NATIONAL QUALITY FORUM

Moderator: Sheila Crawford
April 24, 2014
2:00 p.m. ET

Operator: Welcome everyone. The webcast is about to begin. Please note, today's call is being recorded. Please standby.

Wunmi Isijola: Thank you. Good afternoon everyone and welcome to the Second Surgery Steering Committee measure Q&A call. I just wanted to get started and introduce our team here. My name is Wunmi Isijola the project manager here. We have Amaru Sanchez as our project panelist. We also have two of our senior directors, Melinda Murphy and Karen Johnson.

But before we get into the evaluation process, I just wanted to get a sense of who's on the call. So could you just shout your name just to let me know who of our committee members are on the call with us today?

(Barry Marksman): (Barry Marksman).

(Keith Olson): (Keith Olson).

Wunmi Isijola: Hi, (Keith). Thank you. I heard some other names.

(Barry Marksman): (Barry Marksman).

Wunmi Isijola: Hi, (Barry). Thanks for joining.

(Barry Marksman): Thank you.

(Bob Salem): (Bob Salem) from Seattle.

Collette Pitzen: Collette Pitzen.

Wunmi Isijola: Thank you, (Bob). And I heard Collette?

Collette Pitzen: I'm Collette Pitzen from Minnesota Community Measurement.

Wunmi Isijola: Hi Collette. Thank you for joining us. Is there anyone else?

OK. So with that being said, we're just going to jump right in and I'll turn it over to Karen Johnson. And as mentioned in the previous e-mails, these are informal discussions. Karen Johnson will just kind of give an overview of some of the criteria. And if you do have any questions at any point in time, please do jump right in. So Karen?

Karen Johnson: Thanks, Wunmi. As Wunmi mentioned, this is going to be very informal. So I don't want this to be a monologue by any means. My advise to you is make sure that we answer any questions that you might have going into this project. Just one housekeeping detail, as we start learning your voices, if you would mind just identifying yourself when you ask a question or say something, that would be appreciated.

So let me start off the call just by asking if you had any general questions at all about the project SharePoint site for about any of the materials that we made available in the SharePoint site.

(Inaudible). Has everyone (inaudible) at least get to this site and (inaudible).

Allan Siperstein: Hello.

Female: Hello.

Allan Siperstein: Hi. Allan Siperstein in Cleveland.

Wunmi Isijola: Hi, Allan. Thanks for joining.

Male: Hi. One little trick about the SharePoint site that if you click a tiny little triangle just to the right of the measure name, it will allow you to basically

download in one click, the entire folder onto your computer as supposed to having to download each of the several components.

Karen Johnson: So you had that (inaudible)?

Amaru Sanchez: No.

Karen Johnson: Yes, that's a great shortcut. Amaru is bringing that up now. Yes. So can you go down to the measure documents? OK. So if you click right there …

Male: Yes.

Karen Johnson: OK, so you can download everything in that set. OK, great. Thanks for the shortcut. I'm still learning SharePoint.

Frederick Grover: This is Fred Grover. I just got on the call.

Wunmi Isijola: Hi Fred. Thanks for joining.

(AJ Yates): Yes. And this is (AJ Yates). I'm on the call too.

Wunmi Isijola: Thank you, (AJ).

(John Handy): (John Handy) also. We're going to do a roll call.

Wunmi Isijola: Thanks, (John).

(Tony Asher): (Tony Asher) from Charlotte.

Wunmi Isijola: Thank you, (Tony).

Karen Johnson: OK, thank you. And now you should already see there on the SharePoint page, you have your surgery workgroup and measure assignments document there. So hopefully, you have had a chance to look at that and see which measures you've been assigned for workgroups. And those calls are coming up very soon.

So, I just wanted to see if you had any questions at all about your workgroup assignments, about what they're expected to do for those, these preliminary evaluation, any of that stuff.

Male: Let me ask you this, we're supposed to be on all the calls whether or not our workgroup is assigned to present that day, is that correct?

Karen Johnson: Not necessarily. We ask that you be on the call that you had been assigned to and then if you would like to dial in and listen to the other workgroup calls, that would be great because that will give you closure to these other measures and the discussion but that's not required.

Male: OK. And then say, for example I've had a lot of activity with the SDS over my career in ACC, (inaudible) measures are those are discussed, is that a call that I should recluse myself from or am I welcome to at least listen in or what's the policy you have on that?

Karen Johnson: That's a great question and I think you've built out a disclosure of interest form specifically about the various measures. So we will have that and we will get to you before the workgroup call to let you know for sure which things we would like you to recuse yourself from. You will be surprised, I'm sure.

On the workgroup calls, you're certainly welcome to listen in on that call if you, you know, if they're discussing a measure that you're recuse from, you know, certainly listen in and certainly participate on those other measures that, you know, you might not be recuse from, so we don't want you to absent yourself from the entire workgroup call. And in the in-person meeting, we will probably ask you to leave the table when that particular measure is being discussed.

Male: Yes.

Karen Johnson: So you would (inaudible).

