NATIONAL QUALITY FORUM

Moderator: Family-Centered Care Person and
November 13, 2015
12:00 p.m. ET

OPERATOR: This is Conference #: 79916599.

Welcome everyone, the webcast is about to begin. Please note today's call is being recorded, please stand by.

Suzanne Theberge: Good afternoon everyone and thank you or good morning depending on where you're dialing in from. And thank you for joining Person- and Family-Centered Care Off-Cycle Review webinar.

This is Suzanne Theberge, I'm the Senior Project Manager on the team. And before we get started with the rest of the introductions I'm going dive right into our usual housekeeping announcements. The first of which is, if you are dialing in on the phone and also connecting on the webinar, please be sure to turn off the sound on your computer so we don't get feedback on the line. I'm hearing a little bit of an echo, so I think someone's got the sound on their computer on.

Also I want to note that you are a committee member or a developer and you wish to speak, you will need to also dial in to the phone, we can't accept audio input from the webinar so please do dial in. And we also like to ask folks to keep their phones on mute if you're not actually talking just to reduce the interference on the line.

Now I'd like to introduce the Project team with the next slide. Oops, sorry next slide. As I said this is Suzanne Theberge, we've also got Sarah Sampsel, an NQF consultant on the line and Kaitlynn Robinson-Ector, our Project

Analyst. And now I will turn it over to Kaitlynn to do the roll call of our committee members.

Kaitlynn Robinson-Ector: Great, hi everyone I'm just going to go through the committee procedures here. So first we have Lee Partridge?

Lee Partridge: I'm here.

Kaitlynn Robinson-Ector: Great, thank you. Then Chris Stille?

Christopher Stille: I'm here, good morning.

Kaitlynn Robinson-Ector: We have James Merlino?

James Merlino: I'm here.

Kaitlynn Robinson-Ector: Great, thank you. We have Katherine Bevans? OK, then we have (Samuel Bierner) ? (Rebecca Bradley)?

Rebecca Bradley: I'm here.

Kaitlynn Robinson-Ector: Thank you. (David Silva)?

(David Silva): Hi.

Kaitlynn Robinson-Ector: OK. (Sharon Cross)? (Dawn Dowding)?

(Dawn Dowding): I'm here but I'm going to have to leave at half 12:00.

Kaitlynn Robinson-Ector: Thank you. (Sherry Kaplan)? Brian Limberg?

Brian Limberg: Here. And you're not coming across as loudly as the first speaker just so you know.

Kaitlynn Robinson-Ector: Thank you. Next we have (Sherry Lobe)? (Anne Monroe)? (Lisa Morris)? (Elizabeth North)? (Esy Newerth)? (Leonard Parisi)?

(Leonard Parisi): I'm here.

Kaitlynn Robinson-Ector: Thank you. (Deborah Saliba)?

(Deborah Saliba): I'm here.

Kaitlynn Robinson-Ector: Thank you. (Peter Thomas)?

(Peter Thomas): Present.

Kaitlynn Robinson-Ector: Great. And (Karen Vansel)? OK. Thank you, and now I'll turn it over to Sarah for the rest of the presentation.

Sarah Sampsel: Great, thanks Kaitlynn. I'll check for Katherine Bevans again? OK. So I see Katherine on the web, but I didn't see here called in. But before we move on first, you know, I'd like to welcome everybody and appreciate your participation this Friday afternoon. And, as we go through the agenda make some notes and some comments regarding where we are in this project overall, as well as some past notes unto why some of these measures are on this off-cycle review and what we're looking at coming up in 2016.

But I'd like to also acknowledge that a number of folks on the standing committee had reached the end of their first two-year term. And we did have a couple of folks decide that their schedule just didn't permit them to contribute to participate, or in one case we have a member who really didn't feel that she was contributing as much as she really would have wanted to which NQF is evaluating how we insure that the consumers and insuring that they know that they're valued on these consumers.

So, just want to acknowledge the contributions of (Carol Levine) and (Esy Newerth). I don't think either of them are on the call today but both decided to step off the committee and pursue other opportunities I think in continuing to improve quality.

I'd also now acknowledge staff-wise that as many of you may realize, Mitra has left NQF. So, we have that (hole) in the project management but Mitra was really a great assistant on this project over the past year. So, you won't be seeing her name affiliated with the project anymore, and for any folks who are reaching out to her directly, just please know that we ask that you use PFCC

e-mail box, ensure that all of us on the project team receive your e-mail and can follow up as appropriate.

With that, today we have primary goals, one is to evaluate the risk adjustment methodology for measures that we reviewed during phase two, and those are measures 2643 and 2653. And we have (Collette) and Gunnar from Minnesota Community Measurement who will provide some statements and a little bit of overview on what they have done and updating those measures in a little bit. And then, we have a new measure which is the patient activation measure and (Judy Hibbard) and some folks from Insignia will be presenting on that measure. And that will be a full measure evaluation.

A couple of just other comments that will we'll get into further is we do have a need, a total of 12 members of the committee to vote to reach quorum, we're not sure we'll be reaching quorum today but we'll be monitoring that before, you know, as we move forward and get to the voting portion. And we do know we have a couple of members who requested to – the ability to read the transcript and the opportunity to vote off-line or via SurveyMonkey after the call. So, for any folks having technology issues or anything else you will have an opportunity to vote after the call.

So with that if we can go to the next slide. And then I'll just comment, it look (Lisa Morris) has joined and well as Katherine Bevans. (Lisa) are you there?

(Lisa Morris): I am here, can you hear me?

Sarah Sampsel: We can hear you now, yes.

(Lisa Morris): OK …

Katherine Bevans: And this is Katherine, can you hear me?

Sarah Sampsel: We can hear you, Katherine.

Katherine Bevans: OK, thank you.

Sarah Sampsel: And do we have any other standing committee members that have joined that missed roll call or we didn't get to during roll call?

(Lisa Morris): Can I just say that when I called in initially the operator could not hear me and disconnected me, so there maybe some technical issues.

Sarah Sampsel: OK.

(Sherry Kaplan): And this is (Sherry Kaplan), I wasn't sure because my computer was on mute and I wasn't sure you could hear me, or at least I think it was on mute.

Sarah Sampsel: OK, great, thanks (Sherry). OK, so with that our goal today and starting with the evaluation of the risk adjustment methodology. In this section of the call, I'll be providing a brief overview of risk adjustment methodology guidance as well as to tell you why these measures are back before you so early.

I will then turn the call over to (Collette) and Gunnar who will provide an update on what they've done with these measures and the risk adjustment strategy. And then the call will be open to the Person- and Family-Centered Care Standing Committee for any questions and discussion you may want to ask (Collette) and Gunnar. And then finally we'll be asking you to determine if this associated information is acceptable to remove the condition of endorsement.

(Off-mike)

Sarah Sampsel: So, if we could have everybody mute their line.

(Off-mike)

Sarah Sampsel: OK. So, if we can go to the next slide. With the Minnesota community measures, and for those of you who were participating during phase two of this project, which is the phase that we were or just finally completing right now and we meet earlier this year in-person, you might recall that both of these measures came forward to you, and during the in-person meeting, consensus was not reached on all of the critical factors of the measure and specifically around scientific acceptability.

There was additional information provided during that public comment period as well as you all reviewed the public comments, and there was strong public, you know, kind of interest in moving these measures forward because of their importance in the measure industry and typically in assessing phase and functional status post the knee and lumbar surgery and, kind of getting engage of where patients are.

What you might also recall is that even through these measures were (fully) and some testing was done on risk adjustment strategy, at that very specific time, Minnesota Community Measurement was also waiting for some additional data to come in to finalize the risk adjustment strategy.

And, you know, as part of that they had some – some idea on which way the risk adjustment methodology would go, and they have a technical expert panel that was providing input to them. But they wanted the benefit of a full year of data collection to finalize that strategy. In the additional review of information, I think in looking forth of the conversation at the in-person meeting and the post-comment call, you all when you voted decided to go ahead and fully recommended endorsement of the measure but as the – of the measures, but as the measure went through the NQF process, we really had to assess effective, kind of put the breaks on a little bit.

And while the measures are considered endorsed, we put conditions of endorsement on the measure. And have been in contact with Minnesota community measurement and specially put (inaudible) to talk about, you know, when would this data be available and would we – would they be able to provide an update and a final risk adjust methodology within one yea of endorsement? Timing wise they were able to meet the deadlines for this off-cycle review. So we really asked and appreciated them gearing up their efforts that they would be able to finalize the risk adjustment strategy.

