NATIONAL QUALITY FORUM

Moderator: Person and Family-Centered Care
May 18, 2016
3:00 p.m. ET

OPERATOR: This is Conference #: 5359585

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

Suzanne Theberge: Good afternoon, everyone. Welcome to the orientation for the Person and Family-Centered Care Standing Committee Project, this is phase 3 of the work. So welcome to our returning committee members and our new committee members.

Before we get started, I'd like to just go through a couple of quick housekeeping reminders.

As with all of our calls, you do need to be dialed into the phone line if you wish to speak. The computer streaming version on the slide will only stream audio, you can't participate verbally.

We also do request that for those of you that are on the phone and on the webinar, please do turn your computer speakers off, so we don't get feedback. And finally, we ask that if you're not actually speaking, that you put your phone on mute to reduce background noise.

So, welcome to the call. I'd like to start by introducing the project team. And then we will have the Standing Committee, do a quick committee roll call. Sarah?

Sarah Sampsel: Sure. Hi, this is Sarah Sampsel. And I'm the senior director on this project. I have officially been at NQF actually just since the beginning of this year but have been working with NQF for a number of years, most recently as a consultant.

So I've been involved in all three phases of this work and really excited to see how this portfolio is growing, but also the work of the Standing Committee and how it's evolving and really helping NQF overall think about some challenges with measurement and how it applies to other types of measures as well.

Suzanne Theberge: This is Suzanne Theberge. I'm the senior project manager on the team. I'm excited to be back for another phase of the work.

Kirsten Reed: This is Kirsten Reed, the project manager here at NQF. This is officially my fourth week here. So this is my first project and I'm very excited to jump in and work with all of you moving forward.

Desmirra Quinnonez: Hi, this is Desmirra Quinnonez. You might see me in e-mails as (Desi). I am the project analyst on this project. And I'm excited to work with you all.

Female: All right …

Desmirra Quinnonez: At this …

Female: Go ahead, (Desi).

Desmirra Quinnonez: OK. At this time, we're going to go ahead and take roll call. So, if – when I call your name, if you could just let me know whether or not you're here, please.

I'll start with James Merlino? Lee Partridge?

Lee Partridge: Hi, I'm here and welcome to everybody, including our new members.

Desmirra Quinnonez: Hello. Christopher Stille?

Christopher Stille: Hi, this is Chris, I'm here. And again, welcome.

Desmirra Quinnonez: Awesome. Beth Averbeck?

Beth Averbeck: Here.

Desmirra Quinnonez: Katherine Bevans? Samuel – excuse me, Samuel Bierner? Adrienne Boissy? Rebecca Bradley? Jennifer Bright?

Jennifer Bright: Hi, everybody, I'm here.

Desmirra Quinnonez: Hi, Jennifer. David Cella? I know he's upstairs. Let's see, Sharon Cross?

Sharon Cross: I'm here.

Desmirra Quinnonez: Hi, Sharon. Dawn Dowding?

Dawn Dowding: I'm here.

Desmirra Quinnonez: OK. Nicole Friedman? Stephen Hoy? Sherrie Kaplan?

Sherrie Kaplan: Here.

Desmirra Quinnonez: Hello. Brian Lindberg?

Brian Lindberg: Here.

Desmirra Quinnonez: OK. Linda Melillo?

Linda Melillo: I'm here, thank you.

Desmirra Quinnonez: Hello. Ann Monroe? Lisa Morrise?

Lisa Morrise: I'm here.

Desmirra Quinnonez: OK. Elizabeth Mort?

Elizabeth Mort: Hi, I'm here, good afternoon.

Desmirra Quinnonez: Good afternoon. Lenard Parisi?

Lenard Parisi: I'm here.

Desmirra Quinnonez: OK. Debra Saliba? Lisa Gale Suter?

Lisa Gale Suter: I'm here.

Desmirra Quinnonez: OK. Peter Thomas? And Carin van Zyl?

Carin van Zyl: I'm here.

Desmirra Quinnonez: Hello. Thank you.

Was there anyone else who joined and did not hear their name? Or anyone who joined after I called their name?

OK. Thank you.

Suzanne Theberge: Hi. Thank you, everyone, for joining us. Before we dive into the agenda for the call, wanted to just mention that we have our three co-chairs back, Jim, Lee and Chris, and wanted to see if either Lee or Chris who are on the call today wanted to make any opening remarks before we get started.

Lee Partridge: This is Lee. No, as I say, welcome to our new members. It looks like we're going to have a lively time. And I think rather than take time today, I'd rather just let you proceed.

Christopher Stille: And I think I agree with Lee. So, thanks.

Suzanne Theberge: Great, thank you.

Well, we – as we just said, we do have a packed agenda for the call. I'm going to briefly go over the CDP, the Consensus Development Process, and then Sarah is going to talk a bit about the portfolio of measures. We're going to talk about the project timeline. And the roles of the Standing Committee, co-chairs and staff. And we're going to look at the criteria and the SharePoint.

So, for those of you who are returning, this might be somewhat familiar, but it has been a while since we were last together reviewing measures. So the project team wanted to go over some of the other information again.

So, next slide. As you all may recall, the Consensus Development Process has eight steps. We have completed the call for nominations and we have seeded you all as our standing committee.

We've also completed the call for measures. And we have 13 measures that we're going to be looking at in this phase of work, which is the candidate consensus standards review.

Following the review, which is going to take place at the in-person meeting on June 6th and 7th, we'll be putting everything out for public and member comment. And then we'll ask the NQF members to vote on your recommendations. We will bring the project to the Consensus Standards Approval Committee for their review, and then finally, the Board will ratify the project – the measures and we will close out with a 30-day appeals period.

We'll keep you posted. As you might recall, the bulk of the committee's work is right now in the measure review portion. And then we'll bring you back together after the comment period to discuss the comments that were received and draft and responses to those comments.

And I will now turn it over to Sarah to talk about the portfolio. Sarah?

Sarah Sampsel: OK. So, we have a slide in here that I'm not sure why we have in here. But, let me just mention briefly, actually go ahead and go back and let's talk a little bit about the Measures Application Partnership just because that is something that overflows with this committee.