Male: So that's what I would expect yes, but for the workgroup calls, we can listen in and …

Wunmi Isijola: Most definitely.

Karen Johnson: Yes. We would ask you not to participate in the discussion of course but certainly listen in.

Male: Yes, OK.

Karen Johnson: Any other questions about the workgroup calls or the preliminary evaluations? Anything there?

Collette Pitzen: Hi. This is Collette. I just have a question. Is it true that not all the measures are loaded yet?

Wunmi Isijola: That is correct.

Collette Pitzen: OK, thanks.

Wunmi Isijola: That will be loaded before the end of the week. So please check back by tomorrow.

Collette Pitzen: Great. Thank you.

Wunmi Isijola: No problem.

Karen Johnson: OK, and before we go on, do you have any other questions about the process itself? The upcoming meetings that we're going to have, anything that – as you were going through and looking at your materials that you didn't quite understand or you couldn't find? Again, this is a very general informal Q&A session here so anything you want to ask?

OK, sounds like nobody has any questions. I don't know if that's good or bad, hopefully, that's good. Hopefully, everything is very understandable. Just out of curiosity, how many of you guys have served on NQF steering committees in the past?

Frederick Grover: Hi. I'm Fred Grover.

Karen Johnson: OK.

Allan Siperstein: This is Allan Siperstein. I was on the surgery measures committee last cycle.

Karen Johnson: OK. OK.

Colette Pitzen: This is Colette Pitzen. I have not but I've brought measures forward for endorsing.

Karen Johnson: OK. OK. That should be interesting for you to sit in the table and be on the other side.

Collette Pitzen: I'm excited. Thanks.

Karen Johnson: So Allan and Fred, depending on when that surgery project happened, I'm thinking it was probably back in 2010. Some of our processes may have changed somewhat and actually some of their criteria may have changed somewhat. So that background is going to be very (inaudible) but just be aware that, you know, some little things may have changed.

Male: Sure.

Karen Johnson: And so if you have any questions about, you know, how things used to be and, you know, why are they the way they are now, you know, feel free to ask any of those things and we'll try to explain things.

Male: OK.

Karen Johnson: All right. Well, again, this was simply a very informal call. So we wanted to make sure we took your question first and we thought we'd spend the rest of our time just going through one of the measures and very briefly walking through the algorithms that you might want to use and pointing out a couple of things to you.

This might be similar, I'm not sure exactly what you did on your orientation call, it might be sort of similar to some of the stuff on the orientation call but actually looking at a real measure.

So with that, why don't we go ahead, we picked a measure, an outcome measure this time. We picked 533 and Amaru has brought that up on the screen for you and I don't expect you guys to have looked at this before and that's fine. I just wanted to go through some basic points on it.

So first of all, just in case you haven't really (inaudible) your legs around or your mind around what in these forms – really, these forms here that we bring up for you look at the main forms are – there's really three pieces to it. So there's actually – there's actually four pieces. And the first part is the brief measure information. We have that in front of you, that comes directly from the information that we develop were just submitted.

And then if you go further down in the form, you get to the headers that are in that orange color there. And those sections of the form where the developers filled out are online forms. So certain parts of the important criteria and the feasibility and use and usability sections are all part of this form.

And then after that is what we call our evidence attachment. So that's the questions that we ask developers to use for describing the evidence and that's further down in the form. And basically, we did that as a Word document because it's easier for developers to give this information in a Word document instead of an online form. So we made that available for the evidence portion as well as for the measure testing portion. And the measure testing attachment is at the very end.

So it may take you a little while to use – navigating your way back and forth in that form. But, wow, this one has a lot of stuff in it. I'm really just going to take you to the testing form. So we have those hyperlinks in there so if you're looking at it online, you can click the form and then if you want to get back to where you were, you need to use the alt and left arrow key and that will take you back to where you were. So remember, alt left arrow and that will help you navigate through the form.

OK. Let me stop there, see if anybody has any questions about the forms in general.

Frederick Grover: I'm having trouble getting on your webinar – different link than it was the last time?

Karen Johnson: Yes, it is.

Wunmi Isijola: Yes, it is.

Frederick Grover: Could you pop it to me or it's …

Karen Johnson: Yes, Fred.

Male: I'm connected to the webinar right now but I'm just seeing a static stream. If you were going to a form, all I'm seeing is just the surgery workgroup general document displayed but really nothing else to this point.

Male: Yes. Go ahead.

Karen Johnson: OK. Amaru is looking at that right now trying to reload something.

Male: Hopefully …

Karen Johnson: Are you seeing something that looks like a form now?

Male: No. I'm just seeing a blank page with a bunch of multicolored squares.

Wunmi Isijola: Well, if you can go, can you actually go into SharePoint at this point in time just because we're actually pulling documents from the SharePoint site?

Male: OK. Let's try that.

Karen Johnson: Yes.

Wunmi Isijola: We are in the measures 0533.

Male: Yes.

