NATIONAL QUALITY FORUM

Moderator: Suzanne Theberge
June 6, 2012
4:00 p.m. ET

Suzanne Theberge: Good afternoon, everyone. And welcome to the Neurology Workgroup 3 Conference Call. This is Suzanne Theberge with the project. I'm here with Karen Johnson, Heidi Bossley and Jessica Weber, other members of the here at NQF. And before we start the call, we just like to have the steering committee members and the developers who are on the line introduce themselves.

We would like to go in the order of the agenda so if the steering committee members could just say their name and where you're calling from quickly in the order of the measures that we'll be discussing and then we can just go to the developers so we know who's on the line.

Mary, can you start?

Mary Van de Kamp: Yes. I just want you to (inaudible) this is Mary Van de Kamp. And I'm with Kindred Healthcare. And I will be describing the two outcomes: 0444 and 0445 Functional Communication, Spoken Language and Comprehension.

Suzanne Theberge: Jane Sullivan, are you on the line?

Jane Sullivan: Yes. Sorry. I was waiting for my agenda. This is Jane Sullivan. I'm from Northwestern University. And I represent the American Physical Therapy Association. And I will be talking about Functional Communication Measures for Wording and Motor Speech.

Suzanne Theberge: Thank you.

Ann Barrett: Hi, everybody. I'm Ann Barrett. And I'm a Behavioral Neurologist representing the AAN and also the Kessler Foundation, my organization. And I'll be presenting 0443 and 0446 – oh, I'm sorry, 0443 and then 0449: The Swallowing and Attention Functional Communication Measures.

(Tina Kiernan): Hi. This is (Tina Kiernan). I'm from South Carolina. I am representing AANN and I have 0441 Assessed for Rehabilitation.

Suzanne Theberge: We have (Dr. Sather), Dr. Eisenstock on the line.

Jordan Eisenstock: Yes. I'm Jordan Eisenstock. I'm a Neurologist at UMass Medical Center in (Lester) and I have 0448: Functional Communication Measure for Memory.

Suzanne Theberge: All right. Measure developers. Can you just let us know who's on the line?

Deidra Joseph: Hi. This is Deidra Joseph from AMA-PCPI. I have – I also have members of the testing and specification staff on the call.

Rob Mullen: This is Rob Mullen from the American Speech-Language-Hearing Association, the measure developer for the Function Communication Measures for 444 through 449.

(Toby Finmerk): And also (Toby Finmerk). Also with Rob.

Suzanne Theberge: Great. Thank you, everybody. The joint commission is going to call in later, so they should be online for processing their measure.

I would like to turn the call over now to Karen.

Karen Johnson: Thanks, Suzanne. Thank you, everybody for joining us in today's call. And special thanks to the committee members for reviewing all of these measures and doing this preliminary evaluation.

Just to give you an idea of how it's going to go today. We are going to ask each of you to summarize a measure. And for that, what I'll ask you to do is just very briefly introduce the measure and then go in order and summarize both your thoughts and the thoughts expressed from the preliminary evaluations. And just to highlight areas of concern and then we will open the call up to the other workgroup members to further discuss these points or any others that they feel need to be addressed.

We will have developers answer if you need them to any questions that you might have. And then as NQF staff, for most part, what we will be doing is just clarifying when necessary any of the criteria or the guidance. And also try to keep track of time. Now, we we'll be discussing 10 measures on this call and that means we don't have a whole lot of time for each individual measure. So, please forgive me if I have to kind of cut in on occasion and hurry things along.

For today, because of a mistake on my part actually, we don't have a lead discussion for our first measure so I am going to be the lead discussing that. And better off to give you a pattern of maybe how we will go forward in terms of discussing the measures. So, if we have no questions from anybody, I'm going to go ahead and get started on our first measure, 0224 from AMA-PCPI.

OK. This measure is Consideration of Rehab Services. And it looks at the percentage of patients 18 and older with the diagnosis of ischemic stroke or intracranial hemorrhage for whom occupational, physical or speech rehab services were ordered at or prior to in-patient discharge, or patients who have documentation that no rehab services indicated at or prior to in-patient discharge. This is a process measure and is computed at the clinician level of analysis. And it uses data from administrative claim as well as other clinical data, AHR data or registry data.

So the preliminary evaluations will start first with importance to measure and report. And we didn't have – by 1 P.M., we didn't have a whole lot of evaluations but we did have two folks who evaluated this measure. And pretty much on impact there was split between the two of the impact with one saying high and one saying moderate. And then for performance gap, there was also a split.

And in terms of impact one member noted that the data doesn't specifically address the impact of stroke we have. For performance gap, the workgroup members noted discrepancies particularly amongst particular group – subpopulation, subgroups, raise and gender that sort of thing. And they also noted that almost half of patients were not offered rehab services.

And then going on to evidence, both members who filled out the evaluation for this measure correctly noted that this is not a health outcome so they would need to look at quantity, quality and consistency. And again, there was a little bit of a split there on feelings, on ratings, on evidence. They note several studies reviewed. They note that there is an AHA, ASA guideline for this that supports this measure. And I think I'll stop there and see if other folks, other members would like to comment on impact, gap, or evidence for this measure.

Is everybody pretty happy that there are – that there are – is a room for impact or that it's an important measure to look at?

Jordan Eisenstock: Yes, I'd say so.

Karen Johnson: OK.

Anna Barrett: I think it's important as implied in looking at any results of rehabilitation.

Karen Johnson: OK.

Mary Van de Kamp: I think then I think it is as well. And I think it's almost, you know, as we start to look at the outcomes, ways in which, you know, the claims, electronic data support the identification. That's really important that we're able to draw, that I think the challenge we have of, you know, many times the full EMRs are not in place. And so

sometimes you are left with, you know, more financial claims data then maybe some of the other pieces but it's absolutely necessary; we begin to have this as we've had it in the past.

Karen Johnson: OK. Any other concerns at all about any of the evidence, the consistency of the evidence? Anything at all in terms of importance to measure on report? OK. If not, let's go on to scientific acceptability of the measure properties. The developers for reliability showed data element reliability and capital values roughly around 0.75 in that range. For validity testing, they gave results of their systematic phase validity testing.

And in terms of the comments from evaluators, I really deal with no negative comments at all. One thing I did want to point out, both reviewers noted a high rating for reliability and I just want to remind you that in terms of the ratings reliability and validity, remember that what you're looking at is the specifications themselves and then the levels at which the testing is done in addition for validity, you're thinking about the threats of validity and if those are assessed well.

So, for these key measures since they did data element reliability, this measure would be eligible for a moderate rating if you were happy with the results that they provided, again, the capital values were around 0.75 for that. And then for validity, they used phase validity so from that – for that, this measure would be eligible for a moderate rating for validity because, again, they did – they did phase validity. They didn't do both data element and measures for validity. So, I just wanted to point out how the ratings there should work.

So, is there any question or any discussion about the method that they use or their results? Everybody loves the scientific acceptability. No concerns at all?

OK. Great. This is – this is great. Usability and feasibility. And for usability, both reviewers gave it a rating or high usability. This measure is used in the PQRS system and it's also used with JACO primary certification program and the guidelines program and also the Paul Coverdell registry. So ample demonstration of usability and then the same for feasibility both evaluators gave it a rating of high. So, there's anyone have any comments about usability or feasibility of this measure?

You guys are the quietest group yet. I don't know if that's good or bad.

All right. So, for the preliminary assessment, both evaluators said yes that they would feel comfortable in recommending this measure for endorsement. So, is there any disagreement with that, any concerns about this measure that you would like to talk about at this point.

Jordan Eisenstock: No.

Karen Johnson: OK. All right. So, everybody likes this measure. Great. And just so you know at the end of the call, we will be hearing about the joint commission measure and it is very similar to these measures. I think that will be interesting as well.

So, what I wanted to do is start out the call with a fairly easy measure. It is a process measure. The next eight measures that we will be discussing are outcome measure, again, from the American Speech Hearing and Language Association folk. And I'm actually going to do these measures a little differently. First, I would like to go through a set of questions that we sent to Rob Mullen this morning.

I would like to address the questions and answers that they provided first. And these questions really address all eight of the measures in some way or another. And then after we kind of finish the discussion about those questions we can go into the individual measures. So, I'm not sure, hopefully, everybody was able to take a peek at those questions in Rob's answers but just to give it on record, we will go through data.

So my first question was the numerator in your measure is the number of stroke patient who make progress in a particular area. So my question was how do you define progress. So Rob, would you like to go ahead and just answer that for us?