I think many of you are aware and some of you participate in the Measures Application Partnership which is another arm or project area of the National Quality Forum.

And so, as part of the MAP process, NQF informs a selection of performance measures to help achieve improvement, transparency and value. So basically,

what the MAP does is provides input to health and human services, specifically CMS programs during pre-rulemaking on the selection of performance measures for use in public reporting, performance-based, accountability and other programs.

And what this Measure Application Partnership looks like is a number of committees, including a coordinating committee, there's a post-acute care, long-term care committee, duals, and actually some of these are workgroups, clinician and hospital. And basically, these workgroups provide the input for the government as well as identify gaps for measurement testing and endorsement.

The reason I wanted to bring this up, and we can go ahead and go to the next slide, is the fact that many of the measures that come to the Person and Family-Centered Care Committee are also going through the MAP process as well. And typically, the cycle – and in this MAP, this is thanks to Reva Winkler. But basically, what this MAP talks about is the fact that the endorsement process as well as the pre-rulemaking recommendations should be working together.

So, when we draw information on recommendations from the MAP and the MAP draws recommendations from the CDP process as well. And so how that typically happens is the government releases their measures under consideration or MUC list on an annual basis that's posted on NQF's website in early December, always by November – or December 1st, which means that we're all kind of screening right around Thanksgiving to make sure that the draft MUC list is posted.

In order to inform kind of measures that will be put into rules and pre-rulemaking for inclusion in the federal programs, a lot of times, some of these measures as an example during our last phase of work, some of the functional status measures had appeared on previous MUC list before coming to the endorsement process and then sometimes they come to the endorsement process before they go to the MUC list or go through the MAP process.

And what we're trying to create and what this picture is trying to signify and what you'll hear us talk about a little bit more are feedback loops, and ensuring that you all know some of the feedback that's provided when the measures are being considered for inclusion in programs. And then the MAP, on the other end, is considering measures that you have all looked at for endorsement.

The MAP process does not go into the level of detail that you all go into on the endorsement. So they're not looking at scientific acceptability, they're not looking at the importance and rigor, et cetera. They're actually looking for a best fit for program to meet the goals and demand to the program as they are specified by CMS.

Deb Saliba is actually – I know – I don't think she's on the call, but she is actually the co-chair for the PAC/LTC, where measurement workgroup or the measure advisory panel workgroup. So we do have a voice there. And between Deb and I who was also – who also staff the PAC/LTC workgroup, just trying to keep those connections in place.

So, we wanted to make sure that you all understood that this is going on and we certainly saw this play out in previous phases of work when it came to the IMPACT Act, because a lot of the measures under consideration that we're seeing as part of the MAP process are actually outflows from the IMPACT Act and some of that implementation.

And we can talk about this a little bit more during the in-person meeting and where those project flows are coming into place.

I will say that based on the measures that I'll go into in the portfolio, we have not – for – we do not have any measures under consideration for endorsement for this third phase of work that came out of the most recent MAP cycle of work.

Next slide.

So, what this project is, is it will evaluate measures related to person and family-centered care for use and accountability in public – and public

reporting for all populations in all settings of care. When we did our call for measures, and that call for measures had opened last October and then ended in March, we were specifically – we were actually trying to clean up and try to catch up with some delays we had had because of the number of experience of care and functional status measures.

During that six months period, we were made aware that there are a number of measures that were not ready for their maintenance review. There were some measures that we thought were coming that were new that did not come. So we ended up opening up this call for measurement to health-related quality of life, functional status, shared decision-making, symptom and symptom burden, and experience of care. And we originally thought we were going to have closer to 25 or 26 measures again, and as Suzanne mentioned, we have 13 and I'll go over those briefly.

But basically, we're going into, you know, still almost a clean-up phase, but at the same time, have the opportunity to look at some new and novel measure of approaches when we talked about shared decision-making and some of those processes with some new measures coming through.

We currently have more than 60 measures endorsed within this area. And one of the other things that we will have for you during our in-person meeting is the full measure list of the portfolio. I think it's helpful to have all of that, so we can think more about gaps and have an informed gap discussion. So that's on our to-do list for the in-person meeting as well.

Next slide.

So, of the measures that we will be considering in this next phase of work or this current phase of work, we have both measures for maintenance, as well as new measures.

We only have one new measure and that's the symptom and symptom burden measure, the measure number 0420. It's a pain assessment and follow-up measure. This is a CMS measure, but is – the developers that were working with on this are Quality Insights of Pennsylvania. And we've been having discussions with them so far and we will also be orienting you to kind of our

new maintenance process in how that will work and play out, and we'll talk about that a little bit more when we go through our criteria. But we do only have one maintenance measure and then the rest are new measures.

If we'll continue.

I wanted to also bring your attention that we had – as I mentioned, we originally thought we were going to have a lot more measures for review for this project. And this is an example of kind of considerable back and forth between NQF staff and the developers and stewards on readiness for evaluation.

Some of these measures, and specifically these, the C-CAT measures as well as the 0700, the health-related quality of life for COPD patients. These measures have not been through person and family-centered care. They're actually – they actually came out of other committees that are no longer standing committees. And when we started taking a look at them during the preliminary analysis process, really felt that they weren't ready for a full maintenance review.

So, we've started working with the developers in doing some technical assistance to get those measures ready so that they meet the current NQF criteria.

As you might recall during phase 2, we had a couple of measures that came through where I think committee members had raised issues on, are these really ready or not? So those are some of the standards we were applying here that we wanted to make sure that you all had the appropriate level and amount of information to make a clear determination and recommendation prior to getting to the meeting versus trying to do that during the actual meeting cycle.

So, on mentioning on the C-CAT, and the C-CAT is actually a cultural competency tool, it's an organizational tool. And they're composite measures assessing organizational readiness for cultural competency in communication. They recently transferred from the American Medical Association to the University of Colorado, and many of you know Matt Wynia who is the overall steward for these measures.