Karen Johnson: Yes. It's in the link to Fred. OK. All right. So Fred, I think that was you that said you couldn't – you didn't have the right link for the webinar and Wunmi is going to send that to you.

OK. So this measure, let's go ahead and just delve into this measure here. It is measure 533, post-operative respiratory failure rate and it is a measure supported by (inaudible) and the brief description of the measure is that the post-operative respiratory failure, mechanical ventilation (inaudible) per 1000 elected surgical discharges for patient 18 years and older, that's what the measure is about.

If you go further down that brief measure information form to the middle there, right there at the bottom of your screen, it tells you that this is an outcome measure and the data source is administrative claims in the level of accountability for this measure or the level of analysis at facility level.

So we won't go into the details that are there in the brief information just realize that the stuff that is – the brief information section of your document is repeated in the rest of the document somewhere (inaudible) really some really important points to have it (inaudible) so that you could easily find some of the key points to the measure.

So, when we're going through our evaluation, we will start talking about evidence, so that's the first step that you're in under importance to measure and report that we'd look at. And because this is an outcome measure, we treat it differently in terms of rating it than we do other types of measures. So Amaru, if you would bring up our evidence algorithm, just a reminder, we have an algorithm to help you rate the measures. So we have an evidence algorithm and Amaru is bringing that up now.

You can find this algorithm in a couple of different places. It is located in the steering committee guidebook and if you happen to have that open, that would be on page 38 of the guide book. It's also in a document that contains the criteria and guidance. So other documents that there in – listed document at the top of your SharePoint page. So there's a couple of different ways you can find this algorithm.

When you're looking at the evidence algorithm, the first thing that it says, there is the green box there, does the measure assess performance on health

outcomes (inaudible) mortality function (inaudible) such as quality of life, symptoms, (inaudible)?

So, the first question is what kind of measurement? So this particular measure isn't outcome measures, actually looking at complications. So the answer to that is – that question is yes, that takes us to box two. And then to evaluate the evidence for outcome measures, it's actually pretty simple. We want to know if there is a rationale for the measure itself and specifically that there is some sort of a relationship between (inaudible) and at least one healthcare action is stated or diagramed or something like that and it could be any kind of action (inaudible) process and intervention, something like that. So, it should be pretty easy to evaluate evidence for outcome measures.

And if you look actually at page – let's flip back to our forms from (RF), if you go to page 24 of the form, so again, that's in the evidence attachment, let's see how these going down – go down to page 24 and you will find under 18.2, go back up just a little bit, you will see the rationale and you will also see under 182.1 the various intervention that the developer had suggested can be done to reduce these complications.

What's a little bit interesting, so this is really all you need to look at in this part of the form to evaluate evidence for the outcome measure. This developer went a little further, and they gave you a little bit of information that you don't really have to have but it's interesting stuff if you're interested in looking at it.

So if you go a little bit further down, Amaru, go to section (1.8) kind of keep going down, what they did – actually, go back up a little bit. Keep going up. What they did is they actually had provided some information from the systematic view about specific interventions for these complications. So they described the review, they give you the grade with the evidence for the various intervention and then you can scroll down.

This review I think was done in 2006. They went even further – keep going down please. And they – in section 1.8, they actually gave you some additional articles since 2006 talking more about additional intervention.

So again, these are not things that you have to look at if you don't want to. And they are not things that you would use to help evaluate evidence but they are interesting. And if you go down, a little bit more, let me point out, for those of you who are on the phone, this last paragraph here that starts in addition some studies (inaudible), if you take a look at that a little bit later, we won't go through it now, but if you look at that, that actually gives you some information about the validity of this measure.

So, what that shows you is that sometimes, really good information can be intersperse on the form in interesting places. So if you happen to see that, I'll bring that to you. OK, any questions about thinking about evidence for the outcome measures?

Male: None.

Karen Johnson: OK. Let's go back up to performance gap so performance gap is the second of the three criterions under report to measure report. And basically if you recall performance gaps can be demonstrated either by having – showing poor performance overall, large variation of performance or variation in performance among certain population (inaudible).

So, if you look at these measures – keep going down a little bit, Amaru, go to 1B.2 and we start to see here things that you expect to see, you see some numbers in this part of form and you actually see that they've given you a couple of different things here in table one, they have given you several of the performance scores over time and then also underneath there, the very bottom of your screen, it says performance score distribution 2011.

So for 2011, they actually gave me the distribution (inaudible). And you look at those and decide – you can look at the overall rate itself and see if you think that that's demonstrates poor performance, you can look at the distributional statistics if you think that that demonstrates variation in performance.

And you can also go further down to section 1B-4, and they actually give quite a few different statistics by various population subgroups that you see in

age and gender and income statistics, geographic statistics and it looks like payer information as well so a lot of different things there.

And in case you are wondering why it's kind of ugly, that's really NQF call the online submission form that they – that developers use to enter that information. It's plain text so it doesn't allow formatting. And that's why sometimes, (inaudible) but you will use when you're evaluating this criterion, we'll use our generic rating skill.