Rob Mullen: Yes. Basically, it is a seven-point scale so progress is defined as movement from the patient's low-added mission to one or more higher levels on that scale.

Karen Johnson: OK. Thank you. OK. What is the standard time period for measuring progress for improvement? So, for example, is it from admission to discharge or do you look at a particular period of time such as, you know, the first 30 days or the first 60 days?

Rob Mullen: The way that the measure is expressed for NQF purposes and this is the way that we didn't use most frequently is from admission to an SOP case load and then discharge – I'm sorry, Speech-Language Pathology and then discharge from the Speech-Language Pathology case load which obviously typically but not one (inaudible) at the time coincides with admission and discharge to a particular care site.

There are – there are some folks who use the measure for other periods of time. They will use it to come back and to, you know, 30-day follow-ups to see what progress the patient is making but the specification submitted NQF file for measurement at admission and again at discharge.

Karen Johnson: And I would like to follow up on that real quickly, Rob, just to make sure that I understand. Let's imagine a hypothetical patient A who is getting treatment for both swallowing and spoken language, for example.

Rob Mullen: Right.

Karen Johnson: So, for their swallowing time period. Is it admission from treatment for swallowing to discharge of treatment for swallowing or is it admission from either or both, you know what I'm saying?

Rob Mullen: I do. And it is – it is from treatment – from the initial treatment of swallowing until swallowing treatment how it's included. So if you're treating for swallowing and motor speech and you complete your swallowing treatment before you complete your motor speech treatment, you do the discharge score at the time that you complete the swallowing treatment.

Karen Johnson: Right. Thank you for clearing that up.

Rob Mullen: Sure.

Karen Johnson: OK. Then the third question was about your denominator exclusions that you mentioned. And probably the main one is those who have only one treatment session. So, my question was how many patients this is involve? But can you just elaborate on no data are available on that and also talk about what happens to folks who come in and they are assessed at the highest level of functioning. What – are they included in the measure? Are they not included?

Rob Mullen: Sure. Well, the first part of that question is the one denominator exclusion that we do have is patients who are – who receive only one treatment session. And the question is how many of these are there and we don't have any way of knowing that because by definition, they're not reported to us. We don't know how many patients aren't reported to us. Basically, the way these measures are most commonly used is to essentially a registry maintained by American Speech-Language-Hearing Association.

And so, we've been using these eight measures and about 10 others since 1998 and the outcomes reporting system. And that's one of the rules we use for these measures within – from reporting was we only want patients reported to us on these measures who have had more than one treatment session for that particular disorder.

And so, for patients who's seen for evaluation for evaluation-only or for some reason for an evaluation plus just one treatment setting, those aren't reported to us. And so, we don't have any way to know how many of them are out there. I don't know if my colleague (Toby) has any anecdotal information about how likely she thinks that is, that there are many people out there getting just one session, but in terms of higher data, we don't have any.

(Toby Finmerk): Yes. This is (Toby). I don't have them a guess on how many are – what I would consider evaluation-only. And I would just anecdotally, I think, you know, most individuals that are referred to SOP services are referred because they need treatment. So, I don't know how to answer that.

Karen Johnson: OK. That you're saying that in general, even some inundated to the registry folks only do that when somebody has at least two visits or two treatment session set.

(Toby Finmerk): Correct.

Karen Johnson: OK.

(Toby Finmerk): And then on the highest level of functioning, usually somebody at that highest level of functioning again, does not need Speech-Language Pathology services, so they wouldn't be treated. And again, therefore they would be excluded if they're at the higher level.

Rob Mullen: Well, I would actually care – I actually resist that's in the notion of reference of these as exclusions because there are not really exclusions anymore than saying a patient with broken legs of exclusion because we're not treating them. I mean, these were never people that we should be treating in the first place. So, I think the very specific usage of the term exclusion on this context as I understand it this isn't an exclusion in terms of – because these people are people who would not be candidates for treatment.

Karen Johnson: One thing about just a follow up on that a little bit is in your submissions when you talk about the denominator detail, either staff – generally, I think it's seven levels.

Rob Mullen: Right.

Karen Johnson: So, it was a little confusing because I guess theoretically the denominator there would only be folks who fell into level one through six.

Rob Mullen: Right. And...

Karen Johnson: Am I correct on that?

Rob Mullen: And the reason we did that actually is because of NQF. When we went through the approval, the endorsement process before, when, you know, we submitted it as an admission that can be levels one through six and at discharge that can levels one through seven but they found that very confusing. And so, we were explicitly asked to list all seven points of the scale to avoid confusion on the other direction.

Karen Johnson: OK. I guess, it's kind of what because these are people at different times.

Rob Mullen: Right.

(Karen Pace): This is (Karen Pace). And, you know, after we kind of go through these questions, we'll probably have to work on getting this cleared up a little bit because it's a little bit confusing the way it is now and I'm not sure about what, you know, what the steering committee recommendations may have been generally, you know, the specifications need to be specified as (Yuler) have developed, tested and implemented the measure. So, we'll work with you on that.

Rob Mullen: OK.

Karen Johnson: OK. The next question, question for probably the most difficult question that we're going to ask you and that is, from what you're pointing in your submissions we're still a little confused about how you control for a patient severity at (day five). So, and let me just tell you how we're getting confused.

You mentioned doing statistical risk adjustment but you illustrate different what looks like strata that you would report at. So, for example, for the lighting measure, you talk about level of severity and time (post struck) as what looks to be strata. So, we are just kind of unclear about is this a stratified measure, is it a statistically risk-adjusted measure? Can speak to that just a little bit.

Rob Mullen: Sure. We use the function communication measure at admission as our measure of severity. And so, when we divide the risk-adjusted groups in every case, they – it is adjusted by the start score. So, for anyone of these and

you can see in section 2B4 that when we prevent – when we present the risk-adjusted groups in every case, we report separately for each function communication measure at admission.

Karen Johnson: OK. So, let me – let me run on two things there, Rob. One is when you showed like for the writing line you have that scored admission and then date for discharge. That looks like they're stratifying not only on level severity but also on time since stroke.

Rob Mullen: Right.

Karen Johnson: So – which is a – is it only level severity or is it level of severity and stroke – the time since stroke?

Rob Mullen: The latter.

Karen Johnson: OK. So, for the writing measure, you're saying that you would stratify – you would have 12 different strata for this measure.

Rob Mullen: Let's see, let's see, we four, five, six (inaudible) 12, yes.

Karen Johnson: OK. OK. The other thing I – you mentioned that you're putting your risk-adjusting groups in each of these strata so we're confused about what you mean by risk-adjusted groups. How have you risk-adjusted?

Rob Mullen: Well, the risk-adjusted group are these 12 – we defined their criteria buy doing regression analyses to identify the factors, to enter into our risk model. And so, based on that, we develop in this case for writing these 12 risk-adjusted groups.

Karen Johnson: OK. So, let me make a statement and tell me if I'm correct or not.

Rob Mullen: OK.

Karen Johnson: You've done some regression analysis and that's giving you an idea of how you should stratify.

Rob Mullen: Yes.

Karen Johnson: OK. So, you really haven't actually applied any risk adjustment – you haven't done a statistical risk adjustment of anything. You just use that analysis to tell you how to stratify.

Rob Mullen: Correct.

Karen Johnson: OK. So that helps. And getting down into the leads a little bit. Do we want to getting the leads of this or that we want to do this a little later?

(Karen Pace): Well, I guess, one of the questions is, you know, and you provided some performance data on one level but it indicates that there was any distinction about times and the timing issue that you've talked about. And if you have each measure developed, well, I don't know if they're all in the 12 strata. But do you actually have enough patients to be reporting 12 scores for every, you know, clinician or every facility to – I mean, don't you get some really – how are you dealing with the number issue?

Rob Mullen: I'm not sure I fall about 12 scores per clinician or per facility.

(Karen Pace): Well, you're getting performance scores for a clinician or facility and you're saying that you've broken it – the way you're handling or controlling for risk is to report the score by strata. Is that correct?

Rob Mullen: Right, right.

(Karen Pace): And if you have 12 strata because of these combinations of level of function and the time, it's just seems like you're going to get into a very small number issue of how many patients are in each of those strata to have performance scores that will be, you know, very stable without a lot of...

Rob Mullen: Right. Well, what we do is we – when we compare a facility's performance on a particular measure with benchmark data from across the (sensory) but based on that particular strata so we don't – we don't try to roll all of this out, for example. So, where the facility has a sufficient number of cases in that strata then I can make meaningful comparisons to benchmark data for that strata but we don't – we don't combine them.