We're asking you to consider as (Collette) and Gunnar provide an overview of what they've done, would be the correct criteria for the risk adjustment methodology. And those are on slide number seven. For outcome measures and other measures were indicated, these are outcome measures, we're looking for a risk – an evidence-based risk adjustment strategy based on patient factor

that can influence the measure outcome and present at the start of care, with demonstrated discrimination, calibration, or rational adjustment. In this case, so we are looking for an evidence-based strategy.

I also – just a little of the context that risk adjustment is considered a threat to validity, and therefore in this consideration as well as in you path consideration, what we're asking you to do is look at and ensure that the developer provided an analysis of anything that could present a treat to the validity. And, how it really impacts the measure overall, the specifications and then more specifically the validity of measurement over time.

So with that, next slide. The next slides, just the overviews and reminders of what the measures are, and for that I'm going to go ahead and turn it over to (Collette) and Gunnar and ask you to, you know, remind the committee about the measures and then what you've done since they meet earlier this year.

(Collette): Great, thank you very much. This is (Collette), can you hear me OK?

Sarah Sampsel: We can hear you, thank you.

(Collette): Fabulous. I'm a measure developer with Minnesota Community Measurement. My co-worker – as Sarah had referenced is actually on her honeymoon. But with me is my colleague Gunnar Nelson, a health economist with MNCM who facilitate the Risk Adjustment Committee Methodology and Process. We appreciate the opportunity for an ad hoc review to discuss risk adjustment. The two measures for risk adjustment review today are 2643, Average change in functional status following a lumbar surgery, lumbar fusion which is a patient reported outcome based measure using the Oxford Knee Support tool, and we are assessing those patients, we're looking at the change between preoperative and postoperative function at one year which is defined as 9 to 15 months postoperatively.

The second measure is the average change in functional status following total knee replacement surgery, another pro-based measure and this one uses the Oswestry Disability Index tool. During the committee's discussion and recommendation for endorsement earlier this year, we have presented a plan for risk adjustment that included the method and variables for potential

inclusion but did not yet have enough data to test the model. These are newer measures both recently completing multi-faced testing and the first wave of risk adjustment testing this fall.

Although these procedurally-based measures reflect different patient populations, they're very similar in measure construct and rely on patient reported outcome tools to measure functional status. The patient functional status is accessible preoperatively and one year after the procedure. And the absolute change between preop and postop functional status score is calculated then averaged to computer practice level, average change and functional status. Minnesota community measurement publicly reports data that is risk adjusting using an actual to expected methodology, which allows the unadjusted rate to be preserved and displayed but also displays unexpected result for comparison.

As part of the measure development process, the workgroup recommends potential risk adjustment variables for the measure on measures, and (sort) of a committee measurement collects the standardized set of demographic variables that can be considered and tested for inclusion in the model. The workgroup recommends additional clinical variables that are important to consider for the measured population. These variables are included in the public comment process prior to pilot testing for feasibility and then included in the data collection to test the model.

The standard demographic variables are gender, age, (zip), race ethnicity, country of origin, primary language and insurance product as a proxy for socioeconomic status. Variables recommended by the workgroup included for total knee replacement, the initial preoperative functional status as measured by Oxford knee, BMI, comorbidity of diabetes and tobacco status. For lumbar spine fusion surgery, it was the initial preoperative functional status as measured by the ODI, the BMI, the clinical condition or reason for procedure, the history of prior back surgery and tobacco status.

Potential risk adjustment factors were empirically evaluated for the appropriate inclusion in the model is on the ANOVA, F-Test to test the likelihood that a category is unique with the P value of less than 0.01.

Variables that were tested and failed to meet the significant test included for total knee, these again these variable failed, age, tobacco status, gender, and diabetes. So they were found to not have an empirical. For lumbar surgery the variables that failed to meet the F-Test were age, tobacco status, gender, history of prior back surgery and clinical condition reason for procedure.

The variables that then were tested for performance in the risk adjustment model included insurance product, initial preoperative functional status and BMI and this – actually the same variables for both measures.

There was a strong association with an F-Test P value of less than 0.3. So these variables were run through our model. The preoperative functional status was the only variable found to demonstrate a strong consistent empirical association with the outcome that's measured.

It's worth mentioning again the construction of these measure is different, it's a continuous variable reflecting the average change and function following the procedure. Patients with a poor preoperative function – physical function have the greatest potential for absolute change, model the measure the average change need to take that into account. It's easy to then ask about the strata of patients who have higher preop function while they're having surgery.

However, this is not a measure for appropriateness. It's important to remember that the decision to proceed with surgery is a multifactorial one and functional status is only one part. During the development of the knee measure the workgroup gave the following example. You could take five patients all who have the same (Inaudible) scale, the same Tibia femoral angle, the same pain score, same Oxford knee score, but it's only appropriate to proceed with a TKR for only three of those five patients.

It's hypothesized by the measure development workgroup or the risk adjustment committee performed as we had expected, which is why we continuously evaluate the variables over the next few years as we accumulate more data. These are relatively new measures and have a lower volume of

eligible patients as compared to measures of outcome, for example the diabetes population. So it will take an iterative process as the model evolves.

It's also important to keep in mind that our methodology for public reporting, of outcomes data on the website always displays both the expected and the actual results but practice. And with that I will open it up for discussion and questions of Gunnar and myself. Thank you.

Sarah Sampsel: Thanks, (Collette), and I'm going to ask Jim, Chris, and Lee if you could (lead) the discussion at this point.

Female: Sure. Yes, of course. Why don't we just open it up to all of our members? (Sherry)? Are you there (Sherry)?

(Sherry Kaplan): Yes, this is (Sherry), my thing was on mute. I have a couple of quick questions, one is, if you start, you know, regression means and issue when you start with extreme value. So if you start with people with functional statuses very low, the measure associated with that end is tail of the score is that you then have problems with regression mean. And, the attribution, the causal attribution back to quality has to, you know, hypothesis-driven.

So if the hypothesis is, you would just for the baseline think because – what is the hypothesis with respect to adjusting for the baseline state, vis-à-vis equality.

Gunnar Nelson: As soon as we got right here – sorry, the operator cut in right in the middle of your question. So I caught the very end of it. It was the worst timing possible for the operator to cut in.

(Sherry Kaplan): OK. So, the question is, you know, the business of the regression of (mean) and you can score low enough, you know, you're going to have the movement at the low end of – so, the question is, what – for quality with respect to …

(Crosstalk)

Female: Operator, can you mute that line.

Operator: The operators are isolating that now.

Female: Thank you.

(Sherry Kaplan): Are we good? I think I'm now meant to ask the question.

Gunnar Nelson: Maybe now that's OK, maybe you're being told.

(Sherry Kaplan): OK so one more time. My concern is regression to mean and attribution to quality when you – I'm wondering about what the issues is with respect to adjustment for baseline state with respect to quality. And my concern is only, you know, you're going to get some movement for those extreme values back to the norm, at the low-end and at the high-end there's not going to be much opportunity for improvement with respect to quality and surgery. So, were you arguing that that's an adjuster for – well, help us out with the hypothesis with respect to quality adjustment.

Gunnar Nelson: If I understand and the operator doesn't cut in again, if you have a high-functional status, so reasonable – the (inaudible) then you're already starting a 30, you can only go so far. It's still a legitimate surgery that we legitimately need to measure. And so – and actually the risk adjustment if I understand is, will help the concern that everyone kind of goes to the knee. We were accounting for the fact that it's not the provider's fault that they're starting on a patient who is already higher functional status, it's still – being able to compare those across the functional ability. Is that what you're asking?

(Sherry Kaplan): Not exactly but …

Gunnar Nelson: OK, OK.

(Sherry Kaplan): So say you got two hospitals and some of the surgeons at one hospital are operating on people who have really high functional status, and as you say the movement there can only be so far, you know, one would wonder if that's a good measure of the success of the surgery if there they're already doing very well versus something like pain or other kinds of symptom. Then on the lower end you're going to end up with regression of the mean problems. So, if you're getting a change score that takes into account, you know, the

difference, if the dependent variables, the change score, what additional are you gaining – I mean I guess the statistical principles but what additional substantive principles are you gaining by including the baseline scores?

(Crosstalk)

(Collette): This is (Collette), I, you know, I don't know if I completely have the answer to your question, you know, the risk adjustment committee talked about when you have a measure that's looking at additional functional status of patients undergoing the procedure. That initial functional status is going to be different by patient and again it's not the only determining factor to decide to proceed toward surgery. But, without having that as a risk adjustment, you have different side. So, it's counterintuitive, sometimes just think about risk adjustment with that in mind but it's necessary for the floor effect of the change score to warrant this approach.