Actually, what we've offered to Matt and to his assistant, (Heidi), is to come to the meeting and on day two, they will be talking about the suite of measures and trying to get some early feedback from you as they start preparing and updating their submissions to go through the full maintenance review, possibly later this fall of – or early next year.

Next slide.

OK. So the next – this slide, in addition, we had a number of measures that we had been postponing maintenance review and these have been conversations back and forth with the appropriate stewards and developers. And so, when you see the final and full measure portfolio list, you'll notice that a lot of measures are losing endorsement and this full set the vast majority from AHRQ, and one measure from NCQA. And here, the communication is not that these measures don't meet criteria, but it's more the onus of the steward and with the management of urinary incontinence in older adults measure. This is a health outcomes survey measure.

We've been working with NCQA for sometime as they've been working with CMS and the full health outcomes survey. And they've communicated to us that after testing and trying to move this measure forward that – and keep it in the endorsement cycle that there's still additional work that needs to be done so they have asked for removal of endorsement while they continue to look at different approaches to measuring management of urinary incontinence.

With the AHRQ measures, all of these measures actually were first reviewed and, in all honesty, were probably endorsed more as surveys versus measures. And as you're all aware, and we'll talk a little bit more about, NQF does not endorse these surveys.

And so, because these measures are not in public programs right now and AHRQ just doesn't have the resources in order to devote the time to maintain measures that aren't being used, they still, obviously, support the survey and the survey is still a part of – the surveys are still part of the CAHPS Consortium, but the measure is related to the survey are losing endorsements.

Next slide.

So, these are the new measures and you'll see that we have those split up between – among experience of care so we have three nursing home experience of care measures from ACA, so the folks from ACA will be joining us in June.

We also have a measure, a CMS measure, that was developed and contractor was Truven, and that is the home and community-based services experience of care measure. This looks a lot like CAHPS, so we'll take – I think all of these measures will take everybody back who's been on the panel to phase 1 of our work, where we looked predominantly at experience of care measures.

We then have six functional status measures. And these, of course, will be reminders to everybody about phase 2 of the project. So, our colleagues at UDSMR wanted to have their measures for motor scores, self-care and mobility assessed for additional settings. And so they have submitted the measures for skilled nursing as – and as well as long-term acute care facilities.

And then finally, kind of these new and novel measures that we're really excited to see come through are the shared decision-making measures. We have a measure form Healthwise, the shared decision-making process Measure, 2962, and then one from University of Massachusetts and Massachusetts General Hospital, the informed patients in our hip and knee replacement surgery.

So, really looking forward to some conversations about all of the measures and how they all kind of look the same, different and you have the criteria work with each of those.

Next slide.

So, I will go ahead and turn it back to Suzanne.

Suzanne Theberge: All right. Thank you, Sarah.

I just wanted to briefly touch on the role of the Standing Committee and the staff throughout the project.

We've brought together you all as a group of experts to evaluate the measures and make recommendations to the NQF membership and our governing boards on whether the measures should be endorsed or not.

So, you will be looking at measures against the criteria, which we're going to go over in a minute, responding to comments and responding to directions from the CSAC.

Next slide.

We do ask that all members review all measures. And I just also wanted to note that your in-person obligation for this round of work is limited to the June meeting, but we might have to hold a follow-up call after that. But, again, everything after the June meeting would just be over conference calls.

We also ask you to oversee the portfolio, discuss harmonization issues, identify gaps and consider measure issues that arise throughout the year term on the committee, such as ad hoc reviews or off-cycle reviews, which I know all of you returning folks are familiar with.

Next slide.

We do have three co-chairs on this project and we ask them to help facilitate the committee meetings, meet with staff and assist us to achieve – you know, getting everything ready for the meeting, achieving the goals of the project, co-chairs, we ask them to keep the committee on track, and to represent the committee at CSAC meetings and then, of course, to participate as a full committee member.

Next slide.

The NQF staff is – our role is to provide all the background support to help you all do the work. So we organize and (staff) the committee meetings, the

conference calls. We provide guidance on the CDP on the key questions about the NQF criteria or about NQF policy.

We do the preliminary analysis of the measures. We write up the results that all of you are – discussions in the project report and, of course, we ensure that everybody is communicating with each other or is being kept in a loop for that upcoming projects work and project milestones.

Next slide.

Really, our – one of our biggest roles is communication. And that would be ensuring that you all have the materials that you need working with the developers to ensure that they are getting us what you need, communicating with the public and the NQF membership, et cetera, et cetera.

So, with that, I think we should just pause quickly here and see if anybody has any questions before we dive into the – our measure evaluation criteria.

OK. Hearing none, Sarah, back to you for the criteria overview.

Sarah Sampsel: Sure. Just go ahead and keep going. OK.

So – and I think I should also – we should also mention at this point, so this panel now has 25 members, seven of the members are new. And we felt it was important to hold an additional call for new members or nomination process for the Standing Committee for a number of reasons.

One, we lost three members that chose for various reasons not to re-up further terms which had expired at the end of last year. But also because in looking at the portfolio of measures, looking at how the portfolio has been changing overtime, the types of new measures that we are getting in, we really wanted to bolster some of the expertise areas.

One of the things that we're really challenged with, with the Standing Committee is that you really can't say that it focuses on one area like all physician measures or all health plan measures, or even all SNF measures versus acute hospital, et cetera.

So, we're really doing our best to balance the expertise knowing that we only have a finite number of members that we can include on the panel. And you know, that's where there's opportunities for public comment and other options for folks to provide and put on the process and having those public comment periods.

But with those new members, we hold a very kind of brief abbreviated orientation call not only to get to know them a little bit better, so that they could hear staff and then we did do a very brief overview of kind of the next steps, the expectations and the criteria.

So, I – you know, I'm not going to spend an entire hour going through the measure evaluation criteria. I would like to remind folks that, you know, as Suzanne just called, we really would like you all to be familiar with all of the measures. There's 13 of them.

I should have mentioned with that health and community-based services evaluation or experience of care measure, while it's one submission form, it's 19 measures. So, that's a lot to grasp. But, I feel the folks at Truven and CMS really did a nice job in filling out their forms and providing tables that could help pull that information together for you.