So Amaru, can you open up our generic rating scale? Go to the criteria and go to table three. I know you've seen this before, this is a tiny little table here. Yes, there you go, thank you. And basically, it's pretty generic. It's nowhere near as detailed as the evidence and reliability and validity algorithm. It's much more on high levels than that. But basically, we just want you to consider the information that's been submitted and see if you think you have a high or a moderate or low confidence about whether or not the criterion has been met.

So again, for performance (inaudible), we'd look at the numbers that have been provided to you and then decide if you are highly confident that there is a gap or an opportunity for an improvement or if you have low confidence that there is a gap or a moderate confidence, that's sort of thing.

And again, notice that the scale like our other – like the other algorithms pretty much need a high or a moderate rating in order to have (inaudible), the word high and moderate folded in that table. So let me stop there and see if anybody had questions about performance gap and rating on performance gap.

Male: I have a mechanical question. If you could just click back on the screen on the web page on the presentation and show me where you got that form?

Karen Johnson: Yes. So there is …

Male: Now, where did you click to get that?

Karen Johnson: OK. So if you opened up the measure evaluation criteria guidance, it's 2013 document.

Male: Right.

Karen Johnson: That's a 30-page bible, if you will, of all of their criteria and guidance. And that's going to come open. Yes, (see that in there) and he clicked on table three, if you'll show us that, right there, table three.

And can you go back and show us again, it went so fast, sorry. Not as easy though. Anyway, table three in there, it is also in the steering committee guidebook and it would be on page 41 of the steering committee guidebook.

Male: Now, are those the forms we're to use for the preliminary measure evaluation?

Wunmi Isijola: Actually no. If you scroll to the left hand side of the screen that we're looking at, it says surveys.

Male: OK, so it's surveys. OK.

Wunmi Isijola: That is correct. And based on your corresponding measures, you'll have to answer some questions based on that. But again, as Karen mentioned, you can use some of these materials as guidance when you're reviewing each measure.

Male: But don't leave that, go back to surveys because when I got in the surveys, it just says it doesn't give me a survey to fill out.

Wunmi Isijola: OK. So are you seeing what we're looking at?

Male: Yes – no, I've gone into there and when I click on that, it just said that it's due a certain date. Can you show me one of the surveys?

Wunmi Isijola: Yes. And then if you see at the very top, it says respond to this survey?

Male: That's when you get to do it. OK.

Wunmi Isijola: Exactly.

Male: So it's not a PDF, it's respond and then enter online?

Wunmi Isijola: Correct.

Male: OK, that's what I needed to know. Thank you.

Wunmi Isijola: No problem.

Karen Johnson: OK, any other questions about anything we've covered so far?

Collette Pitzen: Hi. This is Collette. I have a philosophical question.

Karen Johnson: OK.

Collette Pitzen: Historically, how have the workgroups or the steering committees dealt with measures that have a very low rate of occurrence in terms of variability and opportunity for improvements?

Karen Johnson: You know, sometimes it really kind of depends. I mean, this is like a low frequency thing kind of like these complications?

Collette Pitzen: That's correct.

Karen Johnson: Yes. (Inaudible) definitely see that and really, you're trying to getting a really more into the priority sub-criteria there.

Collette Pitzen: OK.

Karen Johnson: But – which we were going to talk about next. But …

Collette Pitzen: OK.

Karen Johnson: … surely there's lots of different ways that we can consider something a priority. And it doesn't necessarily have to be large numbers, it could be severity if high quality care is not given for example. So …

Collette Pitzen: Great. We can continue to have that discussion. Thank you.

Karen Johnson: OK, OK. So let's go to priority, go back to the measure information form and go to page five. So here's high priority. Again, this like performance gap, you'll use the generic rating skills and we'll get priority. And the form there

even tells you what we're looking for either it addresses the specific national goal or there is some way that they had demonstrated high end.

So when you look at what's provided by the developer here, they say, well if that's (inaudible) number is this leading cause the morbidity and mortality, and it's high resource use. And that's how they see this measure as being a high priority measure.

And then in 1C.3, they shouldn't be giving you some actual (epi) data or some other kind of data to back up their (claim). So I think you're right, if they say there's about (5,656) of these adverse events were estimated to have occurred in the US community hospitals in 2008.

So, you know, maybe this remains on there or may not have checked that large number as you have to decide if you think that's large enough. But then you go online and the next paragraph talks about there is extra 21 percent – mortality, lots of extra days in the hospital which of course translates to dollar.

So generally priority, it should be a fairly, easy sub-criterion for a measure to get back. So you still want to talk philosophy here or are is that kind of giving you stuff to work with?

Collette Pitzen: That's OK. Thank you.

Karen Johnson: OK. OK, now let's get into the meat of this measure. So I've kind of covered some of the first – I think Reva covered in the first Q&A. Let's talk about the reliability and validity and how you would think about that for an outcome measure subject.

So when you get ready to evaluate the scientific accessibility of these measures, you'll be looking at reliability first and then at validity. So reliability, there's two pieces you have to think of. So one is the specifications themselves, and then the other is testing to demonstrate reliability as a measure.