Gunnar Nelson: And also – this is Gunnar – keep in mind that since we use an indirect or actual to expect this methodology, the providers who are operating on the – any of the disabled patients are going to have a larger score if the value were – is different. So they're going to come up with a 20 as oppose to the people who are doing, you know, work at the sports center are going to come up with a five, then we evaluate that five might be pretty good and the 20 might not be all that impressive. So we keep – the whole point is we keep that raw score there. You'll see the bigger average change score on these providers who specialize in a more disabled patient.

(Collette): This is (Collette). I just wanted to add another point to Gunnar's excellent summary. We've had a philosophy here at Minnesota community measurement. We've been reporting outcome rights for many years. Prior to a push towards risk adjustment, and our committees and board have gone through a really thoughtful process and they really do value the actual to expected, you know, the display of both rates. So we're not trying to hide or cover something up.

And Gunnar has told me in the past if you keep risk adjusting long enough, you'll risk adjust away everything. So, I think it consider that philosophy of, we will always be displaying both rates.

Sarah Sampsel: OK, other questions from our members?

Brian Limberg: This is Brian Limberg. This could be a little off target here but the last question related to something that I have been trying to figure out. And that is, the comment was included about. And – if someone could just remind how as we evaluate this area, particularly the back surgery for example. How – what are we doing the along the lines of pain as oppose to dysfunctional status, because there is some evidence that trying to address back pain is one of the drivers for surgery. And knowing whether – address that as well as functional status. So I just maybe missing where we're getting that piece of the puzzle in this area of back surgery.

(Collette): This is (Collette), may I respond?

Sarah Sampsel: Yes.

(Collette): So, part of the component of both of the tools for functional status do have a heavy focus on pain and your tolerance in terms of different activities of function and their relation to pain. But I wanted to share with the committee for the spine infusion population, we also have separate measures of back pain and leg pain that we are calculating based on a 1 to 10 pain scale, that's not – this particular measure we have not brought those measures yet forward for endorsement but we actually have a suite of measures for both of these population of patient that also includes health-related quality of like as well. So we're trying to capture all three, it's just not all in the same measure.

Brian Limberg: That is exactly what I was looking to hear, thank you.

(Collette): You're welcome.

Sarah Sampsel: Further questions?

(Peter Thomas): I have a point of clarification, are we commenting on both of these measures now?

Sarah Sampsel: Yes.

(Peter Thomas): OK. So on the first one the spinal Lumbar spine infusion surgery. I read some comments on page 9 of the materials that were distributed. I'm just tying to get a sense for exactly who is saying these. I think it was the pre-evaluation comments. It said there is no risk adjustment, small sample size, no risk adjustment, at least that was the opinion of, I guess one the reviewers, at least two, but no risk adjustment stated here. And I thought that that was our main goal is to figure out, is the risk adjustment of this measure, you know, what is the methodology, is it appropriate and if not, is it – is there an explanation as to why it's not necessary? Isn't that correct?

Sarah Sampsel: So, this is (Peter)?

(Peter Thomas): Yes.

Sarah Sampsel: (Peter), this is Sarah. So, yes, we circulated the original materials that you looked at last January. But then what (Collette) and (Jess) and Gunnar presented would be on a Word Document that finalizes that risk adjustment strategy. So, you're looking at kind of past information that we did not update for this call.

(Peter Thomas): OK. So that just confused me as to exactly what the story is – was with that, and I appreciate you clarifying that. Can – this might be – again, might expose my ignorance of risk adjustment, but if it's a patient reported measure, does it really matter what the risk adjustment methodology is? I mean if I'm a individual who has relatively low functional status but by – say my knee replacement, you know, I'm amazed with how well I can now do, it maybe less – I maybe less functional than many other people that may have the same interventions. But, if I'm reporting the outcome and it's a major improvement, why isn't really need to be risk adjusted I guess is my question. Am I misunderstanding this fundamentally?

Christopher Stille: Yes, (Peter), this is Chris, I agree with you and I think that's – when they said they were going to report both unadjusted and adjusted numbers I think that's really is a great answer, because sometimes it matters and sometimes it doesn't.

Gunnar Nelson: And this is Gunner, to clarify we report the unadjusted and the expected value – essentially if everybody else in the community have the same patient you had. So it's not really adjusted in that way, it's what is – what is average for you patient, that's what the expected …

(Crosstalk)

(Peter Thomas): OK.

Gunnar Nelson: Which is just technically slightly different than what you had stated.

(Peter Thomas): Right.

Gunnar Nelson: OK that's helpful.

(Sherry Kaplan): This is (Sherry) again. Can I ask one more question? What was the association, it's a little bit weird because age is usually associated with functional status, granted it's not always a real strong positive association but it's usually correlated, what was the – in the concern about things drooping out of model one always worries that that's because one of the variable is soaking up all the shared variance and if it wasn't there, and it's only kind of edged in to the computer by whatever logic and algorithm it chooses, you know, among correlated variables, which one gets the lowest P value.

So, is there – in the – well, two questions, one is, is the expected score computed overall of the hospitals in your area or region or – what's the denominator there? What's – what units are being compared and how is the expected score derived? And second is, is there a concern about association correlations among the independent variables in the risk adjustment model that would cause one to drop out because there's too much variance.

Gunnar Nelson: This is Gunnar. First of all we test each variable independently does, what it stands on its own, and then as we start seeing which variable have some – some impact and then you start doing the combinations of adding age on top of functional status, and we found that if there wasn't a strong association between initial functional status, it's there but it's not enough to change anything, and the other questions is how it's calculate is the (inaudible) variable, that how we handle this one as we (inaudible) this way of running it as a linear regression.

So, for you to have a functional status of eight, you would expect an outcome of (small) points.

(Sherry Kaplan): No, but you're derived expected values or across, everybody in your population, right? Or across whatever units you're comparing not within the hospital.

Gunnar Nelson: Correct, correct, I'm sorry I forgot that. It's – all the data submitted for that measurement period, providers who are being reported and providers who are ineligible to report because of sample size, they're all (can) create the expected value. So comparison to your market.

(Sherry Kaplan): Thank you.

Sarah Sampsel: OK. Do we have any further questions? Comments from Minnesota?

(Off-mike)

(Becky Bradley): I'm sorry this is (Becky Bradley), I have two questions.

Sarah Sampsel: Sure. Go ahead.

(Becky Bradley): Could you remind us what the intended use of these measures are? Is it for format improvement, is it for payment – and I apologize I just don't have the material in front of me and I'm sure that was mentioned in some of the previous materials. And then the other question was, as you were looking at variables or modality or therapy – not considered as one of the intervening

variable slide because they got physical therapy versus sessions that did not get physical therapy during that period ...

(Off-mike)

(Collette): This is (Collette), I can talk about those. These measures are intended for use so we're reporting at the practice level. They're intended for public reporting on our Minnesota Health Course website, they're included in the Minnesota department of health statewide quality and reporting measurements system. So, again, both for accountability for quality improvement.

And the second questions about therapy as a modality, during the measure development workgroup process early on. That was discussed as, is this is a component of our measure, is it something that we're going to collect and capture? And we do – the practices in Minnesota are submitting patient level data to us and we're constantly conscious and aware of feasibility and burden of the data collection and what's been submitted to us. And, through a lot of discussion it was determined that therapy was not going to be one of the components that was submitted.

Gunnar did you have something else to add, I'm sorry.

Gunnar Nelson: No, just on the P.T., there are no requirements for a risk adjusted variable by both (instance) community of measurement and by the NQF recommendation is – it has to be independent of what the provider – of the provider's action and recommending P.T. is not independent of the provider's action. So, that would not make that would not make that cut anyway even if we hadn't collected it.

(Collette): Thanks.

Gunnar Nelson: Thank you.

Suzanne Theberge: Further questions? OK. If we concluded our discussion of the Minnesota, we should thank you for your hard work and continued submission of data and I think Sarah, we want to turn it over to Chris to co-chair the next section.

Sarah Sampsel: So, I just want to make a couple of note on this, we had hope to vote online and have the committee determine if – what was presented by Minnesota committee measurement and the update satisfied any lingering questions regarding risk adjustment methodology. And unfortunately we just lost quorum. So we – we'll have to do and we'll send out a SurveyMonkey which will include both the determination of removal of the condition on these two measures as well as for the full PAM measure evaluation which was addressed momentarily.

But we'll be sending that SurveyMonkey out and all these – having a question on these measures online. So, just wanted you to prepare for that and (Collette) and Gunnar, thank you and we will obviously follow up with you and provide the outcome as soon as we possibly can.

(Collette): Fabulous, thanks for the opportunity to join in today.

Gunnar Nelson: Yes, thank you.

(Peter Thomas): Great, thank you very much.