So hopefully, it's not too daunting. You will also realize that since we had only 13 submissions, we have about three or four people assigned to every measure. And we'll talk a little bit more about that process after we go through the criteria, but want you all to be thinking about kind of the process, what you'll be looking at and how – you know, where you might have questions with the criteria as I go through these next few slides.

So, just as a reminder, go back, please. Our evaluation criteria for endorsement are from measures for accountability applications as well as quality improvement. You know, as you all are aware, we really are moving toward more measurement for accountability, but there are still these measures for quality improvement that are used internally within organization. So we ask you to kind of, you know, be able to pull apart, and you find that

typically in the use and usability section on what the application for the measurement is.

The standardized – we do have the standardized evaluation criteria. One thing you will notice this year when you open your measure evaluation forms is that we have, I guess, really beefed up our preliminary analysis of these measures. So, you'll not only see some standardized language that NQF staff are using in trying to draw out of all of the measure submissions, but we're also doing some recommended ratings.

You do not have to agree with our ratings, we just wanted to put some food out – for thought out there. But we also wanted to recognize that there some folks on every single NQF Standing Committee that may not have the level of expertise in reviewing measures and specifically when we're going in through reliability and validity.

So we're really trying to help folks think through things the way that we are suggesting you go through the criteria and also through the algorithms.

Wanted to recognize criteria have evolved overtime in response to stakeholder feedback. I'll be giving you an update shortly on how, really, the work of this committee has gone to the CSAC and there've been additional conversations internally to make sure that the – that our criteria are working for all types of measures and where we've struggled in times in interpreting criteria when they haven't been as clear for the types of measures that we look at on this committee.

And then, again, that our measurement enterprise and specifically within NQF but across the standing committees is constantly growing and evolving, and we are using that growth and your expertise to evolve internally as well.

Next slide.

So our criteria are the same. We have importance to measure and report where you're looking at evidence and performance gap and disparities. This is a must-pass element in our portfolio and review this time, you know, the things to pay attention to are, there's a different between outcome measures

and the level of evidence required and the process measures, and we spelled that out in our preliminary analysis.

There are also opportunities in the event that it's an extremely important clinically construct, but perhaps, there isn't true guide – there aren't guidelines or (USPSTF) recommendations, et cetera, for a measure. There could be an opportunity for considering passing with exception.

Reliability and validity are the scientific acceptability of the measure, both of these, in addition, are must-pass. I think this is where you'll see our preliminary analysis has really been beefed up and where we're walking you through how we're interpreting the algorithms.

And I do – you know, for all new members and we do have screenshots coming up of some of the algorithms, but we have this printed it out for you at the meeting. And they really are the cornerstone and important for folks to pay attention to in the meetings. And that's also, again, where we draw from how we might have to evolve the criteria in those algorithms in the future.

Feasibility and usability, again, feasibility is about burden of measurement and what approaches there are for measurement, where a simple claims-based measure, obviously, tends to be quite feasible and easy to report when you go into surveys, which a lot of our measures are or patient-reported outcome measures, which again, we have a number of measures.

Feasibility tends to go down a little bit and that's acceptable, but it's something that we all need to consider.

And then, usability and use. So, our goal is to use for decisions related to accountability improvement and thinking about the overall usability and usability of the measure, and we're really starting to see as measures have evolved that if a measure is coming back for maintenance review and it's still not in standard use or the – you know, it's not in use in HEDIS or it's not in use in the federal program, then we really should be starting to question, you know, is use and usability challenged here, or what are the plans for this

measure to be implemented, therefore, should it be reported at the national level.

And then comparison to related or competing measures, of course, we have related and competing measures in this project. We have the functional status measures, as well as some of the experience of care measures that we'll have to discuss in our deliberations in June.

Next slide.

So – and, again, when I went over some of the points in the last page, I've already touched on some of these, but importance to measure and report is our first criterion. That's where you're looking for, one, the importance of the measure and that it is the evidence based.

You vote on things and we'll be looking for your feedback on the evidence that supports the measure, realizing, again, there's a difference in the level of evidence which you'll find on the algorithm between a process measure and then outcome measure.

We also ask for you to look for opportunity and improved – opportunity for improvement, and when you're considering opportunity for improvement, we ask that you consider disparities and care across populations group.

So, perhaps, you have a measure that is already reported at 98 percent, 99 percent in a broad population, but perhaps, if you look in a specific population whether it's a racial or ethnic disparity or perhaps in a different setting of care, if the developers provided that information, you know, maybe they're reporting at that level at 60 percent, which may be irrational to keep the measure because there are still areas for room for improvement.

Next slide.

Again, the outcome measures, we're looking for how the outcome is influenced by health care processes or structures. When you look at your evidence form, you'll realize this is only two questions on that evidence form and the rest of the evidence form is not applicable. However, a lot of times,

we do have developers that provide additional information evidence, et cetera. But, really, this is the overall rationale for how the outcome is influenced by health care process or structure.

And our question to you, as the committee, is, is there really a reasonable process structure or intervention that can affect this outcome being measured, and was the rationale stated so that you can understand that relationship.

And then the process – and intermediate outcome measures, this is where we're looking for quantity, quality and consistency of the body of evidence. This is typically where we're going to see some guidelines and, obviously, hope that those are current guidelines. This is where we're looking for USPSTF recommendations if they're available. And if neither guidelines or USPSTF recommendations are available, then we are looking for a systematic review of the evidence with a rating of the quality – the quantity, quality and consistency of that body of evidence.

And this is really important for the process measures. And again, this is in our preliminary analysis. When you start looking at your measure evaluation forms, you'll see that staff has really been spelling this out and showing where we got to how we got to our recommendation by walking through the algorithm for you.

Next slide.

So this is the example of the evidence algorithm. And basically, what you will find in those preliminary analyses is the readings from the algorithm section, where we basically go through step by step saying in box one, "Is this an outcome measure, yes", and you go to box two and your choices are pass and no pass. If it's no, you obviously go down to each of those boxes.