So in terms of specification, can you go to the page six please that show you where specifications are. And the specs are exactly what you would expect.

It's a lot of information about people (inaudible) a little bit. And so down a little bit more (inaudible) 13 numerator statement and then if you go down a little bit more, numerous detail.

So notice that in the numerator statement, they describe the basic ICD-9 post for different things, and then in the details, they actually give you the ICD-9 codes that are used in this measure. And I use the same thing for denominators, for any inclusions if there are any, that sort of thing.

So when you're thinking about the specification sort of measure, you really are thinking about under – at least, under reliability, you're thinking about how precise are your specifications. So the idea there is if two different groups try to implement the measure, our specs precise enough so that both (ridge) can get basically the same kind of (inaudible) or outcome.

Then you go on to think about reliability testing. So when you're thinking about reliability testing and the results that the developer has given you, they – you need to think about really three different things. You think about the method that they use to do the testing, you think about the sample that they use to do the testing and then of course, you have to consider the actual results of the testing so all of those things together work to give you a flavor about how reliable to do our measurement.

So you will see under the testing portion, that actually is in the testing attachment so Amaru, (inaudible) here. Another thing to point out is what we – we call this entire thing, this document and Amaru is going to – a lot of times, what we refer to this is the submission form, you know, it's basically all the stuff that developers have given us. You know, often they will also provide other pieces of information, and that's very true. You know, these guys actually gave you a lot of ICD codes in the actual form but I think they also provided in codes and different things in a cell spreadsheet and some other documentation.

So what we hope that they have done is provide what you really absolutely have to have is hopefully in this form that we have. And then everything else

hopefully is supplemental, you can look at it if you want to and like to. It's just a reminder that it is there.

OK, so measure testing, let's get down. We ask early on in the measure testing form is going. OK, this particular measure is based on administrative claims, so that's what they have said. Specified four administrative claims and (inaudible). That's what we want to see, we want to do testing that reflects the specification. And 1.2, given an existing data that was used, identify the specific data sets that this is basically saying tell me about the samples that you used to get a testing.

And often, when developers do a testing, they'll use the same data to do the reliability testing and the validity testing, not only they don't have to but that's usually what they do. So (Art) here is describing the data that they used to test this measure. It's basically the age (gap) data. And they give you a couple of paragraphs describing age (gap) data and what it is, how many patients that covers the measure. So you would look at that to see about the sample.

Now let's go further down. We asked them to tell us – this is even more details about the sample, we want to know how many measure entities are looked at and how many patients, and that's what they show here. Notice that they – in this age (gap) samples in 2011, there is more than 4,000 hospitals and I actually remember that this is a measure (inaudible) facility level.

So they're looking more than 4,000 hospitals and looking at as many as 3 million patients. Now, I don't think – I could be wrong, I'm not sure if those 3 million patients' going to full age (gap) data set for that year or those 3 million patients who had surgery, I suspected from the former but that at least gives you a flavor of the patient.

OK, go down. So here's a section on reliability testing and we asked at what level was the testing performed at. And one thing that we ask upfront is because the algorithm branches off depending on what level of analysis will be done. So recall that we allow testing at the data element level and/or at the performance score level.

The measures that are tested at the four levels have a potential to get a high rating again, depending on the sample used and the result whereas measure that are tested only at the data element level, their highest rating should be a moderate. And that's reflecting our feeling that we really would like to know about the score that's going to be used and if the scores that are going to – that facilities are going to be prepared on. So we really want to know about the reliability scores.

And ideally, we would be testing at both levels because that gives you different kinds of information. But in this case, they tell us under (2A2.2), that's at the bottom of the form, that they used a method of testing called the signal-to-noise ratio. And you can read that, they tell you what that means. But that is – it is an appropriate methodology to look at reliability of a performance measure score and basically it quantifies the amount of variation and performance that's due to actual variation between providers compared to what is the measurement error, random error.

So, you want more signal than noise is basically the idea. And what you end up with is the statistic for each facility and it will range from zero to one and one is better, zero means that all the variation is due to random error and one is all the variation is due to actual differences in performance. So generally, you'll see (inaudible) reliability statistics somewhere between zero and one, right?

And in general, there's no best threshold that, you know, if it's above this reliability level, it's good and if it's below this one is bad. But there is usually 0.7 if it's regarding minimum accessible reliability value. Again, that's just by, you know, some folks may say, no you need 0.85 or something better than that but again, 0.7 is fairly seen as a minimal.

So now, if we go down, so this is our method and now we look and we see in there their sample from the age (stuff) methodology, the signal to noise analysis. And then here is their result and they give us this table and we won't go through and study this table in detail here but basically, they have been in the reliability statistics overall. So, for all of the facilities that they tested.

And they also show information according to the site of the hospital, really the number of patients (inaudible).

And so overall, that last one there, overall average signal-to-noise ratio is 0.80. So again, remember for this particular test, each facility has a value and a average (fit) and the average is 0.8007.