(Karen Pace): Right. That's what we're understanding. So that was our question. But we'll, I guess, move on – we'll need to probably get some examples of beyond just the one level that – we'll move on, we just want to make sure we're understanding what you're actually doing with the measures if that was – if that's the last question.

Karen Johnson: Just the one more question. We had noted that one of the supporting documents that you had provided with that measure specification doc. And it didn't quite match some of the things that you put in the submission. So my question to you is should we disregard that one? And so, how would you take that, Rob?

Rob Mullen: I guess I would say essentially, yes. I mean, we understood that that's the mission document was part of the – our requirements that came with store endorsement in 2008. And so, we didn't want to change that documents and not have but we consistent with what NQF endorsed in 2008. But in the subsequent use of data collection, we have been able to refine our risk adjustment. And so, yes, we do have updated information which would be more appropriate to put out there in that document.

Karen Johnson: OK. Yes. I think either – give us the most current or have it actually even better within submission.

(Karen Pace): Right. The specifications need to be in the submission form and in our database so I think we would just not want it. We don't really want specifications in separate document unless it's like code list or, you know, the actual coefficients and formulas for risk adjustment. So, we can probably disregard that.

Karen Johnson: OK. That helps. And one more, this is a little in the lead that in that old document you had talked about, for example, for the speech language comprehension measure for whom at least 40 percent of their Speech-Language Pathology treatment was focused on language comprehension. Have you kind of gotten rid of those timing requirement?

Rob Mullen: Yes, we have.

(Karen Pace): OK. Thank you.

Karen Johnson: All right. Thank you very much for being willing to address these questions. I think they were – they were things that we were having trouble with understanding here.

So with that, what we're going to do is ask Mary Van de Kamp to start with, I think, measure 0444: The Spoken Language Expression Measure and introduce the measure and then we'll just go kind of in order the way I did. So, we'll talk about just the first criteria and importance to measure and report. And then have any discussion about that and then proceed to the rest of the (criteria). So, Mary?

Mary Van de Kamp: Yes, great. Thanks very much. Yes, the one 0444 is a measure that describes the change and the functional communication status subsequent to speech and language treatment related to the spoken language expression.

The numerator as we say is the patients who make progress of the spoken language and the denominator are those number of patients have scored on the spoken language mission. And these are called SCMs, just – maybe as we go forward just as if I say that, that's the functional communication measure.

And a grab address to some of the exclusions, et cetera. It is an outcome measure and not a process measure as we've worked through. I was able to add some survey results this afternoon. So, as I speak to the summary of the data that you have, I will add my information to that as well. So the following evaluation are specific to the impact.

And there were too high and the performance gap was in medium and a low and I had a high so I think we've evenly split across the three. The rational and is that patients receiving speech and language services retreated, and nearly 30 percent field progress so no data performance gap indicated disparities and the outcome.

Evidence-based, there were three – there would be three yeses. And it is a health outcome, again, three. So, as we looked to the quantity and I think maybe I have insight information on that to some degree of our company participates as one of our many outcomes that we track. And so, I assume the quantity and I think that's – from a, you know, user standpoint, the quantity supports the ability to make inferences, et cetera, from that.

And the quality again within the – one each and I had high in that area as well. So I think we – since three of us are split, any comments and questions specific to the impact and maybe more even neurology gap?

Anna Barrett: I guess I'm trying to – I'm – this is Anna, I'm just trying to understand the performance gap split a little bit better. It looks to me like there – as you say there's a number of people who don't like progress and then there are some variability across clinicians as well have not tremendous variability. Does that the source of the performance – the difference in terms of the judgment of the performance gap?

Mary Van de Kamp: I think so. I think – I agree with you. I think it's a little – when you're looking at the performance gap specific to these measures, it could be interpreted to be, you know, that's exactly as you said it. I don't know – Karen or – do you have any thoughts on that maybe understanding that definition better than I do?

Karen Johnson: I'm sorry. Can you repeat that question, Mary? I'm sorry.

Mary Van de Kamp: Sure. They were – the question is around the gap in determining, you know, that is it because there's nearly 30 percent to field progress, is it because there's just – is there some therapist decision process in case in that, or how would you see the gap of, you know – how you would see that – this piece to the overall measure?

Karen Johnson: Well, I think if I'm understanding your right, what they're saying is they're kind of talking about an overall poor performance that there is just an overall gap so in other words just not enough folks are doing well on this measure. So, that would be indicative of a gap. I think...

Mary Van de Kamp: I think one of the greater said that was insufficient.

Karen Johnson: One way they're noted insufficient evidence that we'd be interested in knowing why...

Mary Van de Kamp: Well, and I think this speaks to one of the questions we had. I think the information that was provided was to take this five clinicians and five facilities and report scores for one level of the stratification. So, it probably is something we need to get a little more information on. But maybe have the steering committee member on that talked about insufficient evidence for performance gap.

Anna Barrett: This is Anna again, I didn't note that for this measure but I would agree because it appears to me that I'm not getting kind of a fold stance of what their relationship is between failure to recover and the clinician's effort to treat.

(Karen Pace): All right. And let me go back and just explain what we're talking about in terms of performance gap. I'll get back to that first question is we're really interested in the performance on the measure as specified. And what we mean by performance gap is that on that actual measure so on the outcome that there's either variability and performance across those being measured so the pathologist or the facility or that there's overall poor performance.

And so, you know, it's a kind of typical – there's opportunity for improvement if there's quite a bit variability. And though there are some outcomes where there may not be a lot of variability but there's just overall poor performance that needs to be addressed. In this case, we may not have enough information given that, you know, they gave us one score for five different clinicians in five different facilities but also just on that one level.

So, that's something that we can – what we like to see here is kind of a distribution of the scores across all those entities being measured so that you can get a better sense of the variability and performance. So if it's – or to this just overall poor performance.

Mary Van de Kamp: And maybe you could help me because I had a question that I had in across all the measures that I look at. So, tell me the difference – tell me – if you say poor performance, explain me – explain to me what you mean with that a little bit more.

(Karen Pace): We're talking about the results on this particular measure.

Mary Van de Kamp: Right.

(Karen Pace): So, we're not measuring the processes or the – that went into getting this outcome but the idea is, is there variability on this outcome such that it looks like there's room for improvement, or that, you know, that does not making much progress across the board and everyone needs to improve.

Anna Barrett: Right. And I think they got you. So for example.

Jane Sullivan: Let me clarify – sorry, this is Jane. So, when in many of these where there's a statistics for the percentage of patients who failed to progress, that was my interpretation about, you know, I mean there was room for improvement when there were just significant number of people who failed to make progress on that particular measure. Was that a current interpretation?

Mary Van de Kamp: Yes, yes.

Jane Sullivan: And the other issue with the disparities across, you know, different categories of patients.

Mary Van de Kamp: Correct.

Jane Sullivan: OK.

(Karen Pace): Yes, it can be very (delayed) in performance across those being measured or the disparities issue.

Jane Sullivan: OK.

Anna Barrett: Can I ask you a question though with regard to the 30 percent of unprecedented failed to progress since we don't have data on whether these people – how many of these people were at level six and then failed to progress to seven or we're in different stratification groups that may affected outcome. This may lead some to judge that the data is insufficient or that it would benefit for more data. Is that correct?

(Karen Pace): Yes. I mean, you know, that obviously, that's a substantial number. And – but what we're saying is just in terms of what we're asking for our per year review is to have a bigger picture of not just that one number but kind of what the distribution is.

Anna Barrett: Definitely. Yes, because I'm wondering how – what the percentage who are level six and if that was 25 percent then, you know, maybe that was the percentage that failed to progress.

Mary Van de Kamp: Right. I think, you know, and I think just in using them frequently, if there – that is I'm (inaudible) yes, taking someone to the outcome analysis in getting on the leads myself. But I think you're right. There's a lot of variables to, you know, that's what – that's why there's value to measuring something that you can then look at what were the – what were the reasons for that. Was it, you know, the complexity? Was it the multiple comorbidities, the side stroke, was it – was it the fact that they were very high (radiant) and, you know, they didn't get or they were so severely impaired that you didn't see them for a long time because there wasn't improvement but at least you had a measurement that bury that.

So, that's kind of where I'm getting caught. But, I mean, that you're saying, Karen, take it to the more outcome analysis as oppose to how it's used within the settings and how (inaudible).

