then they have to be referred to an ophthalmologist. Is this more of a question of care coordination?

And then one of the other group members asked, is this is a fair measure for most practices who don't have an ophthalmologist with them? So then I thought there must be some – we must not be clear what exactly we're expecting out of the measure.

So I just wanted to ask that back with the developers.

Female: All right. I have the exact same questions. I'm just not sure it's – I don't know how feasible this is.

Ingrid Duva: My impression of it is that the primary care provider makes sure that the screening is done but it's really a referral to the ophthalmologist. That was my impression. But – and then I think maybe it was you and I have the impression that the ophthalmologist had to be part of the practice.

So it just seems like we're a little bit unclear.

(Mary Barton): So, this is (Mary Barton). So if I could say quickly, the – you know, those measure specified for health plans as for all physicians. And the idea I think is that the responsible party is responsible for seeing that it gets done, right?

And so while true that someone who's trained in internal medicine or family practice is not expected to also be an ophthalmologist, they're expected to encourage the patient to go to the referrals to perhaps facilitate them making the appointment, and then out in the backend to check that it got done. And I think that that's what the implication of the measure is that says, you know, the entity that's responsible for taking the care of diabetics is responsible for seeing that these things happen.

Ingrid Duva: OK. I just wanted to – I mean, that was impression, this is Ingrid again. When I read it, that it was kind of the coordination of care, making sure that the referral was made and that the patient had, you know, the appropriate access.

So I mean, I just think that something we discussed when we talked about – when you were evaluating these components of the measure, keeping that in mind that that's what it's measuring.

(Mary Barton): Yes. I mean yes. And that's – I mean, from my perspective, I'm not sure that that's a realistic expectation to have for a physician based on personal experience and how the process works of somebody who does have retinal disease.

So I mean – there's only so much control you have over whether or not somebody goes to see an ophthalmologist which is probably one of the most, you know, eye appointments or appointments you could possibly imagine in terms of trying to work in those schedule and – I mean, we could go on for a while about that and we don't have time. But, I mean, if we're measuring ophthalmologists, fine. I'm just not sure that it makes sense to measure my endocrinologist based on my ability to comply with those or not.

Ingrid Duva: But then, when we get together, we'll discuss from maybe the health plan perspective?

(Mary Barton): Right.

Ingrid Duva: So that – thank you, (Mary). That was my question. I'm glad that you kind of recapped that for us. And then I think we only have one more measure left, all right – since I've bought it in, I just want to make sure you know I'm (vetting) back out.

Bill Taylor: And this Bill. (Inaudible) is that the nephropathy that I've noticed (inaudible), I do have a couple of questions to highlight and maybe ask the developer.

Female: Perfect.

Bill Taylor: Great. So, you screen for nephropathy obviously, so you can find people who are, you know, have early nephropathy and if you treat them with an ACE inhibitor in our view can, you know, decrease the rate at which their adrenal functions deteriorates. And that's well-shown.

But the part that is not well-shown is how we – what's the optimal way to identify these people. If we know that we can identify them with, you know, like albumin test in their urine. But the thing we don't know is (inaudible) again. So you're checking the – it says within the current year, has this test been done? But I'd love to know some evidence if there were some on how often it makes sense to do this test.

I don't think there's much out there that helps us with that. So it may just be that that's unknown question. But we ought to at least highlight it. My understanding is the cycle at which these evaluations take place is yearly because that's how long it takes the earth to get around the sun. And so we looked at that rate. Maybe it makes sense to look twice in a year or every five years, or something else, you know, to at least, I think (inaudible). Is that part of the recommendation or something that …

Female: Was this the one that has some contradictory evidence related to how many years – that might have been the (inaudible) exam, I might be forgetting …

Bill Taylor: Oh, I think this one had no evidence.

Female: No evidence, OK.

Bill Taylor: No.

Female: For the timing, OK.

Bill Taylor: Yes.

Female: So let's go look back at that.

Bill Taylor: But I think …

Female: Yes it is. I think it's an interesting point and I mean, if somebody who has kidney disease as well, you know, there are periodic fluctuations and, you know, protein, you know, albuminuria or micro-albuminuria from one part of the year to the next or two, you know, three months physician to the next three months physician visit.

So that's an interesting point. I guess part of my question was why are we just doing urine analysis as opposed to urine analysis plus serum creatinine. And I don't know if that's just an issue of, you know, ease of collection or cost. And I would be interested in getting that information from the developer.

Female: Does anyone from NCQA on the line to provide some comments?

Male: It sounded like the question was is there a reason we did not include serum as opposed to nephrology concept for (inaudible).

Female: (Inaudible) was the urine test and then the question of whether creatinine was also included. And I'm just trying to – are the specs there on the webinar now for this measure?

Female: So the ADA guidelines suggest both urine and serum creatinine. And yes. I mean, I was just wondering if there was a logic behind that or was it just – I would like to know what the logic is to that. I mean, maybe it's not relevant but …

Female: We'll come back to you just a minute on that, OK?

Female: Sure.

Female: Good question. Were there any other burning questions on (0062) that while NCQA folks that are on the line that you could ask?

Bill Taylor: This is Bill. Those are the big ones from my point of view.

Female: OK. Great.

Female: All right, great. Now, if we could take a brief opportunity to open all the lines for a public member comments?

Operator: Thank you. At this time, all lines are open.

Female: Are there any comments from the listening?

OK. Well, we thank everyone for taking the time to participate in today's workgroup call. Over the next couple of weeks, we'll continue to hold the remaining workgroup calls. And everyone should have received their travel arrangements and registration link. Please let me know if you have not received that. And our meetings department will follow up with you.

We also really want to encourage you to reach out to us, give us a call, or send us an email as you're reviewing the measures. If you have any questions about the criteria, we can help walk you through any of your questions.

We are more than happy to work with you.

Female: And also remember, we have three more workgroup calls coming up even though you weren't assigned to this workgroup, if you would like to listen and that might be a useful exercise if you have the time and availability to do that just to get more practice and see how other folks walk through that algorithm and that sort of thing.

So feel free to that if you would like.

Female: And our information can be found on the website and on the SharePoint site (inaudible).

Female: OK. Thank you to everyone from NCQA for joining us as well. And we will look forward to seeing you all at the end of February.

Female: Thanks very much.

Male: Thanks.

Female: Thank you.

Male: Thanks to all of you. Thank you.

Female: Bye.

Female: Thanks, bye.

Female: Bye.

Operator: Ladies and gentlemen, this concludes today's conference call. You may now disconnect.

END