Sarah Sampsel: Yes. No now we are going to move, we're going to move to the next slide, actually let's just go ahead to another one.

We have a series of just a couple of slides to remind folks first an introduction to the performance measurement evaluation review, of the PAM and then a few slides just to get everybody's mind, you know, even more reminded about the NQF evaluation criteria. And I'll – I'm actually going to move through those as quickly as possible just in later time, but at the same time, you know, if you do have questions on interpretations of criteria we really would like.

And, just wanted to mention that Helen Burstin is on the phone as well, so that if there are some specific questions that, you know, we all are struggling with, with review of this measure or globally, it would really be helpful and Helen would chime in.

So, I'll provide a brief overview and reminder of the (excess) and then we will turn the call over to (Judy Hibbard) and the team from Insignia. I know at

least (Jim Hornish) is on the phone and I'm not sure of others to make any introductory comments as we do during an in-person call.

And then, Jim Merlino is going to guide the discussion based on the NQF criteria and as a reminder as we do during the in-person meeting, we need to be mythological about that and ask that we talk about importance. So evidence and performance gap first. And then move on to scientific acceptability and feasibility, usability, and use in any overall concerns about feasibility for endorsement.

As I just mentioned had previously intended to do the vote online and have this software available, unfortunately since we don't have quorum we will not be able to do that. And you'll receive a SurveyMonkey afterwards.

So next slide. I think I actually covered this. So just as a reminder, with quorum, we are looking for 12 folks, we only have 11 on the phone so we do not reach quorum. But we do go ahead and do the full discussion on the measure and folks that – who were unable to participate on the call will have the advantage of using both the call transcripts and all of the materials that were distributed prior to the meeting. And all of those will be posted on the committee's SharePoint site fro you all to refer to, and we'll send out a SurveyMonkey. And then the interpretation of those votes will be greater than 60 percent on any of the criteria, the passing or recommended vote.

We still have the gray zone, which is the consensus not reached and then the do not pass, not recommended vote, would be under 49 percent. Next slide.

So as you're thinking about evidence in performance gap, this is to help outcome measure, and so what we are specifically looking for is for the committee to assess that the relation between the measured health outcome and at least one health care action is identified and supported by the stated rational and then all of the evidence in the evidence format was submitted behind the measure. With performance gap, we're looking for demonstration of quality problems and opportunity for improvement.

Typically those things would be identified through considerable variation of the performance metric, if there's overall lesson optimal performance based on the data provided. And then any comments made about disparities and care across the population group.

Next slide. With scientific accessibility which is reliability and validity, with reliability we are looking two things as the measure is well-defined and precisely specified, and that testing demonstrates the measure data elements are repeatable. As a reminder this is a Patient Reported Outcome Performance Measure or PROPM, therefore we are looking that for both reliability at that data element level as well demonstration of reliability at the computed performance score.

For validity, you're looking for measures specifications that are consistent with the evidence to support the focus on measurement, and that testing demonstrates that the data elements are correct. Again the PROPM, we're looking for tested – validity testing at both the data element level and the computed performance score level. And then, reminder under validity, you are looking for assessment of threats to validity and those are things such as risk adjustment, missing data, exclusions and all of those that are listed under that part of the NQF criteria.

Next slide. And then finally, just a reminder, with feasibility, usability use, and overall feasibility, the first three criteria that I just went through are all must pass element, feasibility, usability use and (over suitability), we look for your overall votes, they would not stop the measure from going forward but are obviously an indication to NQF staff and the community that there maybe some concerns. But with feasibility you're looking for extenuate specifications including the measure logic required data that are readily available and can be captured, and can be implemented. Usability and use extenuate potential audience are using or could use performance results.

And then the overall suitability for endorsement is your overall recommend measure based on all NQF criteria and whether they're meet or not meet. I'll just note here that we did not include the other consideration here as related and competing. And we did not claim any related or competing measures

currently in the NQF portfolio or where if any development or measures coming forward that are related or competing to these measures. So, we just did not bring that up during the slides.

Next slide. And then just of course an overall reminder about the PROPMs and just for the committee discussion with (Judy Hibbard) and the Insignia team regarding the PROPMs, and the distinction between service instruments and tools and of measure. And so even through the elements perhaps of the Patient Activation Measure or the PAM, you do want to see testing of those because of your data element, we are not endorsing the PAM itself, we are endorsing a measure that is derived from use of the items collected.

And that NQF performance measures are intended for use in both performance improvement and accountability applications, that could be public reporting, that could be Pay For Performance. I've already mentioned the repeated levels of testing, both the patient level or data element level and the computed performance score. And then, the special item that is specific to PROPM is that developers are asked to provide as it ends to the target population value, the measure patient reported outcome and find it meaningful. And Insignia – (map) based on some external testing that was done.

Next slide. Next slide. So with that, I'd like to turn the call over, and believe (Judy), you are going to present the measure or if not, if somebody else on your team is going to present the measure it's your turn to talk.

(Judy Hibbard): Thank you. Yes, this is (Judy), I am going to present. And thank you for the opportunity to consider. So, the measure that we're talking about are changes in the PAM score overtime. And just to remind everyone, the PAM measures and individual's knowledge, skill, and confidence and their ability to manage their health and their health care.

The rationale for this is that, we do know that the PAM score is predictive of health behavior, clinical outcome, many measures of utilization or costly utilization and overall cost. The underlying assumption is that if patients are

getting high-quality care, they should be gaining in their ability to self-manage overtime and that's what the change in the PAM score would tell us.

It measures an individual on a 0 to 100 scale, it is – it uses a – either 10 item or 13 item survey. We have seen that changes enhance scores are length with outcomes – all of the outcomes that I mentioned in the expected direction. And so that's an indicator of validity. We're proposing based on examination of data from several sources that the numerator of the measure would be the aggregate change in PAM score for a defined population, and the – over a 12-month period but not less than a 6-month period. And the denominator would be the patients in that facility or that panel, and who would be exclude – or who would be included were patient who have at least two visits during that time period.

We would exclude people who have diagnosis codes indicating cognitive impairments, we would exclude children under the age of 14. And we also exclude people who at baseline measure at the highest level of activation as they are unlikely to increase overtime.

In looking at different sites, we observed that in one site where there were 295 primary care providers – oh I'm sorry, I forgot to say that we are proposing that a passage score would be an aggregate positive change of three points on that 0 to 100 scale. And an excellent score would be an aggregate change, positive of six points. And looking at large delivery system with 295 primary care providers, who were not doing anything targeted to increase activation, 62 percent would pass using that criteria.

So, at this point there are organizations who are using a measure or are planning to use a measure similar to this as a performance metric. And so the – defining what constituted good performance. I think it's very important to do it at this juncture. And that was all I was going to say and we have questions then I'm happy to respond.

James Merlino: Thank you, (Judy), this is Jim Merlino and that we appreciate your presentation. And so I think we can open it up for the discussion for the order of topics. Are we ready, Sarah?

Sarah Sampsel: We are. Just a reminder that we need to go through the criteria in order, but I think what we have noticed in the past is it help focus, start out the discussion on as any member have questions about the overall construction of the measure and insuring people understand exactly what the measure is. So, why don't we start there and then move to (importance).

James Merlino: Any other – any overall questions that the group has?

(Peter Thomas): I have one, this is (Peter).

James Merlino: Hi.

(Peter Thomas): The Insignia, a couple of documents towards at the very end, the – it's got the Insignia Health logo on it. A series of 10 questions, is that the actual measurement tool that's being used?

Male: Yes.

(Judy Hibbard): Yes.

Male: Yes, that is.

(Peter Thomas): And that's the – what does that done over the course of time, every integrals or something?

(Crosstalk)

James Merlino: Or it's just a one-time thing.

(Judy Hibbard): For the measure, people have to have two scores to see a change. So, the measure requires measurement at two points in time. And we're proposing that that be a once a year but not shorter than six months.

(Peter Thomas): OK.

Katherine Bevans: This is Katherine Bevans. I wanted to ask question about kind of the nature of the score for the performance measure. Just to be clear, is this – would this be a continuous score or is it going to be the categories that clinicians or

whatever aggregated unit are put into with regard to, you know, the passing or excellent category. And also, could you just explain quickly how those criteria were determined that, and increasing three points is equivalent to passing a particular score were selected for categorization into the excellent group?

(Judy Hibbard): Yes, so the – three points, there's actually two rationales for that. In study we've seen that a three-point change is related to changes and behavior and sometimes (best clinical indicators).

And, the – we look at the three points also as what would be reasonable in terms of how many clinicians would pass. And the other – and the excellent would also, based on how, you know, how many would reach that level of performance.