(Karen Pace): Right. Because this is an outcome measure. And you're right, when you're looking at your own data in the setting and looking for areas of improvement, you're going to have to dig down and look at all those factors. But in terms of the outcome measure, you know, if there's variability that indicates there's room for improvement and that we do ask for performance data on the measure as it specified and being implemented.

So since this measure is stratified, you know, it gets at least part of the question that someone was asking about, you know, where is this. So, we can follow up with the developers on that.

Mary Van de Kamp: OK. All right. If there's no more questions then the scientific acceptability of the measure properties, there were three yeses including mine and the reliability would be the two high-end and one medium. The rational, I think developers for the finale is on the evidence of convergent that's coming a constructive validity, consumers' questions are related to patient satisfaction of service not from around the measure. I think customer satisfaction is, again, as we speak to those is influenced by – as we all know a lot of other issues.

So, I think as the issue look at the scientific properties, if there are any questions, since we all rated high on that. And then maybe if there's any questions on the reliability since there were some that maybe had a question on that piece.

(Karen Pace): So, this is (Karen Pace) again. And one thing just to keep in mind here that the reliability and validity information they presented were for the actual instruments and we consider that the patient level data that are going into this performance measure. So it would really only be eligible for a moderate rating. And also, because these are outcome measures, the whole issue of risk adjustment or re-stratification and demonstration of the adequacy of that affects the validity rating. So, it's something that we need to follow up on as well to...

Mary Van de Kamp: Right. So, as a pre-survey – begging for more information before adequate response if that what's you're saying in that. So that would be the same for all of those in this eight group, (Karen)?

(Karen Pace): I believe so. We just need to clarify that analysis that Rob talked about and how that relates to, the final stratification that they chose.

Mary Van de Kamp: Understood. So, as we move forward then, would you just state again, (Karen), the fact that it would rate only a medium and just restate that again.

(Karen Pace): Right. So, our – the NQF guidance on rating, reliability and validity is on a scale of high, moderate, low or (inaudible) evident. And for a high rating, measures have to have been tested at both the data elements and the performance measure score level for both reliability and validity.

Moderate rating, we allow testing it either data level or the performance measure score level and of course, either way, the testing have to have a reasonable result. And then low would be if the testing actually indicated that it wasn't reliable, wasn't valid.

Mary Van de Kamp: That's very helpful.

(Karen Pace): OK.

Mary Van de Kamp: Anything else specific to number two? Usability – now you've got me doubting myself on this one, (Karen). Is that – because I read those (HMNL). Is this another one that would have a two-fold?

(Karen Pace): No, this is – this one, you know, the – it's not as complex so the high, moderate, and low are really kind of your assessment of, you know, whether it's usability.

Mary Van de Kamp: OK.

(Karen Pace): So, your...

Mary Van de Kamp: All right. All right. I make sure I understand (inaudible). So usability two is at the high, one is at the medium and the usability factor – oh, is I think impacted as stated here, is that – that was about from private stakeholders and is used in reporting and recommended by CMS as part of – as part of the initiative to start to implement outcomes across the continuum.

Karen Johnson: This is Karen Johnson. I did have one question here. I think on the – on all of the measures under criterion 3B usefulness for quality improvement. I think the developers actually didn't put anything down for how these

measures are used in quality improvement. So, generally, we would ask that either tell us how it's being used or maybe the progress of it's being used if it's not actually in use. So, that might be something Rob – that you guys could add in.

Rob Mullen: OK. That's 3B?

Karen Johnson: Correct.

Rob Mullen: OK.

Mary Van de Kamp: OK?

Rob Mullen: Yes.

Mary Van de Kamp: Anything else? All right. The feasibility, and this is made up of the data, generate in the process, electronic data susceptibility and accuracy and unintended consequences and the data collection strategy. Any comments from the – from the committee on that?

All right. And then in summary, the preliminary assessment as (civil) for endorsement adding mine. We have three yeses and obviously, we still have some committee members to weigh (NMS).

So moving I think hopefully quickly to the comprehension component of the spoken language functional measure 445. I think as I'm looking through all of these eight measures certainly there's similarities in some of these areas. The descriptions physically though it describes a change in function, functional communication status, subsequent to Speech-Language Pathology treatment related the spoken language comprehension.

And again, the numerator and denominator would be the same as in the expression. And I think the measure, again, is an outcome and the data source as identified earlier is a number of areas of Electronic Medical Record and registry as also paper medical record.

And preliminary evaluations rated the impact, all three at high and again the gap. And I think we talked a little bit about that. That probably have some similarities to the discussion we had previous. Is there anything else that would – that's more or like to add to that or that's obviously a follow-up?

All right. And I think there's on the quantity side, I think we're probably dealing with again, with the small subset that was given an example. And I know going back to look for the larger number representing the bigger number of outcomes that are entered. Consistency, I see that we have a question maybe from the committee on that piece of the impact? No?

Karen Johnson: And, again, this is Karen Johnson. Just to clarify, in terms of these measures being outcome measures, it's great if the developers want to give you evidence about, you know, information about the quantity, quality and consistency of the actual evidence behind the measures but we don't require that for outcome measures.

So, you know, what we do ask for, for outcome measures is that there – they had at least some rationale for why they think their measure might relate to and improve health outcome. So...

Mary Van de Kamp: So the numbers as more the intent as oppose to the numbers that have demonstrated that?

Karen Johnson: Right. So, we – yes, exactly, because outcomes are, you know, there's a – the name of the game really and quality. So, and there are many, many different processes and things that could affect outcome. So we really just are looking for at least more in rationale. And I believe in these measures what Rob have you doing is they are giving your some evidence showing that there is a treatment effect in that. And...

Mary Van de Kamp: Both from them.

Karen Johnson: Right.

Mary Van de Kamp: But I'll (inaudible) treatment effectiveness. Right.

Karen Johnson: Correct. Yes. So, you really don't have to talk so much about quantity, quality, and consistency for outcome measures.

Mary Van de Kamp: OK. Right. Thank you. Thanks.

Karen Johnson: Sure.

Mary Van de Kamp: And then on the scientific, I think, again, if I'm understanding the previous question back to Rob. This is an area that you will get for the information specific to the risk facility, right? Is that correct? Or I'll – is this an area across all the measure, Karen, that that question will be need to be answered?

Karen Johnson: Correct.

Mary Van de Kamp: OK.

Karen Johnson: Correct.

Anna Barrett: This is Anna. Can I just ask a question about what was just said that, you know, under structure process outcome for this measure, I just don't see that there's a treatment affect relationship discussed. Is there a treatment like noted for this measure?

Mary Van de Kamp: Let me pull up in my hand a second.

Anna Barrett: Or maybe that's not as important as that it's just an important measure by some standard with some gap.

Rob Mullen: This is Rob. I think either 1C1 addresses that.

Anna Barrett: For this measure – this is Anna. I'm just looking at to the spoken language comprehension, correct?

Mary Van de Kamp: Correct.

Anna Barrett: And I'm looking at in the sense as it – this is strictly an outcome measure that types of evidence requested are not directly relevant.

Karen Johnson: You're right again. This is Karen. And I think hours of treatment and relationship to present making progress was actually noted in the – in the writing measure which is the one that I'm concentrating on. So, Rob, I'm not sure if you have kind of similar information for all of the different measures or not.

Rob Mullen: We do. We can provide that.

Karen Johnson: OK.

Anna Barrett: That would be great.

Mary Van de Kamp: And it is inherent in the managerial but it's not pulled out I think as well specific to that.

Karen Johnson: OK. Yes. Having that would be interesting if nothing else especially for folks to maybe aren't familiar with speech language treatment.

Mary Van de Kamp: All right. Anything else? Good thoughts, usability, again, comparables to the previous measure and that it was developed with input from stakeholders and has been used in demonstrating outcomes across payers obviously and providers.

Feasibility, I think, you know, we talked a little bit about that previous any additional questions relative to the comprehension, a strong outcome.

All right. And then in preliminary assessment, the endorsement of the three of us that have weight in to in favor 1 verbally agree to at least preliminary to endorse that. So, Karen, those are my two measures.

Karen Johnson: Thank you, Mary. I appreciate that. The next set of measures I think we're asking Jane to take for us. And Jane will do the major job a little harder or a little (inaudible).

Jane Sullivan: Right. I think we've made it a little easier. It seems it's still most of – or many of the clarifications that have already been made are – just or, you know, any of the things that might come up have already been clarified. But I have two measures. The first one is 0447 which is another one of the Functional Communication Measures that's specific to Motor Speech and added before it's the proportion of individuals who make progress on this measure over the number of people who receive treatments from the speech-language pathologist for difficulties and motor speech production.