If you look at the signal-to-noise ratios up in the upper part of that table, you could see that for the hospital, the facilities that had larger numbers of patients, the reliability was better. So you may want to think about this measure when you're thinking about it, you know, what happens if you're looking at really small hospitals or hospitals with a very low number of surgical patients. What happens to reliability there and what are the implications for the reliability of that measure?

And those are the things that you'd take into account when you're waiting and evaluating the measure for the reliability. So real quickly, if Amaru, if you could bring up the algorithm for reliability which we'll get into very quickly. There you go. So the green box talks about the specifications. So first, specifications are foundational, so you have to make sure that you agree that (inaudible). If so, the next question on box number two, what empirical reliability testing conducted using statistical test with the measures specified? In this case, it looks like the answer is a yes.

And then – so that takes you down to box four, where was reliability tested with a computed performance or and the answer of that one was yes. That takes you to box five, why is the method described, you know, appropriate, again, that was a signal-to-noise, so that is an appropriate methodology. And then that takes you to box six and box six there you're looking at again, you're think about the sample that was used and really the results that they got and you choose whether you feel that there's a high certainty of confidence that the measure four is all reliable or you had a moderate certainty or (really) certainty in that wherever we land there, is how you would rate that measure.

And just so you know, if you had answered no in box number four, if the testing and (inaudible) four level but it hadn't been done in the data validity

level and that takes you to those, I don't know, orange colored boxes there and note that the highest possible rating would be moderate. So let me stop there and see if you have any questions.

Everybody is clear is (inaudible). You know, we hope that these algorithms are pretty straightforward for you to use but, you know, I forgot to say, a little bit later, you know, you're going through your measures and you get stuck on something you're not quite sure, you know, give us a call or an e-mail and we'll help you through that (inaudible).

OK so let's go ahead and talk about validity. So validity is the next piece of scientific accessibility. I'll have to rush through this pretty fast, but validity, there's really three major things that you have to keep in mind because you only vote once for validity. We got to think about things.

So you have to think about specifications again under validity. So you think about, I mean in a little bit different way, specifically you think about specs in relation to the (inaudible). So for outcome measures, that really not a tremendous thing that you may have to consider but for example – and I have no idea if it will be the same for this measure or not but sometimes, people who really understand plane data may know that, you know, certain (IC) codes is never – are filled out right. You know, they're always not so good.

And so that might be something to discuss under validity in terms of specs. It usually had to do with evidence. So for example, if there's evidence that – a particular medication should be done within two hours and you're looking at a measure that has it being done within five hours, you might say hey, that specs are consistent with the evidence. That would be an example of that.

So the second thing that you think about under validity is the actual validity testing. So just like reliability testing, you think about example that you used, you want it to be a representative sample, you don't want it to be so tiny that you don't have a good (inaudible) that this measure would work well out in the real world.

For this measure again, they're using (age gap) data, you know, really into (reference), but that's not going to be a problem. You also look as you do the

reliability and the methodology that was used in the validity testing. So no surprise there. The methodology is appropriate. And then finally, you look at the results.

So, and our list real go quickly to the validity section. This little printer down, yes, and they're telling you and just like reliability, they can choose to test at the score level at the data element level, or so. But they're saying that they did at a score level. And one that I did here in the methodology, they talk about construct validity. And basically, that's the label that's used when people look at different relationships.

So they say, I've got this measure and I think that if this measure is doing what I think it's doing and what I want it to do, I should be able to see a particular relationship with some other – something or another that I choose to do.

So what they've done is they've actually come up with four – five relationships here that they tested. So they've done five different ways of thinking about construct (inaudible) and they – if you go further down, they tell you their thinking behind those. So the first one, volume, they are hypothesizing the higher volume and associate it with better outcome. That a little tricky in this one because their outcomes would be a lower number, right? Lower, right? if you're looking at a complication, right?

Female: OK.

Karen Johnson: So sometimes, you know, I read these and kind of makes my mind hurt but that's what they mean now. And then you can read the rest, if we don't have time, you know, to go through each of these but basically, they hypothesize a relationship and they told you which way they expect it so and then they actually go further down in the submission, they give you the results.

So again, you could throw that a little bit more, and there is the result and they actually interpret the results for you which is good and, you know, the more interpretation they do, the better because not everybody really gets, you know, all these coefficients and see them and stuff. But you would look at these results to see if (inaudible) that that has demonstrated validity of your measure.

OK, so we're getting short on time. So we'll make sure that I won't forget to talk about risk adjustment. So this is an outcome measure. And generally, we expect outcome measures to be risk adjusted. So just to remind you, risk adjustment is the process of controlling for patient factors, so present at the start of care that may influence patient outcome.

So basically, kind of in the vernacular, we talked about leveling the playing field. So we want to look at outcome but we don't want it to be confounded by things like, you know, this facility had sicker patient then let's – for this facility had older patient or for whatever. So, a risk adjustment tries to control to the extent possible so that you can have fair comparison between whatever it is you're trying to measure.