In terms of the reporting, kind of all units can also, you know, look at it in terms of the change, and as a continuous variable, but we were looking at this also as having the option of these levels of change.

Katherine Bevans: And so I guess my follow question to that is, perhaps to the committee leads, are we being asked to consider endorsement for the continuous measure or for the categorization of passing excellent or excellent or both?

Christopher Stille: Right. And Katherine, this is Chris and I'm struggling with that. The way the worksheet seems to have been filled out, it's the aggregate average change in patient scores, which sounds like it's more of a continuous variable that's described in some of the literature that the developers used to support it. So I guess my question from that is, I'm a little bit confused, it seems like there's a lot more literature that was presented to support the PAM itself rather than changes in PAM. And, I'm a little bit confused about, is there a difference between this proposed aggregate score across patients?

And for example, the proportion of patients that have some significant change overtime. And then related to that is, you know, the categories of the PAM 1, 2, 3, 4 versus the absolute change in score. So, I think we – the group could

probably use some clarification and some (windowing) down of exactly what evidence do you have that's specifically supports the measure as specified.

(Judy Hibbard): OK, I'm sorry for the confusion. We are proposing the two, the passing and the excellent performance. And it's made up of those changes. In terms of …

Christopher Stille: So it's – I'm sorry I cut you off.

(Judy Hibbard): I was just going to on to answer the second part of the question …

Christopher Stille: Yes please, yes, sorry.

(Judy Hibbard): So there is evidence that the change in PAM not these levels, so to (say) that we're talking about. A change in PAM is predictive of change in clinical outcomes and utilization and cost. And that is – wasn't presented as a performance metrics but it was a study that was published in health affairs this year. It doesn't apply to this performance metric, is that what you're asking about?

Christopher Stille: I guess what I was just asking is for you to point us toward – I mean there are lots and lots of literature, you know, talking about how good the PAM is and, you know, and I believe all that stuff. But the specific literature supporting that a change in, you know, three points or six points for an aggregate group of patients as oppose to change on the individual patient level, if you know what (I think), it's good, it leads to better outcomes.

(Judy Hibbard): Yes, we do have that. And I believe it's – that we provided the citations for that, but I can point them out.

Christopher Stille: OK, that would be great if you could point those out, thanks.

Lee Partridge: This is Lee, I don't want to keep beating on this point but I am now totally confused. When I read the submission, I understand the numerator was just as (Judy) said at the beginning. The change over a max of 12 month but a minimum of 6 months, and the person PAM Score, from point A to point B.

And when I saw the discussion about passing an excellent, I thought those were sort of recommendations for how a user might want to perceive what was considered good performance or acceptable performance.

I think what if – what I'm now hearing is, we want to put into the numerator those three and six point numbers, am I right or wrong?

(Judy Hibbard): This is, you know, the learning curve that I've been on, had to translate where we know into this performance thing. That I think that we started out with that – the idea that it's a recommendation and somehow got morphed into …

Lee Partridge: Yes. Because, personally, I think the change over time can be used independent of the discussion of what would be considered passing or …

(Crosstalk)

(Judy Hibbard): Yes.

Lee Partridge: But – OK.

James Merlino: So are you OK? We are doing a – we should talk a little bit more about it. I think it goes back to the bigger picture as we were reading the review to, you know, with what Chris pointed out, I think we're all familiar with PAM, and what a terrific pleasure it is but we just need to be very clear about what we're asking the group to look at here.

So (Jane), maybe just one more time, can you just compartmentalize how you're thinking about PAM over as it's proposed for the review.

(Judy Hibbard): So I think that what Lee said that we would that measure a change, and then we would recommended that that's what it means.

Lee Partridge: Yes. It's started like the discussions we've had around CAHPS where we had the measures themselves and then some discussions about protocols. We didn't endorse the protocols.

James Merlino: OK. So …

Lee Partridge: Sarah? (Helen)?

(Helen Burstin): Well, yes. I mean, this is (Helen). I mean, I think would just go by with the numerator statement says which is the summary change score for the aggregate of patients. They then make recommendations about, you know, what excellent might be, but I think you're looking at – we – of course, do not endorse tools. So we're not endorsing the PAM. We're endorsing potentially the measure that looks at the summary score change for the aggregate of eligible patient, is my understanding.

(Crosstalk)

(Helen Burstin): And, (Judy), am I interpreting that correctly? I want to …

(Judy Hibbard): Yes. Yes. Thank you.

(Helen Burstin): OK, great.

James Merlino: So, Sarah, are we good with importance and evidence and/or should we move on to performance here?

Sarah Sampsel: Yes. I guess I would just ask, I mean, I think everybody has asked the questions about evidence, you know, is part of just getting clarification of the measure. But if were – if there's no other questions, we can move on.

James Merlino: Any other questions on evidence? OK. Let's go to performance gap. Any questions or comments? None?

Sarah Sampsel: That is just straight forward, I guess.

James Merlino: All right. So – and moving to scientific acceptability because we're not – we don't have a quorum, so we're not moving on each one, we're going to discuss each one.

So looking first at reliability and meeting Sarah's slide of, you know, having it being well-defined and repeatability, (inaudible) a reliability.

Christopher Stille: Yes. It seems like the there was good data that were presented and individual item reliability as well as not test-retest but the other one. I don't think there's specific test – what retest reliability presented.

But, again, you know, we're not evaluating the measure, we're evaluating the changed over time. And if so, I don't know that there is specific issue or reliability with change over time necessarily, other than they're might be reproducible. I just kind of thinking out loud.

(Judy Hibbard): There is, actually. We did present evidence a test-retest reliability in the original PAM articles for this.

Christopher Stille: Oh, OK, yes. That's OK, great. I think I didn't look at those …

James Merlino: And has a very high internal consistency reliability, on which you (inaudible). I think it was 0.8, 0.9.

Christopher Stille: Yes. Those are all like 0.8, 0.9, yes.

Male: Is that Cronbach's Alpha, my favorite measure, reliability.

James Merlino: Yes.

(Crosstalk)

Male: Very consistent scoring in that, I noted.

Female: This is a comment that I think pertains about reliability and validity as well. I noticed in the instrument level reliability statistics, the Cronbach's Alpha is that more presented that – unless I missed it, there were no reliability or measure about validity or measure about validity but certainly reliability data presented for children. And by that, I mean, 15s to 17s even though the request is for endorsement for 14 plus I believe.

So could you talk to use a little bit about what work has been done both on the measure but then also at the performance measurement level for young people, 15 to 17 years of age.

(Judy Hibbard): Yes, good question. We don't actually have much data in that post area. Chris, you know, if we have data on that page group.

Christopher Stille: I don't. I suspect there hasn't been any work done, especially because there's so much transition especially because there so much transition especially among kids with chronic conditions, there's a lot of transition between parents and, you know, kids and parents in terms of, you know, independents and stuff like that.

I think it would be useful. I think there are a couple of self-efficacy measures that have been used in couple of studies, not many among kids and parents, but not the PAM specifically.

Female: I'm bringing this up in part because we've seen this before where instrument, you know, we're being asked to consider measures for age groups for which there is minimal reliability and validity data. But I'm especially concerned about this issue for this measure because of the content of the items, because of the content.

Christopher Stille: Right.

Female: The meaningfulness of activation looking at the items, they are cognitively difficult first. And also just may mean something very different for child whose, you know, parent or caregiver tends to take primary responsibility for managing their health condition.

So I do personally have some concerns about, I don't know if anyone has any other thoughts.

Christopher Stille: Yes. You know, it hasn't been tested under 18. We probably can't say anything about under 18.

(Crystal): So (Crystal) in here with Insignia. Quite a few (days) we've take – taken in over the years have had, you know, kids in the, you know, 12 and above kind of segments with descent samples sizes, so not published literature, really focusing upon adolescence for instance. But we've been asked a number of

cases by clients to offer an opinion on its applicability to adolescents, you know, to teenagers.

And so, we've look at the difficulty structure of the items, so do – is a 14 or 15 or 16-year-old respond as adult do, as one example into these segments a long the lines for instance of adult.

Again, this is kind of at the PAM level. And the answer from what we've seen in data sets is yes. We start to see some shifts, you know, when you get into below the age of 15, and organization and that context are using the parent PAM, which is gaining increasing amount of publishing, where in fact the parent really is that, they're responsibly party for the child.

But we could share some of the insights and that kind of 15 to 18 range. But you're correct, they're not in the application but we have some confidence at the kind of higher teenage range.

Female: Yes. I think it's helpful to include because it is important to assess these concepts by adolescent report.

(Crystal): Yes. I agree.

James Merlino: Any other – any comments from staff about reliability-validity?