There are two people who reviewed this. So – and this is an outcome measure not a process measure as we discussed with the prior functional communication measures. With respect to importance, it seems as though that there is a sizable percentage 24 – just a little over 24 percent of individuals who failed to progress on this measure which certainly there's evidence of room for improvement as well as the disparities that are noted on outcomes based on some of the categories that are listed.

And so, it seems that I'm a little – I'm a little confuse because an importance to measure on report, there's only one yes but in evidence and whether this is a health measure there or two yeses. So, it seems as though the reviewers of this measure were consistent in viewing this as importance with opportunities for improvement and performance gaps. Anybody have anything to add to that?

All right. So, in terms of scientific acceptability, two of the reviewers said yes and both reviewers rated this as a moderate. The one question that I raised which I think has been clarified with the time window so that was made clear earlier on. But I think the discussion with regards to reliability and validity is pretty consistent with what we had on the other functional communication measures.

I don't see any specific issue that identified that have not already been raised. Would everyone agree?

All right. In terms of usability, one of the reviewers rates this as high and one as moderate and I think this was my comment and part of my rationale and I think I was the moderate one. And I believe in the (descriptor) of this category, there was something to the effect that the results could be meaningful and understandable so I didn't feel that that was specifically addressed.

I mean, it's sort of assumed, but that was, I think, my reason for rating this as a moderate and not as a high.

And feasibility, again, one person rated it as a high, feasibility, the other as a moderate. And I don't see any specific questions raised there. And both reviewers said yes in terms of the suitability of the measure for this preliminary endorsement and a (need) discussion.

Well, so that was quick.

Karen Johnson: This is Karen. Rob, did you have any comments about Jane's question about a demonstration of the results or meaningful and understandable? Did you have anything that you wanted to mention here, just in response to that comment?

Rob Mullen: No. I'll go back and take a look to see if there are other – if there are other things that we could add. I'm just trying to look – I'm looking right now what we've submitted and so we can take a closer look at that and see if we can come up with specific examples of how and whether these were – these data have been found to be useful.

Jane Sullivan: And I guess – I sort of had this question about a number of these – you know, there was information provided about intended actual (lagged) use and current use. I didn't see and maybe it's assumed just because some things used doesn't mean it's meaningful or it's useful. So that was – that was my question. And I'm not sure if I'm, you know, overly interpreting the category or not.

But if it's in fact that was what we were looking for, I did not see that.

OK. So if no other discussions, I'm going to move on and do the second one which is 0442 which is another functional communication measure with regards to writing. Again, this is looking at change in functional communication status in writing, in people who are treated by speech-language pathologist. Those – the proportion of those who change but all those who received treatment for that dysfunction.

And I seemed the only thing that was different or additional in here was this exclusion which was patients using and augments this alternative communication system. It seemed reasonable to me for an exclusion.

This also is an outcome measure. The reviewer – both the reviewers were consistent in reading yes for importance, one high, one moderate. Just a little over 30 percent of people failed to make progress on this measure which certainly is (inaudible) undesirably high percentage.

And disparities were just identified when you started to stratify the categories. So further study of those disparities, I think, is warranted so there was an agreement across reviewers that this is an important measure.

In terms of evidence, again, because this is a health outcome, it – but their evidence wasn't rated. Moving on to scientific acceptability, both reviewers rated this as a two and my only question has already been addressed, again, about the time window. Both reviewers rated both reliability and validity as a two or moderate.

In usability, this was rated as high by one reviewer and moderate by the second. And, again, not – I don't – I don't see anything substantially different in the comments here than what have already been discussed as well as in feasibility which seems pretty similar, one reviewer high and one reviewer moderate and not an agreement across the two reviewers that the preliminary assessment with the measurement was suitable for endorsement.

So I kind of rolled through that very quickly. Did I (inaudible) to any discussion or – because no one – no one jumped in, does that mean that everyone is in agreement.

Anna Barrett: I have a damn question about the validity, I mean, agreement with everything is that this is on – but with regard to the validity, we've documented in which we had presentation of the development of validity only used validity based on item or test in disease. Is there other data that's available for validity in the four years of use subsequent?

Rob Mullen: There are not, no.

Jane Sullivan: Well then I – that's a point, then I would probably raised the same question about reliability because the reliability data was about the development of the tool.

Anna Barrett: Is there any such data?

Rob Mullen: There certainly are for reliability, yes, but not for validity.

Anna Barrett: All right. This is Anna Barrett, Anna Barrett. Are we ready to go on to measure 0443 swallowing?

Karen Johnson: I think so if nobody else has any question about the other measures. OK. Go ahead, Anna.

Anna Barrett: All right. So this is, you know, I'll do an abbreviated session because we've heard about a lot of these. Seriously, it's a health outcome measure looking at inpatient and outpatients who are undergoing speech-language pathology treatments for a patient and data source is noted previously through ASHLA. Patients are only excluded as a result of one treatment session.

And with regard to the evaluation, we had three data points and the most – I think those data points was mine, I think, partly based on this understanding of some of the impact and performance gap issues requirements, so this may be very straightforward.

With regard to impact, there was relative agreement. There's a large percentage of people who – among the group of people who've been assessed so far, who've been reported to have swallowing abnormalities and – or

large group of seven – a group of 7,240 patients and then 33 percent didn't make further progress so that's potentially a good group of people to understand better given the potential importance of that outcome.

Some of the other comments though that were made was that we – there are some concern about the validity of the measures. I think those go back to reference kind of the discussion we've been having about on-going evaluation of the measure because there's a – it says, one of the comments was that the outcomes might be selected to better.

I think there was actually a note in the materials we got from a test development that there was a positive bias, so.

Also there's disparity noted (insurance) or in baseline status diagnosis in age. With regard to the evidence that was available, I had voted no and two people voted – two other people voted yes, so if I change my vote because, again, I think that for health outcome measure I was using more stringent criteria than we've been instructed now and that I've now understood, then we'll have a unanimous three votes that the health outcome evidence is appropriate.

With regard to any of the specific aspects of the evidence, again, we don't have a lot of individual information linking this to quality outcomes and that would be helpful but I guess it's not something that can be worked out after this process.

Any other questions about the impact and evidence, impact and performance gap for evidence? Are we good?

Karen Johnson: Good.

Anna Barrett: OK. Sounds like we're good. Good. OK. With regard to the scientific acceptability, the measure properties, reliability, and validity, everyone pretty much went at the – at the moderate level and I think they give in what we've been told that's consistent with the standards by which we're supposed to be measuring and consistent also with the information we have which is very, very good reliability and decent validity of the time and measure development but some potentials for improvement both in the source of the data and in maybe some on-going measures evaluation.

With regard to the usability, I think that the scale that's given, very similar to that use previously was thought by most people to be acceptable one person to be high and two people rated that moderate. I used by this – used in reporting, recommended by CMS. It's built with big holder input.

On-going evaluation of meaningfulness and the understandability of the measures, of course, would be useful. And the feasibility of the measure was also rated to be high, but one person in moderate by two, so it seems to be feasible measure overall.

And now, I was the person who had voted against endorsement based above on, you know, my thoughts about the measurement, the – about the assessment of performance and evidence to support against the quality measures and I think I was not using the standard criteria so I think that would make three votes in favor.

And there are just some notes that I think maybe useful to mention. There were only 7,000 stroke patients measured so something that's one comment are made, that was actually me, was that there maybe some bias with respect to those patients who are being assessed and some kind of a process measure might be useful in conjunction with this but I don't know if there is – if there's any opportunity for that.

And I just stated that I thought that it's adherent to a systemic professional standard was inherent in measure which is great, however, it would be nice to also see links to patient quality care and patient outcome in strengthening over the use of the measure.

Other comments or discussion? Everybody happy with the measure?

Karen Johnson: Anna, this is Karen at NQF. I do want to go back and maybe land this for a minute or two. One of the comments under scientific acceptability and that was the concern that I guess the reliability was based on the clinical vignette and just kind of – is that something that is a concern or is that – everybody is OK with that.

You know, is there any…

Jane Sullivan: This is Jane. That was my comment and I think it came up in the last measure or two with the fact that is these measures have been in used since the late '90s or early 2000s. It would seem like we could have information that is available beyond the development of the measure but the reliability of the measure in use.

Anna Barrett: Rob, you had mentioned that there was such data available but there's…

Rob Mullen: No, well the – no, we just have more data on the use of the vignette in –use of the vignette is on-going to essentially certify participants for our data collection that they have mastered the function communication measures. But there is – but it's still using the – using vignette.