So, in this admission form, there's actually a couple of different places that you'll see information about the risk adjustment set. One is under the specification. So Amaru, if you want to go up to items (F)14 and 15 and 18. So, way at the top. Keep going up. Yes.

So as part of specifications in that shows that same section where they talk about – yes, here we go, where they talk about numerator, denominator. There is a place to talk about what they did and how they did the risk adjustment.

OK. So you'll see some stuff there. You'll of this stuff in the testing attachment. Specifically under item 2B4, so – Amaru, you're pretty good at this. You're getting down – you keep going down …

Male: 2B4?

Karen Johnson: Yes. 2B4. So that's just an FYI that there's information – couple different places, but basically, in specification section, (inaudible) here. They may could tell you that how they develop the model and what they ended up within a model. Because, you know, usually risk adjustment is done with some kind of statistical bobbling techniques.

And anytime you do modeling you are basically questioned, "How good was the model?" So, when you're talking and thinking about risk adjustment off the measures. A lot of the discussion falls around "How did the model went?" And there different things that they can tell you – if you go further down in the form, I'll draw your attention to – going down. These tables are showing what's in the model. So these are the actual factors that are in a model.

And then here they start telling you about discrimination and calibration. So those are fancy words that really just give you a flavor about how good the model is. And fee statistics of – statistics that often used to describe discrimination and basically it tells you how well the model can distinguish those with the outcome versus those without the outcome. So, again for this one, if those who had a significant complication versus those who did not.

And you pretty much, you know, want to have good discrimination in your model and usually, at these just a value of 0.7 or higher is considered good. But you can't look at discrimination in a vacuum, you also have to look at calibration, that's another way of looking at the goodness of it.

And generally, what – if you can go down further. A lot of times you'll see a calibration curves or things like that, they didn't actually draw you the picture but they showed you the numbers there – keep going down, there you go. These numbers, you know, predicted versus, you know, reserve rate, what you are hoping and what they (inaudible) is that you are hoping if you have a good – a well-fit model, those predicted rates will be very similar to your reserve rates throughout the entire space if you will (inaudible) split it up by decile.

So risk adjustment is one of the potential threats to the validity of an outcome measure. So that's the third thing that you think about when you're thinking about validity. So first thing is specifications (inaudible) with the evidence, second thing is validity testing, again, that incorporates samples and methods and results.

And then you also had to consider threats to validity, so threat to validity that we mostly think about are things like exclusions. For this particular measure, if we went further down, we would see that there are no exclusions to this

measure so that's kind of a loop point for this risk adjustment. So how good is the risk model that was used, meaningful differences in performance, comparability of results if there's more than one set of specification for this particular measure that is – there's only one set of spec, you don't have to worry about it, and missing data. There's a lot of missing data that could actually threaten the validity.

For this measure, there are actually – I think there is very little and maybe even no risk (inaudible). So, but that won't be a problem. So let me stop there, Amaru can you bring the validity algorithm so we will have that? And let's see if anybody had any questions about that (spiel) that I went through.

So that's a lot. This is a, you know, this measure is a little complicated. We didn't really even delve into the, you know, the numbers of results. So this is the algorithm for validity. So it forms – the four measures similar to that for reliability. So again, green box has to be with specifications, you think about this first.

You think next about the threats to validity. So again, for outcome measures, risk adjustment is probably the biggie, although other things could threaten validity as well. And then assuming that you make it to box three, it's basically testing – was testing them for the measures specified which is necessary and then get to box six where it ask, is it done for the measure score or the data element level? Can you go down just a little bit for us?

In this particular measure, they tested validity at a performance score level. So you're going to (inaudible) through boxes six, seven and eight and notice that you have your options – you could eventually rate it high, you don't have to. It depends on, you know, what the results but if you could (inaudible), so that's validity. There were, two other criteria as you know, and we're running out of time so we have lunchtime to talk about this that feasibility. And feasibility really has to do with burden of data collection and ease of implementation best described for you in your steering committee guidebook. And feasibility, you'll use that generic rating scale that we looked at earlier

These are usability in use, there's three things under the usability in use. It really has to do with whether the measure is used or not, that really comes into play for measures, it doesn't – it's not that big of deal if it's a brand new measure, that's never come in NQF before. But if it is a maintenance measure, then at some point, we expect these measures to be used. We expect improvement over time and we hope that there are no unintended consequences. Those are just the kind of the things that think about for usability and use. And again with those, you would use that generic measure of scale.

And so I have one minute left. Anybody have any questions? I realize I threw a lot of information at you. But happy to answer any questions you may have.

Male: We can just give you a ring if we really get confused, right?

Karen Johnson: Absolutely. I will say that, you know, if you're new to this, it can be a little overwhelming. Even through the workgroup process, people still are a little overwhelmed sometimes. And often in the in-person meeting, that first measure to – people are still kind of getting the hang of things. But believe it or not, we do get the hang of it and things go faster and, again, you really get the points that need to be discussed and, you know, it really does work out believe it or not.