For those of you not – who have not been to one of our in-person meetings, again, we have this printed for you. We will draw your attention back to these throughout the meeting. And as you talk as the lead discussant for you measures, we ask that you reference this algorithm so that we can see how you're making your conclusions because when you vote on the measures,

these are also the steps that you'll go through in order to come up with consistent voting as well as, hopefully, achieving consensus.

Next slide.

OK. As I mentioned, we have an updated maintenance process and this is only applicable to the first measure on our agenda in June, that's measure 0420, which is the pain follow up – or the pain assessment and follow-up measure.

What happens with the maintenance measures is that we ask each of the developers when they are updating and submitting the measures for maintenance to complete a maintenance checklist. This checklist will tell us if they have changed anything in the measure, whether they have changed – if the evidence has been updated, and therefore, they are providing more information on the evidence form or updated information on the evidence form, if they have changed the specifications, if the specifications has been – have been a material change, and therefore, looking for testing results. And that will then draw our attention and be able to – it also changes our preliminary analysis and the analysis that you would all go through.

So, with evidence, same as with new measures, while we still are looking to see that there's evidence behind the measure, if they are changes in the measure, we would ask you to look at the changes in the evidence.

But frankly, there's a decreased effort – or a decreased emphasis, we're not asking you to rehash something that's already been hashed out by a different standing committee or even by yourself to the three years ago. So, in the event that a measure comes through that there's really no significant change to the evidence, then we will ask you to – if you want, to vote on evidence or not or have any additional discussion about evidence. And thus, there would be a decreased emphasis.

If the developer has told as that they have changed the – or the evidence has changed, whether there's more evidence, maybe it's new evidence, something else that they really feel strengthens their measure, then you would evaluate as

for new measures. And again, we will kind of prompt you during the meeting on what would need to happen during your conversation.

Gap, opportunity for improvement, variation and quality across providers, this actually becomes more important in the maintenance review. So we really want to see data on current performance and gap in care and variation either – whether it's across settings or within population group. But we're, again, drawing your attention to this area because we think this is what's really important after a measure has been out for a while.

You know, if a measure has been out and it's consistently performing at 82 percent and that 82 percent hasn't changed for five years, why not. Is it a problem with the measure and have those problems been identified, or is it an indication that, perhaps, the measure is tapped out at a lower level or whatever level that it's at. And so those are the types of conversations that we'll want you to have.

Next slide.

So reliability and validity, and recognize these are everybody's favorite in kind of trying to understand what we're looking for. And again, this is where in our preliminary analysis that we're really trying to help draw out the information and help to interpret some of the information that's provided by the measure developers during this submission process.

We'd also like to recognize, though, that, you know, NQF staff are also limited as – not only in our time, but in our expertise in some areas that we really do rely on the standing committees whether it's on the importance in evidence side, but really all of the criteria and on reliability and validity to help us understand what was provided or what additional information you might want.

So, in reliability, we're looking for precise specifications, including exclusions. We're looking for reliability testing. This is where – and I, unfortunately, didn't catch this, but reliability testing of data elements or a measure score. This is a change. Actually, we're looking for data elements and measure score.

For those of you who are on the committee in phase 2, you'll remember numerous times during our conversation, there are questions in the algorithm of, I want to see the facility level scores. I want to see the level at the measure reporting level and those scores, and the testing that supports those scores.

We've required them for every single measure coming through on this project. And in fact, you'll have – some of our measures, in fact, we're still waiting for the measure level results based on the feedback from that.

But, you'll also – and I'll talk about it in a few minutes, that this information and this requirement has gone to the CSAC to make sure it's required across measures.

And then on validity, again, we're looking for specifications consistent with evidence. Again, we're looking for validity testing at the element and score level. And we're looking for justifications to exclusions. You'll see that referred to us as threats to validity. So we're looking for analysis on exclusions, missing data, risk adjustment, if no risk adjustment, we're looking for a rationale. We're looking for identification for differences and performance. And then comparability of data sources and methods is applicable.

Next slide.

And this is – so this is just kind of the visual of, you know, when you're looking at a measure, you're not just looking at reliability, you're not just looking at validity, you're not just looking at feasibility. But in looking at reliability and validity, we really want to look at them both together, although both are must-pass criteria. If you fail one, you pretty much failed both.

So we are looking for that far right hand target that we are looking for measures that are both reliable and valid, and that interpretation that are both consistent in how they're measured. But also correct in how they're measured. And you know, that, to us, in order to achieve that target, is one of the reasons that we've really pushed for having the both data element and the score level testing provided for your review.

Next slide.

So the key points in measurement testing, and you – and again, when you see the slides, the slides are out on the SharePoint site, so available to you. All of this information is in your committee guidebook so you can reference it. And I see that, Sherrie, your hand is up. Do you want to ask a question?

Sherrie Kaplan: Yes. I didn't want to interrupt your flow. But I just wanted to make sure I understood back when you were talking about reliability and validity, and it – or actually related to empirical analysis while you're here, too.

So, you know, for some measures, there's not a whole lot known empirically, you know, they're fairly new and, you know, they haven't been really widely tested. And we've talk before and I know the NQF staff, at one time or other, has had – has entertained the idea of sort of the equivalent of FDA phasing, you know, sort of a – how mature is this measure, is this kind of an early stage measure versus it's been around like CGCAHPS for a long time.

And so, the question is, are you asking for empirical evidence for attribution and it relates to, not just asking for the units being compared scores, but also for the proportion of variation in a measure that's accounted for by the unit being compared, the hospital or clinic, the physicians, whatever.

And as opposed to the patient characteristic, and in one of the other committees, the sociodemographic characteristics adjustment, it's kind of one of the features there. You know, so, are we – I know this varies by how mature the measure is, but for more mature measures, are – is NQF interested in the proportion of variance accounted for by the unit being compared versus patient characteristics, and that would argue for, it's what you're doing to them that changes these outcome measures not who you attract.

Sarah Sampsel: And that's a really good question.