(Peter Thomas): No – well, this is (Peter Thomas), I'm not staff, but I do have a question. I honestly don't know where it goes in this construct. So, I'm just going to ask it now, if you don't mind.

James Merlino: Sure.

(Peter Thomas): I did go through the list of 13 questions on the documents (externally) with Insignia on it. And I just have a question about – so I recognize that there are 25 clinical trials and haven been done on this and that there is, you know, seems to be very good evidence about this measure. And I know we're not really looking directly at the measure.

But nonetheless, question five, just as an example. I'm confident that I can tell whether I need to go to the doctor or whether I can take care of the health

problem myself. So you're – the question goes to the individual's confidence, and I could be really confident or something that I could take care or something myself but I'm diluting myself. And in a much more serious health care condition, how do you address that kind misinterpretation of the question? Or potential misinterpretation?

(Judy Hibbard): So, you know, people do – and can interpret the items. But the overall score is the summary of all of their responses, and there – and what the measure the is a – you know, the (Roche) creates a measure that is unidimensional sort of tapping into one underlying idea. And that's the – what's what we're going on that underlying idea.

We think its fair so, concept as a manager of their own health. So if their understanding of their situation, its way off, it's probably way off on many things.

(Peter Thomas): Many of those questions, I see.

(Judy Hibbard): Yes.

(Peter Thomas): OK.

(Peter Thomas): I thought of a question involving gap but I hope you don't mind if go back.

James Merlino: Go ahead.

(Peter Thomas): You had mentioned in your presentation that, you exclude the people who score very highly because there's not a lot of room for improvement? Is that right?

(Judy Hibbard): In the aggregate score, yes.

(Peter Thomas): And I was just wondering, isn't that typically an indication that there's not a – at least for that population that there is not a significant opportunity for improvement, through the use of the measure itself, rather than just kind of using it as an exclusionary, you know, to exclude that data.

I don't know whether that's completely …

(Judy Hibbard): But there is an opportunity to work with that population in a more targeted way, because their challenges are different, so their challenges are more out relapse prevention, and maybe, you know, having some stretched goals, so understanding that the patient is there if useful to that clinical team.

It just that – it sort of disadvantages. It disadvantages the group, the measurement, if they have a lot of patients who are already highly activated, because they're not going to see a lot of change.

(Peter Thomas): All right.

James Merlino: Yes.

(Peter Thomas): Thanks.

James Merlino: OK. So let's go back to scientific acceptability. Any other comments of the reliability or validity, or concerns?

(Sherry Kaplan): This is (Sherry). I apologize I missed the front end of this discussion. Is part of this measure proprietary?

(Judy Hibbard): Yes.

(Helen Burstin): That will come up under feasibility, (Sherry). So we're getting there.

(Sherry Kaplan): Holding on that question.

James Merlino: Reliability and validity, any other comments?

(Lisa Morris): This is (Lisa). I'm also not sure exactly (what's in) – and so, since I'm currently recovering from really bad pneumonia, I am wondering – my experience has been new for me and that, my functioning was severely impaired, especially at the beginning of this acute incident. And I'm wondering if there's room in this measure for taking activation to wax and wane over time – isn't just a baseline and an end for – but in between, there maybe ups and downs. Does that makes sense?

(Judy Hibbard): Yes. And in fact we do see some of that. That patient do go down, and it can be related to some adverse event in your life that may not even be a health event, but that's why people sometimes measure more often, so that they can really be meeting the patients where they are.

But – so, the ups and downs are, you know, kind of factored in and we think about that over all patient, should be gaining, some won't be. But they can – even the ones who are following back if they're appropriately supported, can comeback.

(Lisa Morris): Thank you.

James Merlino: All right. If there's no other questions or comments on reliability and validity, I think we'll move on to …

Christopher Stille: Jim, this is Chris. Just to reiterate, I think when we get ready to vote, if would be really helpful for the developers to point out that the best studies, they're specific to reliability changes in the PAM as proposed.

You know, I because – these are the six that we think are the best, or something like that. That would be super helpful.

James Merlino: Maybe what we can do is, is collect that and then send it out to the committee after the call.

Christopher Stille: Yes. Yes.

James Merlino: They know exactly what (to look at).

Christopher Stille: OK.

Female: Before we move on, just one other point that I was hoping developers could point us in a direction of a literature on, since there is so much on the PAM, but what about evidence of measurement of equivalence across modality. I understand that PAM can be administered on paper or electronically, or even over the phone. Is there research on relative …

(Judy Hibbard): Yes. Actually there is a randomize study where people, we're randomized to the telephone or the web. And for other measures, there was a mode effect, but for the PAM, there was not.

Female: OK.

(Judy Hibbard): Then, we can share that citation and stuff.

Female: Great. Thank you.

James Merlino: Great. Let's talk about feasibility, comments or questions?

(Sherry Kaplan): Well, this is (Sherry). I was asking about what parts of the measure are proprietary, is the questionnaire is self proprietary, is the scoring proprietary, is all of the above proprietary?

Christopher Stille: So hi. This is Christopher Stille again. So, I guess, all of the above would be the correct answer. The surveys and all the PAM versions themselves are kind of registered trademarks if you will, ultimately owned by the university and state of Oregon, the algorithm also proprietary.

We do, on occasion, give clients the algorithm to integrate to their systems, their EMRs in particular. We also, you know, setup score web services where we have the scoring provide scoring tools to clients and so on, but it is all proprietary.

(Sherry Kaplan): And so to use this measure, you would – you would have to sign some contract – we just have the contract with who to make – to use it and in what capacity – what's the nature of your contractual arrangements with folks trying to measure quality?

Christopher Stille: Right. Right. So a various based upon the use, so as an example in, you know, New York State DSRIP, PAM is required but we had the license with each of 14 DSRIP organizations. In England, PAM is become a national vital sign and we have license with NHS England, so at a real macro level. We kind of go all routes, you know, creating license global, hospital level, health system, health Medicare bench plan and on and on. And licensing process,

you know, it usually runs anywhere from a few days to few weeks for most organization.

Brian Limberg: This is Brian, Brian Limberg, could someone explain to me how that would play out if, CMS wanted to use a measure and, I mean, we want – I mean, you want to use it everywhere for example.

Christopher Stille: I can give you a couple of examples of how in the past we've worked with CMS. So PAM was required in the ACA for 3026 which was preventing readmission post-discharge. And what they did there is required its use, but the license in our relationship was through the technical support as one example.

In the current scope of work for QIOs, CMS recommended PAM and we have created relationship with (Telegan) which – is the support center for I believe all of the key initiatives for the next five years for QIO. So we have a relationship from a licensure standpoint with (Telegan). So that's how it work for us in the past, via kind of CMS elected initiatives.

Brian Limberg: And is there a charge for the use of the measure?

Christopher Stille: There is. So it's a charge that includes some training guidance of access to PAM and all translation in caregiver PAM and parent PAM, and scoring services and all of that. And it varies but its right around a dollar, of patient a year, if it's not kind of use this races, those who use. And if it where kind of spread over a population, then it's, you know, 20 cents a year on a population basis.

So we look at pricing in both manners. Usually it larger program and in the context of CMS would be, you know, population basis. So it would probably have reach out to 15 cents, 20 cents, a year, a person, something like that for that kind of services.

Brian Limberg: So I wonder if the NQF staff could just comment on …

(Helen Burstin): Yes.

Brian Limberg: … on that issue and whether we freely endorse. I can't remember what other things we've endorse that there is a charge with.

(Helen Burstin): Yes. This is (Helen), I would be happy to give some historical context.

So probably about seven years ago, the NQF Board decided that there should be a (corridor) to allow measure to come in, that have an associated fee, and we ultimately determined that that associated fee would something that committee would discuss and in fact on your SharePoint site, there is a documents from Insignia that gives the scoring – that gives the resource pricing, which is the requirement of an endorsement. And then, you would factor this in as part of feasibility assessment when you're voting.

And we do have a couple of other examples like this for example. There is a proprietary tool used for pediatric ICU severity adjustment that is built into measures that has associated fee. And there are several risk adjusters also that are built into measure that have associated fee. But again, from where you seated is more an issue of, you know, it gets factored into the feasibility criteria.

Brian Limberg: Thank you.

(Jim Hornish): Yes. In the SharePoint site, there's a complete deck in the business model around PAMs, very transparent.

(Helen Burstin): Yes. Thank you.

(Sherry Kaplan): And this is (Sherry). To clarify is – who is contract or is it Insignia or is it the university?

(Jim Hornish): So Insignia, essentially, where that – the caretakers, if you will, for the state and the university. So we handle that kind of process. I should also add that, this is included some of the pricing materials that we tend to include all of our products, no charge, when we deal with kind of uninsured safety net programs. And then, researchers up to a certain level could use PAM without charge as well. So there are some kind of free options as well if that's have interest.