Male: Good. And you expect us obviously all to have this online and advance or meeting where we discuss it. So we moved right in to …

Male: (Inaudible) accessible via SharePoint and you – we'll hope that you have laptops at the meeting and on the workgroup call so you can just pull up the documents as needed.

Karen Johnson: OK.

Male: OK.

Male: And let me ask a logistical question, so in the measure evaluation criteria and guidance summary tables, is that – as I recall, pretty much all of that is in the committee guidebook, am I correct in that?

Karen Johnson: The committee guidebook is an abbreviated version.

Male: OK.

Karen Johnson: It's kind of a flip note of that document. So the algorithms are on the guidelines.

Male: Yes. That's what I recalled, OK.

Karen Johnson: Yes.

Male: What would you recommend we have, the site or computer that we can refer to as where filling out these evaluation forms? Committee guidebook and what else?

Karen Johnson: That's probably it. There is an a document that we made available called, What Good Looks Like. And that's probably not something that you necessarily need to have with you as you go through but it might be something worth looking at if you're a little confused because that pretty much tries to show you what the kinds of things that we would expect to see from a developer that would fully answer the question. And that might just give you a little bit of a context.

In the in-person meeting, you don't have to worry about having these algorithms, we will actually have a printed-out version of the algorithm for you so you have that right there at your fingertips.

Male: OK.

Male: Thanks. Measure evaluation (inaudible) evaluating evidence (inaudible) testing documents.

Karen Johnson: I'll let them …

Male: Too much?

Karen Johnson: That's too much. We have some pretty detailed reports that go into details about thinking about evidence and think about reliability and validity and that sort of thing. And those are available on our website. We can certainly send those to you. I don't think they're going to (inaudible) on the project page but we could certainly send it to you or show you where those are.

They're pretty dense and a little bit long. If you're interested, then we definitely encourage you to take a peak at it but we realize that, yes, you probably won't have time to do all of that stuff.

I think in terms of doing the preliminary evaluation, we won't ask you to actually submit ratings, so if you're looking at for example reliability, we won't ask you to actually click whether you think it's high, moderate, or low. But on the preliminary evaluations, we would like you to give us some comments.

So I had comments that we're particularly interested in would be things that you see as problematic. So, you know, it's perfectly fine, you know, to say, you know, I think this is high and looks great, you know, but what will be the more helpful is if we see something that really is a concern and what we will do is we will collate those comments prior to the workgroup call so that you can see your comment and everybody else's comments, the comments (inaudible) those comments can form if nothing else, the basis of discussion in the workgroup call.

Male: I have a question along those lines. Is the document – a living document that's being shared between the parties putting together the workgroup presentation or is it individual efforts? Each person does their own and then we compare?

Wunmi Isijola: So just to Karen's point, every committee member will fill out those preliminary evaluations and then we will compile them into a document so that you're able to view your comments as well as your fellow committee member's comments prior to the call. So it will give you a sense of, you know, if your thinking is kind of aligned with your other committee members

or if there are kind of different aspects that, you know, everyone has kind of pinpointed to.

So we will compile that for you and that's kind of why we have the deadlines of the preliminary evaluation, so that way we can have the time to kind of compile that in preparation for the individual workgroup calls.

Male: The reason I asked is supposed to say, a Word document with 20,000 annotations in the margin. That would be something that you guys put together?

Wunmi Isijola: Correct.

Karen Johnson: Right.

Male: Got you.

Male: And what's the deadline again? I missed …

Wunmi Isijola: Depending on your actual – depending on your actual workgroup call, that's actually posted on the SharePoint site and we're pulling it up now.

(Off-mike)

Wunmi Isijola: So if you are part of workgroup one, the survey for that is April 29th, workgroup two is May 6th, workgroup three is May 9th, and workgroup four May 16th. And that's prior to each individual workgroup call.

Male: OK. So the 20th, I think I'm – you know, I'm the first one, the 29th, yes?

Wunmi Isijola: But you have a few days.

Male: Getting close.

Wunmi Isijola: But like Karen mentioned, if you do have any questions or need any guidance with regards to technical assistance, we're definitely here. We can definitely do one-off calls with individual workgroup or committee members if you have any questions. So that's definitely an option available to you.

And if we don't have any additional questions, as mentioned, our next workgroup call is May 1st – our first workgroup call is May 1st. And that will speak to the seven measures that were within the workgroup assignments that was posted under general documents.

We have Melinda Murphy, our other senior director and she will be speaking to that. But come with your questions, if you have any concerns about any of the measures and we'll definitely show it up there.

We also (inaudible) these calls and make it available to the developers so they will be there. If in fact, you need any clarification as it relates to any of the criterion, they can be there to kind of answer any of those questions.

And Melinda, do you want to speak to anything or Karen?

Melinda Murphy: I think that covered it.

Wunmi Isijola: OK. Well with that being said, thank you again for joining us for the second of two question – Q&A calls and we look forward to speaking with you on May 1st.

Male: OK. Thank you.

Wunmi Isijola: Thank you.

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

END