You know, I know – we are – so we are pushing toward really understanding that broader and understanding the unit being compared and thus, you know, and in all honesty, I think you're going to see some of that and some of the

results of that, specifically for the UDSMR measures because those are measures that are 20 years old and they've been collecting that data for sometime, and that's some of the outstanding data that we asked for.

But that – was also in response to, I think, some of the similar questions during phase 2 was a project. So – but then, you'll also notice and I think you even said in your statement that some of these newer measures, we don't have that data yet. And so, no, we're – we wouldn't have that same – you know, I can't say it's not the same standard but we're asking folks to be able to provide us with the data based on the maturity of their measure and if it's just from testing, they may not have the same amount.

And thus, we ask the committees to think about, you know, is what they provided good enough for what we're looking for that we could then, you know, put in the report or establish, we may be looking for more in the future.

Does that help?

Sherrie Kaplan: Yes, thanks. That helps.

Sarah Sampsel: OK. So any – so, you know, basically, what you will find in your measure algorithm is going through some of the questions on empirical analysis and where, you know, if it's a no-question, then you would go to a different point of the algorithm, et cetera.

But basically, what we are looking for is pretty substantial analysis to demonstrate the reliability and validity of the measure as it is specified. We don't have any measures in this project, but there was a project earlier this year, where we had measures tested and these were older measures, and unfortunately, made it through older processes. But basically, the testing that was provided was not the measure as it was actually specified.

And so, we really want you to put a critical eye to that and make sure that the information and the data that has been provided, really, is specific to that measure and not something else, or there aren't additional variables there. At the same time, you know, you may have additional questions on how the level of testing that was provided in the data use, et cetera.

Again, we try to pull that out for you, but you may want to dive deeper because we don't want to redo what the developers did, but just kind of draw those summaries.

Next slide.

OK. Again, you can refer to your guidebook pages 42 and 43, the algorithm is on page 44. These are just the examples that reliability is the measure scores, the proportion of variation in the performance scores. We're typically looking for – and we all typically find signal-to-noise analysis of – in the most – in most of these measures.

And so, in looking for, you know, why there might be variation in performance and making sure the developer understands what introduces variation in performance, you will also see data element testing. And so that's reliability of data element refers to the repeatability and reproducibility of the data. This is where patient level data, where the reliability of the measure score is typically using the units or the hospital or the post-acute care, et cetera.

The data element level is looking at the patient level. We have both inter-unit reliability. We have some inter-rater reliability. We see Cronbach's alpha a lot used here. And so you'll see a lot of those terms used across some of these different measures when you look at your measure evaluation forms.

What we did do and I think it's important to mention, the project team is doing the first review of preliminary analysis and doing a first run through. And then, I look at every measure and a lot of our measures, I have pulled in at least one other senior director to do an additional review. If it's a set of six measures, they may have only looked at one of the six because they were all very similar.

But, we are having additional methodologists look at each of these measures and we have been going back and forth with developers to get more information when needed.

And what we're trying to do is get at, you know, was it an appropriate method used. A lot of times, people go through the NQF criteria and say, "Well, they said I could do this", but if it doesn't fit for their type of measure, you know, we might have some questions about what they did.

And so, again, we have six measures where reliability testing is outstanding and that's because we've already gone back to them and said, "These measures have been around for 20 years. We think the committee is going to want more data." So, they're working on that.

Next slide.

Again, reliability algorithms on page 44, what I want to bring your attention to for both this one and the validity algorithm is you will see box number four was reliability test conducted with computer performance measure scores. Frankly, if it's no, we should have already caught that and gone out and ask for the data so that you do have it.

Of course, you might not agree with what we found or if we said yes, you may say, "No, it's really not what I was looking for", therefore, you would go down to the no.

But, again, we've tried to catch these measures and this is an example of frankly you can't pass reliability and validity if you do not have both that data element or the computed measure score testing, and of the data element level testing. And you will, you know, kind of be seeing that rolling out and reminders of that all the way through our materials.

Next slide.

So again, kind of key points on validity testing, for empirical testing, we're looking at the measure level that some analysis at the hypothesized relationship of the measure result to some other concept assesses the correctness of conclusions about quality. And at the data element level, the correctness of the data elements compared to a gold standard. And sometimes they have to tell us what their gold standard is at the measure score. A lot of times, what we see is a – some kind of correlation analysis of the specific

measure against other measures looking at similar outcomes. But again, that you'll find our analysis of that and the kind of the conclusions of the developers in your preliminary analysis and your measure worksheets.

Face validity is acceptable and face validity is a subjective determination by experts that the measure appears to reflect quality of care.

Be honest, if you were not seeing a whole lot of just face validity anymore, again, it is something that's acceptable. But, we're seeing the trend away from that and really more towards correlation analysis of the measure. And I think in the measure set that you'll be looking at, there are few that have face validity results in addition to empirical testing, so not just one. And so, you know, we're kind of excited to see that.

Next slide.

Again, algorithm and as with the reliability algorithm and the evidence algorithm, these are printed out for you and will be at your seats in June. Just as a reminder, as has already been identified in this presentation, the criteria and these algorithms still say you either have to have the measure score or data element. We are requiring that measure score level testing at this point.

So, frankly, you don't get a pass if only data element testing has been done.

Next slide.

Another big section is the threats to validity. And there are conceptual threats that the measure focus is not relevant, or not strongly length to a relevant outcome.

There's the – if the measure was not proven to be reliable, then typically, it cannot be valid. Perhaps, patients are inappropriately excluded for measurement. There are differences in patient mix for outcome and resource use measures, measure scores that are generated with multiple data sources or methods, and missing data all introduce threats to validity.

The measure forms and measure submissions that the developers have all filled out are checked to make sure that all of this information is provided. There are times when the level of information on threats to validity is not there. But this is where the developers have an opportunity to have, you know, get the feedback from you as well and perhaps, they have that data somewhere else. Perhaps, it's not necessary or it's not something you're as worried about, because they've explained why they don't have the data there.

But, again, this is part of the overall validity of a measure, and therefore, part of the overall assessment and voting when we get to our ratings.

Next slide.