(Becky Bradley): This is (Becky Bradley). So I guess, I'm a little – still not clear how do they just collected. So if you look at the care setting that you've described like in a long-term care facility, or inpatient rehabilitation facility. I can see where you would maybe collect the data admission but then, how do you get it a year later? Who is actually responsible for contacting the patient or administrating the questionnaire? Or is it a mail questionnaire?

(Jim Hornish): We've seen a kind of all of the above in terms of that repeat PAM. You know, it's mailed in home. There have been phone calls made in home. You know, in some settings, you know, it's easier when you have a regular, you know, interaction with patients over the course of the year.

And in this application, we've focus on within 6 and 12 months, but admittedly we have many clients that, you know, think about the ACA provision with readmission post-discharge. PAM was a readministered, you know, on day 28, 29. We have clients that readminister, you know, months, two or three because they may not usually have an interaction that six months or 12 months. But for the purpose of this application we focus on that 6- to 12-month kind of emphasis.

(Lisa Morris): This is (Lisa). It sounds that what you're saying is the provider who get – makes the assessment of point one basically has the responsibility for the follow up generally.

(Jim Hornish): Typically correct. So that organization perhaps not the same individual but organization and through some kind of method or mode of an erection, correct?

James Merlino: Any other questions or comments on feasibility? Shall we move on?

Usability and use, comments, questions?

Christopher Stille: Everything seems fine from – this is Chris – from what I've seen. It's use a lot. It's short. It's pretty easy.

James Merlino: Yes. And I would agree.

Christopher Stille: Yes. The only thing is, you know, to reiterate Katherine's point about – age group usability.

(Jim Hornish): Yes.

James Merlino: Can you, guys, in terms of how you execute in this. It's really – when you're looking at using multimode, correct?

(Jim Hornish): Correct. I would say we have a very even split between patients taking it, you know, of their own accord, whether it's only, mobile, SMS, paper, or being asked the questions, or having it done telephonically. So we've, you know, practically, you know, we tell our clients whatever modes is the mode to use is the mode that will work. Yes. So it is multimode.

(Sherry Kaplan): This cannot be a usability question, it's (Sherry) again. But if you – in the attribution to care as an outcome variable, if one which is to a use of these measures of risk adjuster, for example, if people who are more active in their care get better outcomes, independent of other things that we mean full and see outcomes. If you flip the use, do you the same answers with respect to you usability feasibility, et cetera or even the reliability and validity et cetera?

(Judy Hibbard): (Sherry), this is (Judy). I don't really recommend using the PAM as a risk adjuster because we really want accountable units to focus on improving, you know, helping patients to gain in their ability to self-managed. If you use as a risk adjuster, you sort of wash it out, you don't – you're not attending to it.

(Sherry Kaplan): I get it. I was interested because sometimes where the constituents that a unit of care provision is serving. They can't control some of the folks that arrived there. And they get compared with things that associate demographic set, the easiest one. But this kind of has the feel of something that may not be – it's who you attract not what you do to them, that may have an important influence on the outcome. So I just – was curious about whether or not you considered this as a potential risk adjuster.

(Judy Hibbard): I don't think we have or that anyone used it in that way. We do see that in studies what when you offer patients something special like a patient portal or

special class for chronic disease health management, that it is a higher activated who show up.

(Jim Hornish): Yes. Yes. You know, we have – we do see organization using PAM as a, you know, as a risk adjuster but as a layer on top of their predictive models, so in some pretty good data sets, we found that kind of those retrospective markers, claims data, clinical data, even demographic socioeconomic. They tend to miss on average about 55 percent of patients low in activation. So we do we PAM as prospective risk measure on top of or on in collaboration with kind of those retrospective database models.

So not quite, you know, I was risk adjusting but adding this as a insight, again, this isn't around the – on the outcome assessment here but PAM is being used more in that manner.

James Merlino: Right. Any comments or questions about overall suitability to endorse?

I think we've addressed the issues (inaudible). It seems like we're in a good path, does anybody have any other comments? Great.

So, Sarah, do you open up to public comments?

Sarah Sampsel: Yes. Let me just make a couple of kind of summary statements about the evaluation of that measure. And, first of all, we would like to thank (Judy), Jim and Chris for being available to present the measure and answer questions.

And we, staff, will follow-up with you but I think you heard from the committee that they would just like highlighted a little better in the (field) that you presented, those specifics studies about the questions on reliability and validity and the adolescent study or adolescent data.

I know and if I remember correctly, the best study might be the one that you have attached to one of the PDF, Jim, and that was – is it a health affairs article?

(Judy Hibbard): Yes.

Sarah Sampsel: So that might, you know, and if there's any additional to that, if you could get those to us by Tuesday at the latest, because we'll be sending out a SurveyMonkey by Wednesday when we get the transcript, but we'll follow up with you.

And just wanted to thank you and, you know, I'm sorry we can't give you a vote on the call today, but we do need quorum to do that.

(Jim Hornish): Sure. Sure.

Sarah Sampsel: So with that, (Shawn) or operator, can we please open up to the public comment?

Operator: At this time, if you would like to make a public comment, press star one on your telephone keypad.

And there are no comments at this time.

Sarah Sampsel: OK, great. Then, we will move to next steps, let me see if I could forward the slide so I can remember what they are. OK.

So what I wanted to do hear is just provide some updates on what transpired over the past few months when we haven't met, although there has been kind of a number of e-mail updates that you've received, their project notices on phase two, where we're going on this cycle and then, introduce you to what will happen in phase three.

So just kind of as a recap with phase two, as you might recall we had a total of 28 measures. There was significant back and forth on those measures between the in-person meeting and they're in public comment and additional data provided. And, you know, these were the functional status measures and there were a lot of questions about the care tool that CMS is implementing and deriving measures from – in response to the IMPACT Act.

So we saw a number of those measures come through as well as some measures and competing measures, specifically for inpatient rehabilitation

facility on self care and mobility functional status and change in functional status.

So as we come through these past few months, the final of this decision of the majority of the measures are 24 out of 28 of those measures, the vote had been through the CSAC and through final ratification by the NQF board to endorse those measures.

We did have the four measures that had been identified as competing and so those were two change in mobility measures and two change in self care measures, one being – one set being based on the care items, the others that being based on the use of the (FIM). And those measures have been the ones that have taken a little bit more time and consideration through the NQF process. And they were just – their final disposition was just approved by the board this past week and those four measures are out for appeal right now.

So really what that means for all of you is that, staff is paying close attention to the functional status measures overall and how they fit into the person- and family-centered care portfolio. We anticipate additional competing measures coming forward, whether they are from UDSMR and based on the (FIM).

But we also anticipate additional similar measures based to your items from CMS sometime in the future, and we really just want to – and we will be reminding you all to kind of remember and go back to why there were some concerns and challenges as raise in consideration of the evaluations in the first place. And a lot of those considerations had to do with the newness of the care items, but then also the burden of implementing additional assessment tools on top of assessment tools that are already (elected), but then, going back to the reminder that we don't endorse assessment tools, we endorse the measures.

So we're currently working through this process, we're learning from the process. And as we move forward and see more functional status measures as part of this standing committee, you know, we'll be coming back with more guidance.

But the final disposition of the four competing measures was they were approved for endorsement with conditions. And those conditions were based on discussions on the standing committee and included such things as, you know, if industry is moving forward and can – stressing the – some concerns about the sensitivity of the care tool, stressing concern about of the care tool. We're ensuring that all of that information is shared that with CMS so that as they implement those measures starting in 2016, those are some areas of special consideration that they are – paying attention to and that they continuously update NQF during their annual maintenance process as well as then their three-year maintenance process. Kind of, what's going on with the sensitivity and what concerns are being raised, that even though they provided and met the criteria during their initial endorsement, you know, are their additional things that are coming to light when they implement those measures.

So we'd asked for additional data from CMS on that, we've also asked for additional information from both UDSMR and CMS on burden, and really what is it going to take for facilities and specifically the post acute setting not only in-patient rehab but we'll be seeing measures for skilled nursing and health, any of those facilities that fall under the IMPACT ACT, long term care hospitals would be the fourth is, you know, what is going on with burden how is CMS thinking about that internally, and how is that impacting their implementation of measures into their quality reporting programs. And so, we've asked for updates in that.

We are actively meeting with CMS and UDSMR, we will on an ongoing basis to ensure that we're bringing information back to you for any of those areas of concerns, that were raise during the in-person meeting, that could have been raise during CSAC and subsequent public comment periods.