Jane Sullivan: I guess my comment would be that – and I looked at the vignettes and the materials that were provided and the vignettes gave the information that the speech-language pathologist would use to assign a score. And so that's all well and good. But in real life, the clinician sees the patient and uses their skills to assign a score so I don't think – I don't think it's the same thing.

Mary Van de Kamp: This is Mary. Just, maybe, speak to that because that's a big part of our inter rater reliability discussion when you're trying to ensure that each therapist is using a scoring system that's comparable and, you know, because without a vignette that gives you an example of what that looks like without sort of setting a benchmark is to what a one, a two, a three, or four, five, a six, a seven is a vignette.

You could assume in your own mind that that might be a little bit better, a little bit worse than that. So it's really to set an expectation around the number relative to the performance that's consistent even though as we have discussions with clinician, part of the training, is there is – there is some dialogue around, you know, what this would look like being that other patients may not present exactly the same.

But that – if I'm understanding what you're asking, that's how the vignette is used even though the vignette is change, there is a common – there is a common functionality within that vignette that can be applied to the patient in the outcome assessment.

Jane Sullivan: No, that makes sense. I guess, to me, I was sort of seeing the vignette in the sense of, you know, if you have examiners in the clinical trial, you have a vignette or you have a video that everybody is trained on, but then with the real life individual in front of you, the clinician has to derive the information that – but they may or may not successfully derive in order to assign a score.

So I think it's – I think it's different. It's like training to be able to administer a test versus can you reliably administer the test. And I think those are potentially two different things. But, you know, if I'm the only that's bothered about that, then I'm willing to live with that.

Anna Barrett: Well, I mentioned, this is Anna, and I will say that I was affected by that also but it's not – it didn't seem to me to be out of standard with the number of other examiner or I should say therapist rated tools that are used in the rehabilitation setting which frequently have kind of a simulated reliability training through that process which is not followed by formal continued either inter rated or rater reliability or even, you know, things like test- retest reliability could – are not even part of that equation in the real world setting.

And so I didn't want to hold this measure to an unreal standard one that's – I haven't seen real lives well with other clinical tools.

Mary Van de Kamp: I agree with that. But I think if – I think that it's comparable to many of the outcomes measures that are used. But I do – I do – we do – we do retests and we – so we do build at an inter rater reliability within the outcome, you know, as we have an expectation. So that's not consistent across all measure, I mean, all providers.

But I understand what you're also saying if there's a risk without some of that being incorporated into the implementation.

Anna Barrett: I guess what I wanted to, this is Anna, is whether this is just the database that's being maintained within a professional organization and there's a lot of commitment from the members, is it possible to do such a setup, such a taskforce that's willing to do that kind of reliability evaluation at their institution but in multiple sites. It's then, you know, provide on-going information and that would be something useful as the – as the tool continues to be used.

Rob Mullen: This is Rob. Certainly that's absolutely feasible as, you know, you can well imagine the timeframe on getting that done is not…

Anna Barrett: Yes.

Rob Mullen: …not in the near-term at all, but time is only feasible.

Mary Van de Kamp: And it is happening. It's just to your point, I think, to do it in a more global way would be very helpful.

Karen Johnson: And, Rob, this is Karen, I'm just curious about this, but as I recall, you talked about before you allow providers to enter data into your system, they have to basically pass a test with an 80 percent?

Rob Mullen: Correct.

Karen Johnson: What happens if they don't pass?

Rob Mullen: They're instructed to go to the training manuals once again and we take a different version of the test.

Karen Johnson: OK. So they're not going to learn from their mistake necessarily of what they did wrong on that particular vignette. You're getting them another…

Rob Mullen: Well, we don't – we don't want them to retake the same test so we have them take a different version.

Anna Barrett: And they only have two chances there?

Rob Mullen: Yes.

Anna Barrett: So that's some data actually about the usability, right, of the tester the meaningful – I mean, I don't know. I wonder is how could – is it potential for – if a large proportion of the people who attempt to enter data are able to pass the reliability test on an initial attempt and that suggest that the items are understandable.

Rob Mullen: The past rate on the initial test is maybe something totally, you know, better than I do. But my guess is the pass rate to the first test is about 90 percent, would you say totally?

Jane Sullivan: Yes. I agree with that.

Anna Barrett: Anything else on that measure before we go on? I think I'm passing it on – I don't know if (Dr. Sheff) is with us.

Suzanne Theberge: Is Dr. (Sheff) with us today? OK. So give me just a second. I'll pull up the (Dr. Sheff's) measure which is the reading measure. No surprises on this one. It's pretty much the same measure as the other ones that we have discussed. And it looks like two – I'm just going to be very, very quick on this and see if there's anything in the comments that we shouldn't land on importance to measure report.

It looks like there was on yes and one no. It looks like about a quarter of the folks measured failed to make progress so that's pretty high percentage. And there's also disparities were demonstrated.

And then the comment that was made is that the developer envisions that a benefit of this measure is increasing the proportions of patients who make progress on this measure will stimulate clinicians to think about ways to improve care to increase this proportion.

So I think there was a question about that statement so, maybe that's something, Rob, that you would like to expand on a little bit. Were you thinking anything in particular on this one?

Rob Mullen: I'm not really totally sure I understand the comment because it falls to say and the same we said about almost – about the same measure that has a low percentage of success. Only public reporting and improvement plans are likely to achieve this. That's sort of precisely why I wish some of these measures to NQF such that there will be – it increase likely that they're used in public reporting such that they can be helpful to stimulate sort of the – sort of incentivize clinicians to think about how they can improve these percentages.

I mean, these measures are also part of PQRS and so it's our intent to get some into these public reporting systems precisely for that reason because we think that 25.4 percent of the patients failing to make progress on this particular measure is not acceptable. But whoever made the comment is absolutely right and that just coming up with the measure and if so, it wasn't going to help anything unless we can get it into public reporting to help increase the incentives for people to care about what their – what their performance rate is.

Jane Sullivan: Rob, this is Jane. That was my comment and I think – I think that's – I absolutely agree with you that that's a reason why we do measure things is to identify where we have performance gaps. And I guess a time goes back to a discussion we had a couple measures ago with regards to usability and my guess is that in fact this measure is probably already that the data from this measure, it's probably already being used to help people make those changes.

So it sort of the difference between implicit and explicit information I think.

Rob Mullen: Sure.

Jane Sullivan: I sort of assumed that these things are helpful because we're measuring them but I was sort of looking to connect the dots and, you know, what actually – what has the impact then. It's in fact these measures have been used for some time. One would expect that just exactly what you said might happen is happening.

Rob Mullen: Sure. I guess maybe we were a little bit too reluctant when we filled out this forms. I mean, because we have all sorts of anecdotal information about, you know, X skilled nursing facility in California, looked at their data, and decided to make changes in a clinical pathway or et cetera.

We have lots of that type of information but I guess I didn't – I didn't want to sort of – I didn't feel comfortable representing that this particular facility in Arizona or New Mexico or wherever, used the data to make changes. But I guess maybe I didn't – wasn't sure that was our place as the measure developer as opposed to the people actually using the data.

But if those are the types of examples that would be helpful, we can certainly provide some.

Jane Sullivan: Well, and this may just be my being enough as I'm reviewing this. I was – I guess I was – I was trying to apply what I was reading as the criteria, you know, literally. And so where I – where I may have stumbled was in what's implied and what's – you know, what's implicit, and what's explicit. And I don't – I don't see this problematic but I wasn't really clear what the bar was.

Anna Barrett: Rob, is that…

Rob Mullen: I share your lack of clarity.

Anna Barrett: Well, I guess, what's – again, this sounds like there's an opportunity. I know that this is very effortful to do. But if there are, you know, the X number of facilities that are participating and, again, if your network isn't committed which it sounds like it is to this as a long-term endeavor, one wonders if – just asking do you have any processes or, you know, clinical care plan changes, program changes occurred as a result of examining your quality data and if you can show that 60 percent said yes, then that's – begins to speak to that.

Rob Mullen: We can surely do that. Right now, we don't have any systematically collected data such as that as I mentioned we just have anecdotes, but that's certainly something we can undertake.

Karen Johnson: And, Rob, this is Karen. We talked earlier about the Section 3B, the usefulness and quality improvement. That might be a place where you can add some of that narrative.

Rob Mullen: OK. OK. Sure.

Karen Johnson: If that make sense to the committee members and that seems like a good place that you might want to fix that.