OK. Again, this refers to the new maintenance process. So, with scientific acceptability new measures, everything we went through are the expectations.

With the maintenance measures, there is – the measure specifications, there's no difference. We require updated specifications. We require folks to tell us when they've updated those specifications. And it's actually required annually.

What we're looking for, though, is material changes. So, if somebody, you know, if a code goes out of service or something happens and you wouldn't technically re-specify nor would you retest a measure. But, let's say, you add an age group or do something else that is a material change to a measure, in that case, we would look for updated testing, and updated specifications.

The overall reliability and validity of measures. Again, this goes back to material changes in the maintenance review.

If there've been no material changes to the measure, then we really don't ask you to spend a lot of time reconsidering reliability and validity.

That being said, we are looking to ensure that the measures are reliable and valid, and that is something that's being caught at the staff review level if measures, perhaps, wouldn't meet the current criteria, that's the test. Do they meet NQF current criteria?

But, you know, as – again, as measures are going through now, they've gone through last year, they've gone through two years ago, those are all measures that are really on current criteria. So, when you see those again in a year or two, you may not see updated – or you may not spend as much time with the updated testing.

Again, in the measure evaluation forms, we clearly point out, this is a maintenance measure, this is a decreased emphasis area. However, as the committee, you're still welcome to ask those questions and make sure that the measure meets the requirements for maintenance review.

Next slide.

OK. And Lee, when we get through this, if you have any additional comments, be glad to have you provide additional feedback from the CSAC here.

Based on the last couple rounds of Person and Family-Centered Care standing committees and our projects, a number of issues have come up. And we've been trying to listen to you and, you know, Helen is obviously engaged in these projects as well. In trying to figure out kind of, are there certain issues that we need to be thinking about with the NQF criteria. And I've talked about that a little bit having to do with the data element versus unit or measure score level testing.

But, what we wanted to do was have an additional conversation with the CSAC as well as internally to make sure that we're interpreting things appropriately.

So, the first issue that I want to remind you all or acclimate you to is that the fact that NQF does not endorse surveys, tools or instruments. A lot of – there was some interpretation that – and it's actually publicized on certain websites that NQF has previously endorsed the CAHPS surveys.

We're not endorsing the surveys. We're endorsing the measures that are derived from the surveys.

And some of the confusion we think that has come out of that is the fact that – and you'll see which – with the home and community-based services experience of care measure, it doesn't make a lot of sense for us – to us that somebody submits 19 forms for 19 measures out of one survey.

If they all have the same denominator, they all have the same kind of structure and there is a way to present the information on all of the criteria in one form, well, it may not be pretty, you know, there may be opportunities to provide some additional information in order to ensure we have the right information and not fill out 19 forms.

And so, that was one of the things that we received. And the CSAC kind of agreed to is that, that's true, NQF still does not endorse surveys, tools and instruments.

And so – but that translation doesn't always get all the way across, and we also think one of the things that was happening is that you have to remember that the surveys, tools and instruments are our data elements in these types of measures.

And so, when we were going through and reviewing and approving measures at only the data element testing level, then that was almost an inference that we were endorsing those surveys or tools.

And so, I just want to be clear, that's not what we're doing, that's why we're looking for measure score levels. While there may be some interest in endorsing surveys, tools and instruments, right now, we don't do it.

And go to the next slide.

So, basically, we had a discussion with the CSAC in March of 2016. We're continuing the policy of no endorsement. What we are doing is continuing to provide technical assistance to developers.

You actually have, again, at least, one measure in this portfolio, where the developer is going to be given opportunities to update their description to

make sure that the description isn't all about the survey, that we talk about the measures.

And you know, staff are continuously working with committees not only person and family-centered care, but other committees are starting to see survey-based measures as well, and our functional assessment-based measures as well to make sure that we're clearly communicating that we're endorsing measures.

I'm going to break.

Lee, any other big thing that you think came out of the CSAC having to do with the endorsement of surveys?

Lee Partridge: No, nothing big. I think it's difficult for lots of people to say, "Well, the data here is the answers to these three questions." And not say, "Well, we've actually endorsed the survey that incorporates those three questions."

And I think that's, to some degree, the source of the confusion. But, we keep trying to explain to people that the survey is the tool with which you collect the data on which the measure is based. I don't know that that really helps.

Sarah Sampsel: Sherrie?

Sherrie Kaplan: Let me see if I can – if I understand. I think I understand, but I want to make sure I understand. I think what the – so, for example, the internal CGCAHPS says there – I think it's a three item, physician communication, sub-dimension of that, cluster of things that are measured in CGCAHPS.

And when you attribute sort of between – the rest of things are may or may not have anything to do with physicians who are providing care, for example, large multi-specialty clinics, like the front office or whatever, or an academic medical centers.

So, if you're talking about attribution, you're approving the measure for physician communication and that might be different from the front office staff sub-dimension for the CGCAHPS questionnaire.

Is that what you're saying? You didn't – you wouldn't endorse the whole CGCAHPS instrument. You'd endorse the measures that are used in that instrument. And then they would independently be tested for attribution?

Sarah Sampsel: Correct. Yes. You know, and I think that's, you know, kind of what we've heard more than once is, oh, well, you endorse – and, you know, where we're seeing it more is really – I don't think I've really heard it was CGCAHPS as much as I've heard it maybe with hospital CAHPS or the in-center hemodialysis CAHPS where, you know, and there's been the assumption that we have endorsed the entire survey.

But if you actually look at the survey, there might – or, you know, there might be 65 items of which some items are grouped to, you know, just like you said, to provide the communication measure, which would then be reported on dialysis compare. And so, we really are looking for that full kind of follow through of being able to track the measure and where the data elements are the items. And then back to the specific facility wherever it's been tested.

Sherrie Kaplan: Thanks.

Sarah Sampsel: Linda?

Linda Melillo: Yes. Hi. So, I understand that you're not – we're not endorsing a survey. But, if in the prior example, you have three questions in the physician section and they're all being submitted together, do you then also look at the data to – that it would establish each item as contributing to the overall reliability of that – of this survey tool, or of that section, or, are we looking at just each individual question by itself?