So I just really wanted to provide, you know, kind of very quick overview in what we're doing and that we are in constant communication and that's one of the reason for standing committee within NQF now is that, these are things that you all can monitor over time. But that also have to be prepared for, they may impact future consideration of measures down the road.

So I'll stop there and see if anybody has any questions on phase two.

(Peter Thomas): I do.

Female: Go ahead.

Sarah Sampsel: OK.

(Peter Thomas): This is (Peter Thomas). So you said that they four measure, self care and mobility care tool and (FIM) were in the appeals period or something along the most – are they actually being appealed by someone – are the decision is being appealed by someone?

Sarah Sampsel: So, listen, this is a standard part of NQF process that once measures go out for – once the board approved to ratifies both the CSAC and standing committee recommendation for endorsement. There's a 30-day appeal period. So we won't know until mid-December if anyone is appealing the current endorsement status of those measures.

(Peter Thomas): And just so I fully understand it, the endorsement status currently, we – there was an ongoing discussion about whether they were competing measure and that one should be kind of best in class or I'm not sure exactly the terminology. That was differed, right, we didn't decide – that NQF did not say one was best in class as opposed to the other. That still out for another year of data collection?

Sarah Sampsel: No. So really what happens there is, the standing committee as – you're correct, (Peter), did not come to consensus on designating best in class measures, the measures and what with the CSAC. And this tool some additional time and frankly, the first time the measures went though the CSAC. They did designate a best in class.

But when the measures went to the board, the board asks that the CSAC fully reconsider in light of public comment as well as in light of the standing committee in ability to reach consensus. And so, the avenue that was moved forward is – we did not force the best in class determination while NQF would prefer best in class to be determined.

In this case, we did not force that decision to be made, instead we put the conditions on endorsement and moved out four measures forward and has been having this ongoing discussions with the developers.

(Peter Thomas): Great. Thank you.

Lee Partridge: (Peter), this is Lee. Since I am on CSAC and was very deeply involve in this discussions. We had almost as much trouble with the CSAC level, with these measures and the whole concept, and we had at the standing committee level. And I think, when – if we hit the situation again and say three which I expect we will, we can probably bring back some of the comments that were raise both in the standing committee and then afterwards in the CSAC, to help us make those decisions. Because there was very good arguments advanced on both sides.

(Peter Thomas): Thank you, Lee.

Sarah Sampsel: OK. So with that, if no other questions, then we are working on finalizing the report for phase two. We do have to wait for the closing of the appeal process, but that is anticipated for mid-January, and you will be notified when that report was finalized.

And, you know, really what – staff were struggling with and challenged with is to appropriately documents and, you know, summarize and convey some of these challenges that were face in consideration for the future on these types of measures.

Moving on to off-cycle, which is was we – are working on right now. Obviously, you know, the next steps will be the voting, the closing of the conditions on Minnesota Community Measurement if that's what you all determine to do during the off-line vote. And then, the voting on the patient activation measure, some more information will be coming out of that next week.

But you should be prepared to see a draft report in mid-December that will go out for the typical public comment period. We do – we will need to set up a

post-comment conference call which will happen in late January, so we'll be pulling you for dates on your availability for late January.

And then, with this three which just kicked off, measure submission is currently open. You know, we were very fortunate to move forward with CMS in funding a third phase of this project. We'll just provide this a little bit of staff background and how we're thinking about measures for phase three.

First of all, we kind of have a backlog of measure that need to go through maintenance review that had been, you know, hopefully, we had wanted to look at in phase two but with 28 measures. We just didn't have the time or the capacity.

So there are some measures, I think there are couple of health outcome survey, kind of patient burden measures sitting out there that need to be reviewed. We know that we have some pain measures that are ready for review as a part of their maintenance review.

And so, we're prioritizing maintenance review during this third phase. At the same time, we still have developers out there, developing measures, really interested in achieving an endorsement status. And so we're working with those developers to identify what other symptom burden type measures quality of life measures. Any of those person- and family-centered care domains, we did not look at, that didn't fall under personal status and/or experience of care.

However, I think we still have at least one or two more CAHPS measures that need to come through for maintenance.

So we're going to clean up the backlog, and then – moving forward, we may have more functional status measures in phase three, but we really anticipate more symptom burden type measures in phase three, or quality of life measures in phase three.

More information coming on that but the measure submission period is open until March 31st. So it's going to take us a few more weeks to vet out what will be coming forward and what won't be coming forward.

Female: Sarah, we have some annual update stuff too?

Sarah Sampsel: I did believe we have annual updates. And then, I guess, the other thing that I should mention on that is, we do have new, some new criteria for the maintenance review of measures, and we'll be introducing that to you, and during the kind of Q&A calls prior to our in-person meeting, you know, we'd really like full participation of the committee so that we – and at that time, would have the fullest of measures for review.

Female: OK.

Sarah Sampsel: You can anticipate also receiving a poll to setup a meeting for sometime in June for our in-person meeting. And then the additional kind of overall global dates on how things will continues through our next summer, are stated on the slide.

I would also just like to mention here that both Suzanne and Kaitlynn will be transitioning off this project. So NQF is currently identifying new steps to – not necessarily new overall but new for this project. That will be you'll start hearing from – in the coming weeks and months. But would really like to thank and recognize Suzanne and Kaitlynn for their efforts, in staying on top of things and, you know, coordinating and managing all the work that we've been doing.

So with that, more information to come on both phase – work is off-cycle recall 2.5. So phase 2.5 is when, again, we'd really like to thank all of you for your active participation and really excellent questions and discussion as we've added these measures. I think we're done with easy button measures, globally and performance measurement, and you all are tackled with some of the most difficult measures that NQF is seeing.

So thank you and we're always open to considerations of things that we can be doing differently. So please feel free to reach out to staff and we'd be glad to discuss.

I would also like to mention that we do have opening on the person- and family-centered care standing committee. We're specifically looking for consumers, purchasers and health plan representatives. We also – NQF overall is looking at the other ways to integrate consumers and meet the, you know, sometimes special need these consumers, who don't feel us comfortable talking about reliability and validity, and some of the measures criteria.

But really do have some useful information that could be conveyed to this group, as well as any of the NQF group. So if you have any folks in mind, we'd love to receive their nominations and those nominations are currently open.

Lee Partridge: This is Lee. Before we close, I want to personally thank Suzanne and Kaitlynn for their help and we will miss you. And special to Sarah, because we've also realized how hard we'd work here for CSAC on some of these four measures that we struggled with the standing committee. That she probably put in more sleepless nights that she'd like to count.

So we are very glad you're staying with us, Sarah.

Male: Thank you.

Male: Yes.

Male: Thank you very much.

Male: Thanks a lot.

Male: Your work is amazing. Hard stuff.

(Helen Burstin): And thanks to the committee and the chairs, it's been extraordinary. We know how hard this is, so thanks for your patience with us as we kind of move through some of these newer measures that don't have a script.

Sarah Sampsel: Yes. And, you know, I thank you for that. And I guess, I should have mentioned, there was a transition in the off-cycle that we moved from two chairs to three chairs in welcoming back Jim to the committee. And we would like to continue this kind of three (co-chair) format. I think it works well for

this group. And special thanks of course to Lee, Chris and Jim for picking up some extra effort on this off-cycle project.

Male: If there are people who are interested in becoming a member of the committee, what do you recommend they do specifically?

Sarah Sampsel: So, there is – and we'll be glad to send out to send out that link to you but there's a link on the NQF webpage to this project, and then there would a link to the nomination page. But we can send that out to you, we'll have it.

Male: And deadlines for expressions of interest?

Sarah Sampsel: So, Kaitlynn or Suzanne, do you have that off the top of your heads?

(Suzanne Theberge): I don't off the top of my head but I think it is late January.

Sarah Sampsel: Late January, early February, seems to ring a bell to me.

Male: Thank you.

(Suzanne Theberge): We can send that link along with the survey that we're going to send out next week.

Male: OK. Thanks.

Sarah Sampsel: With no additional questions, again, thank you all. And we will be in touch next week, the transcripts will be available around mid-week. And we would also like to remind you that the materials are out on the committee SharePoint site if you want to rereview those. And I would specifically look at the Minnesota Community Measurements, updates to their measures, and those are separate Word document from their measure submission form.

And then with Insignia and the patients activation measure, we will follow-up with them to direct you to the most appropriate article to look at regarding your reliability-validity questions and then the adolescent data.

So thank you all and hope everyone has a fabulous weekend.

Female: Thank you.

Male: Thank you.

Male: Thanks, Sarah. Thanks, everyone.

Female: Bye.

Male: Bye.

Male: Thank you.

Operator: This concludes our call, and you may now disconnect.

END