Rob, also, real quick, since it is a PQRS measure, do you have – I didn't know this, do you know how many speech-language pathologists are actually reporting on this measure – on these measures? Do you have a flavor of how many?

Rob Mullen: Very, very few and it's not that they are reporting other PQRS measures and specialties. I mean, these are – these eight measures are – that are before you today are the only eight speech-language pathology measures in PQRS.

And because they are not measures that can be reported under Medicare claim form, but have to be reported through a registry and we had answer and also service of PQRS registry and I'm pretty sure with only non-profit PQRS registry.

Firstly, all the reporting of these measures does come through us on our registry and the numbers are very, very small. So – I mean, it could be interpreted that people don't like the measures but I think more accurately, it's the SOPs are not pruned to be terribly interested in and motivated to participate in PQRS and mostly because most of them aren't eligible to do so.

Karen Johnson: OK.

Rob Mullen: But I would say that there are the same growing pant (logs), probably not appropriate to characterize those growing pants. But it is certainly complicated in PQRS which is mostly process measures that when it comes to encountering outcomes measures such as these, it starts to create all sort of problems that aren't there with the decisions about processed measures.

Karen Johnson: OK. Let's go on to scientific acceptability and feasibility. I don't think there was different comments on this measure compared to the other measures although maybe a measure to back. Somebody did mention, if I have the term right, the positive – I'm sorry, I don't remember the term, is the idea that people may be – might be inclined to go ahead and score higher at discharge than they would otherwise, or then it's narrative.

Do you know what I'm talking about, Rob? It was…

Rob Mullen: I do.

Karen Johnson: Can you just talk about it?

Rob Mullen: Yes. We are absolutely sure that concern, we – I mean, this is not in a timeframe that will be able to affect NQF's deliberations about this. But the two areas on which we're working because we realize that absolutely is a concern, yes, and absolutely it affects the validity of these measures is that – and we – I was talking earlier about we really want to get these measures use in public reporting such that people will be motivated to think about how they can increase the proportion to make progress.

But with that comes the structure increase bias and people being motivated to score their patients as high as they think they can. So it's absolutely a threat that we're concerned about. You know, over the medium, it's a long-term the two things we're working on are – one is developing some sort of audit mechanism to be able to make those determinations about whether the score is reported a relative score.

And the second area that we're looking at is to supplement these measures with corresponding patient reported measures, but that's obviously really tricky in the area of communication disservice because the whole notion of patient reporting is in itself an act of communication and we're talking about communicably impaired people, so that's, I think perhaps more tricky in our discipline and perhaps for most others – certainly, for most others.

But – so those are sort of medium to long-term attempts to deal with this, but you're right to be concerned about it and we're concerned about it.

Karen Johnson: OK. Thank you, Rob. In the interest of time, boy, just two hours is flying by for us here. I think let's go on to the memory measure and Jordan, this is yours. And if you would, just maybe focus on things that may be different, if anything, on the memory measure than all the other measures that we've already discussed.

Jordan Eisenstock: Sure. You know, this is obviously very familiar so I think we can run through it pretty quickly. Numerator and denominators are very similar to the other measures for FC and this time memory. And this is an outcome measure of the committee members who rated this one.

There was again pretty much agreement on every measure. In part one with impact, two rated it high and then with performance gap, I think, because of some interpretations that have been clarified this afternoon, there was one high and one moderate. And basically the comments were the same though.

As far as evidence three Ss and this was a health outcome. One of the things that I was just pointing out which is related and unrelated, I suspect at the same time is something put under (1C1) with regard to the hours achievement. I just found it very interesting that for those who received five plus hours of treatment, the number was lower than those who received four.

And I just wasn't sure exactly how to interpret that since the patients who were listed in this particular section all started at the level five. And you just expect, the more time the better they do. And so – that was just on my mind but didn't obviously affect my scoring on this one nor anybody else who submitted.

As far as scientific acceptability, again, everybody was in the moderate level which as others has stated as, in line with that, we'd expect for this kind of measure. Usability, again, there was a split but most people agreed that high or moderate was appropriate, same thing with feasibility and everyone agreed that this was suitable for endorsement. So I don't know if that was quick enough or if we needed to add some comments along the way, so.

Anna Barrett: Impressive.

Karen Johnson: Thank you, Jordan. Does anybody have anything they want to add on the memory measure?

Anna Barrett: I just have a question about harmonization with other measures. I know that's some things that may not – might follow this kind of a deliberation but would need to be assessed at some other point because, for example, it struck me that there are other measures that might apply the same patients and also assess the same (inaudible) like functional independence measure, for example, has a section on memory.

Karen Johnson: Right. We will be looking at that and giving you some more information on things to think about in terms to related or competing measures. Like you said, you need a process. What we would do in the in person meeting is goes through each measure and you'll vote on each criteria.

And only after a measure has been recommended by the committee as suitable for endorsement will we go on to talk about the control related or competing issues. So what that's doing there is just making sure that you don't – we don't spend a lot of time talking about related and competing and then find out that you didn't want to recommend it anyway.

So, yes, you're right and we will be looking into and getting back to you within the next – well, before the meeting – before I enforce the meeting which is right around the corner.

OK. Anna, that leaves you, I think, with the attention measure. So, again, if you can raised Jordan and see if you can – if you can, in fact…

Anna Barrett: OK. Let's see. With respect to importance – impact to performance gap, we were a little bit more spread for this measure with one person rating this measure as potentially low impact. I think that might have been me though with regard to the demonstration of impact.

So I – we rated it moderate, we would be kind of at the moderate and high and performance gap similarly, it's moderate. Other areas in which this might – this measure might stand out from others, again, there was a mention of test-retest reliability with respect to measure properties which we've talked about already to some extent but it's important in assessing improvement.

With respect to usability, again, the plan for how the data might be leveraged to identify risk conditions might inter leave with stratification data. It could be made more available. And, again, just kind of making things more explicit to us in terms of the overall program that this measure supports.

I don't see any others potentials major differences between these points that have been brought up on this measure and those that have been discussed previously.

Are there any other comments or what other would like to bring up, issue?

I did not recommend endorsement initially because there's a holding that the measure to a higher standard for performance and I have learned this maybe the most appropriate so then we have three people from the initial group that would like to – that feel that the measure is suitable for endorsement.

Karen Johnson: OK. Great. If there's no other comments about this measure, we will – we'll go on to our last measure. Thank you guys for being so understanding about dealing all of these measures. Normally, we would never ask to review 10 measures in this two-hour timeframe, but since eight of them were very similar, we though that that would be OK to do.

At this point, has anyone from the Joint Commission joined us?

(Karen Cole): Yes. This is (Karen Cole). This is the Joint Commission.

Karen Johnson: Hi, (Karen).

(Karen Cole): Hi.

Karen Johnson: OK. (Tina), you have been assigned as the lead discussant for the measure 0441 from Joint Commission. Would you like to roll with that measure?

(Tina Kiernan): Sure. Let me apologize upfront. I'm working from a hotel room and I keep getting knocked off the Internet so it's loading my screen for about the sixth times so I can't actually see the measure and it won't let me follow up to your point. But I do have – I can give a description of the measure from part of the material that I brought with me to this conference.

So this is the stroke measure 10, assess for rehabilitation. This measure captures the portion of ischemic or hemorrhagic stroke patients assessed for or who received rehabilitation services during a hospital stay. And this is part of – this measure is part of eight nationally implementing measures that address certain care.

And I can't – I can't see the actual form. It's not loading. So I might let…

Karen Johnson: OK.

(Tina Kiernan): …somebody else read it for me. Sorry about that.

Karen Johnson: No, we can help you with that. Let's see if I find it in my list here.

(Tina Kiernan): OK.

Karen Johnson: I'm nearsighted. I can't read it on the screen that we have then for another, so. So in terms of importance to measure and report – so, (Tina), let me make sure I understand, you want me do this kind of go through what the evaluators said about this measure, is that – since you don't – you can't look for…

(Tina Kiernan): Yes. Unfortunately, I got the e-mail either yesterday or today with all of our scoring on it but it's – I can't pull it off and now I can't pull up your slides either.

Karen Johnson: OK.

(Tina Kiernan): It (loads) in but it won't come up.

Karen Johnson: OK. On this one, it looks like there's three people evaluated this measure and all three rated it high on impact and high on performance gap. And what I'll do is ask the folks who rated the measure if there was anything – there's a lot of comments actually on this measure which is – unlike the other ones.

Was there anything of concern about either impact or performance gap for this measure?

OK. This – I think this is one that we talked about. We used this measure, Joint Commission folks, in case you didn't know it, the measure that we looked at in our tutorial and you note that you have a really high percentage of people meeting this measure, 97.4 percent as of third quarter 2011.

So I think one of the questions that the committee might be interested in hearing about is, you know, is this tapped out? And I know you discussed that from in your submission but maybe you'd like to speak to that just a little bit. Do you feel it's tapped out or is this a function of it being reported by just a percentage of divisions in PQRS?

(Karen Cole): Hi. Well, I just joined the call, Suzanne. This is (Karen) as the Joint Commission speaking so I didn't hear what the other measures and (inaudible) had to say. So for getting tapped out, we would not consider this measure tapped out.

The reason that this measure is included in the set is because – and we know that probably two-thirds of stroke patients could benefit from some form of rehabilitation service whether it's cognitive or PTOT speech whatever and less than a third actually received rehabilitation services and we know that from the evidence.

And we think that this is a very important measure. The focus here is on assessment being that if the patient is never assessed then the rehabilitation need is never even considered with the likelihood that they will actually receive it if indeed it is indicated.

And, you know, what we have found through from (inaudible) operation realization of the measure is basically that, you know, we focused on the assessment being done by a qualified member of the rehabilitation team whether that the physiatrist, PTOT speech, neuro-psychologist, it's felt by the attending neurologist or attending physician that the patient does not need rehabilitation services, we expect that to be documented explicitly in the medical record and we would take that as an assessment that indeed they were assessed but it's not indicated for the patient.

However, what we find actually through using this measure is, oftentimes, what hospitals do want to do is have basically nursing assessments used to meet this measure which we do not allow.

And I think that we're holding it to a higher bar to keep it where we are actually trying to push to have more patients actually receive rehabilitation services by having a completed rehab assessment. So we feel that it's not tapped out and that it's a very important measure to keep in the set.

Karen Johnson: OK. Thank you. Does any of the committee members have any other questions about impact or gap?

OK. Evidence, it looks like all three folks rated this as having high quantity, for quality, one high, two moderate; and for consistency, two high and one moderate. So were there any concerns at all about evidence? Is it fair to say that those of you who evaluated that you feel confident that there is a strong evidence-based behind this measure?

Rob Mullen: Yes.

Female: Yes.

Female: Yes.

Karen Johnson: OK.

Anna Barrett: Yes.

Karen Johnson: Great. Let's go on to scientific acceptability. It looks like three folks rated it as high for reliability and two high for validity and one moderate for validity. And so on this one, the reliability information provided was data element reliability and that would be actually according to our guidance eligible for a moderating rating of reliability.

And the numbers, I think, the percent agreement was quite high. Half of the statistics were not given the percent agreement with high. And then for validity, the developers described several different things if they consider as validity.

We would like to point out that when they did their reliability testing, they had the inter rater reliability, but their second rater, if you will, if I understand this correct, and (Karen), you can jump in and correct me if I'm incorrect on this. But the second rater on that was a Joint Commission obstructer whom you consider to be the sort of head of source.

And so that kind of reliability testing would also count by NQF standards as data element validity testing as well.

Does – was there any other concerns about either reliability or validity testing for this measure?

I do have maybe one quick question. I know we're getting close on time, but, (Karen), I haven't see in the – in the numerator, do you spell out what counts is rehab or is it just any rehab or how does that work? I didn't quite understand that part.

(Karen Cole): OK. Suzanne, well, actually, before inclusion on the numerator, there's only one data element used for this measure and it's assessed for rehabilitation services. And the details of what counts is the rehabilitation assessment is provided in the data element definition that's found in the alphabetical data dictionary and the link to that particular source of information was provided in the submission.

But basically, we provide a variety of notes for obstruction and then we provide inclusion guidelines for obstruction which means – inclusion guidelines mean what counts is the rehab assessment. And we tell them, an assessment or counts was done by a member of the rehabilitation team and the members of the rehabilitation team include a physiatrist, the neuro-psychologist, the physical therapist, the occupational therapist, the speech and language pathologist.

So we are specifying for them who is a qualified rehab provider and who is competent to complete the assessment. We also provide and inclusion guidelines patient received rehabilitation services from a member of the rehabilitation team. Of course, that would be anytime during the stay.

So we assumed that if the patient have, let's say, a range of motion exercises and then find by a physical therapist, then obviously, there had to be an assessment done in order to get their therapist to the bedside.

And exclusion guidelines for obstruction, exclusion guidelines are what does not count, what is not acceptable as a rehab assessment. If there was a request or an order for an inpatient rehab consult and the consult was not performed, it was not completed, that does not count.

So you may have an order for PT in the chart but if it wasn't executed, you would not be instructed to select out through the data element and you will not be included in the numerator.

So the level of detail is in the data element definition for the short answer.

Karen Johnson: OK. Thank you, (Karen). I think for that, we might actually ask you – Suzanne, are the submission still open for Joint Commission?

Suzanne Theberge: I have to check but we can reopen them.

Karen Johnson: Yes. We might go ahead and reopen this measure as we discuss about some of the measures yesterday because we would like to have those specifics actually in the form. If we…

(Karen Cole): Suzanne, we had – we provided the link to the specification and the submission. And if you reopen all those measures for all those data element definitions, that would be quite a large amount of information to extract an importance to your submission form.

The link is provided in the submission and it's easily accessible. They also had that called dictionary is the – it's PDF 1B alphabetical data dictionary. Open that PDF just like we used your PDF through the submissions to view and you can go write through alphabetically, find the data element definition by page, and review it that way.

Karen Johnson: OK. And this is Karen again and I'm just asking for that simply because we have had a couple of committee members asked for that and actually have trouble finding your data element. So just so you know that we have had committee members who had trouble finding those data elements.

But we're almost out of time so let's go ahead and go on to usability. It looks like all three evaluators found this measure to be high in terms of rating for usability. So does any of the workgroup members have any concerns or anything that they would like to discuss about usability?

OK. And then feasibility, looks like members noted two had high and one had moderate in terms of feasibility. And I'm not seeing exactly why the one person would just have moderate as opposed to high. Does anybody have anything to discuss in terms of feasibility for this measure?

OK. You guys are back to being a quiet group again.

OK. So in terms of the preliminary assessment, all three evaluators found that they thought that this measure was suitable for endorsement. And (Tina), did you have anything – I know you don't – you can't probably still see the measure upload, but was there anything that you remember that you wanted to bring up with this measure?

(Tina Kiernan): No, and you're right. They boot me up again so I'm not even – it's trying to log me back in. You know, the only comment that I had and it doesn't – it relates to the measure but it doesn't in fact – even though patients are advised for rehab, if they don't have insurance from any state, then they go home with their disability.

But to me, that was a disparity that this core measure doesn't speak to but related to it and that was my only comment is that we just see a lot of patients who are (stuck) and have a need for rehab in the hospital but they go home with no rehab because they have no insurance and no funding.

So – and we are – it does really relate, it does but it doesn't. That was my only comment.

Karen Johnson: OK. Thank you. So I'm paying attention to the clock and I want to hand it back over to Suzanne to wrap up our (inaudible).

Suzanne Theberge: Great. Thank you, everybody. Now, it sounds like some folks may want to change their ratings and I know some of you didn't get a chance to complete the measure evaluation survey. So if you would like to do that tomorrow or this evening, we're going to be recompiling all of the survey results and putting them all into a document, all the measures on Friday.

So if you want to go back and change your vote, you can do so. Just mark your name with a two after it so we know you've updated the vote. If you haven't completed the survey, we really appreciate your votes, if you could do that because these will go into our final report. So it's helpful if we can have a large number of the workgroup's voting.

At this time, we would like to open the line for public comment. (Natalie), can you open the line and see if anybody – any members of the public have any comment?

Operator: At this time, ladies and gentlemen, if you would like to ask a question, please press star, then the number 1 on your telephone keypad. We'll pause for just a moment to compile the Q&A roster.

At this time, you have no question.

Suzanne Theberge: Great. Thank you. OK. So next steps for the committee members, we, as mentioned in my e-mail earlier, we are allowing some of the developers to make changes to the measures and we'll be sending out updated measure form and summaries of those changes over the next few days. We're hoping to have all those by Tuesday or Wednesday of next week.

So your next steps are to review the remaining measures and we all look forward to seeing you in just two weeks here in D.C. And if you have any questions, don't hesitate to call or e-mail.

And I think that's all for this evening. Thanks very much for your time and have a good night.

Rob Mullen: Thank you. You, too.

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

END