Sarah Sampsel: So, that's not always a clear-cut answer based on how some of these measures are put together. So, typically, in a CAHPS survey, you might have a scale measure and that scale measure is, you know, might have multiple questions. So your communications scale measure might have four questions.

And then, yes, I mean, we're looking for the – you know, we would want to see some level of reliability of those specific questions, because those are the data elements. The items are the data elements.

And then when you get to the overall scale measure and, you know, then, again, you're looking for the data around that scale saying that those three questions lead in to – or that that scale is reliable overall.

Linda Melillo: OK. Great. Thank you.

Sarah Sampsel: And you know, that's – it's not always, again, you know, that you also have kind of more global measures, which is just one question.

And so, you know, that leaves directly to our measure. So that is the data element and the measure score.

And you know, it's not always really easy with some of these surveys and there's always a lot of numbers involved. But, you know, that's also why we're here, as staff, to help you.

Linda Melillo: Great, thank you.

Lee Partridge: Sarah, this is Lee again. I don't want muddy the water here. But, in our discussions, particularly, I think in phase 1, we had some request from CMS in particular to also endorse the measures, but, how they were – the manner in which they were reported. And I assume we're going to stay away from that.

What I'm talking about is, for example, if you're using a CAHPS survey, some people collapse the scales. They'll say, we're looking at people who – the number of people – some will say, "We're looking at the number of patients who scored 10", scored this doctor as 10.

Others will say, "We're going to use the survey to report doctors who scored either at nine or 10. And I think there was some discussion about whether NQF should also get into that area. And as my understanding, we absolutely don't.

Sarah Sampsel: That's correct. Yes, we're still staying out of that.

So, the other thing I'll mention here and it's something we can certainly get out to folks. And I think, Linda, your question goes to this a little bit, too. And

this is one of the conversations that staff has had with Helen Burstin to make sure that we're on the same page is, when – whether it's a tool-based measure or a PRO-PM, there is some level of expectation that that tool is reliable and valid before it ever turned into a performance measure, which is why we really want to be focusing on that measure score level, that data element is important.

We would hope that that tool was – has already been proven to be reliable and valid. We know there are reams of data typically showing the tool are reliable and valid. And I think this even goes back in some ways to your question earlier, Sherrie, is that, if it's a brand new instrument, you know, we probably are going to see in our submissions that, you know, the testing that was done on the actual tool and then the translation into the measure score level.

Whereas somebody's older measures, there might be 20 articles that already say, "This measure, you know, this tool is reliable and valid." And that is our assumption, that it comes to us with a reliable and/or valid tool, but still understand that you probably want to see some of that data.

OK. Next slide.

OK. So, this is where – so this issue is the NQF endorsement criteria been interpreted to set a higher standard for PRO-PMs where testing at both the data element and measure score levels are required.

Technically, this is how the criteria are written right now that the PR – that only for PRO-PMS are both of these levels is required.

So, you know, as you remember from our last conversations, the CAHPS-derived measures were required to submit testing at both instrument scale and that data element level and the performance measure that aggregates the patient level data, which is what we've already just been talking to.

In – or talking about – in addition, the developers of functional outcome measure, so all of those, the care-based measures, the FIM-based measures that we looked at last time were strongly encouraged to present scientific acceptability at both data element and measure score levels.

And you might remember in phase 2, and it's even in our report, kind of the challenges and the discomfort members of the committee expressed when we didn't have performance measure scores at the facility level, both to see variation, both to see the reliability at the way that the measures are being reported.

But, functional status measures aren't PRO-PM. So, you know, I'll talk to say as staff, I felt kind of caught in the middle of, well, but this is what the criteria say.

So, this has been part of the genesis of this conversation as well as to say, whether it's a tool-based measure, a PRO-PM, we're looking for that data – you know, both the performance scores and the scientific acceptability at the measure score level. We endorse measures. That is kind of why this (builds) upon, we don't the surveys, we are endorsing the measures.

Next slide.

As with this conversation with the CSAC, the CSAC also approved that the PRM – the PROM – that thus (fit this) into a PRO-PM or tool is considered the data source for the performance measure.

As such, we need testing – well, the testing of the tool and reporting and reliability, the tool can assist in establishing scientific acceptability. We kind of expect that to be established before you turn something into a performance measure. And it provides information for the Standing Committee, but we require reliability and validity testing at the performance measure level.

Our algorithms have not been updated, the criteria have not been updated and released yet. It's a project we're working on. But, you know, that's something that we'll continue to remind you, but I don't think this committee really needs reminders on it at this point, because it was something that you all expressed challenges within the past anyways.

Next slide.

So, just what we're doing internally, again, clarification of materials, that's not an overnight process. We also have some other changes in NQF processes that are coming down the pike. And so, we're doing that – this all at once. And hopefully, by August, everything will be updated.

So, you know, not only the NQF criteria, but any of the public-facing documents, internal staff education to make sure that this is consistent across committees. This committee has typically been the one that sees almost all of the tool base and PRO-PM type measures, but that's changing.

And therefore, everybody needs to understand this. And then, there has to be education across standing committees, because, you know, otherwise, we'll start having, you know, kind of measures from other (subject fee errors) coming to this one. And I'm just not sure that we can support that all the time.

Next slide.

OK. So, with feasibility, and again, just kind of jumping back into the criteria and I've talked about this a little bit. What you're looking at in feasibility is the extent to which the required data are readily available, retrievable without undue burden and can be implemented for performance measurement.

I think one of the conversations this group has had in the past and actually, most committees have this at some point is, you know, kind of what are those considerations for feasibility, you know, some of it is training, you know, what does it take to train staff for a clinician-based assessment tool?

What is the time and money allocation for distribution of surveys and then the appropriate turning of those surveys into measures? What about conceptual rights and if a measure or a survey or tool or the data collection means are not publicly available?

These are all considerations for feasibility that would come out in your rating. So, if something is an administrated claims-based measure, that data is readily available, you would typically rate that as high. I think most of the measures in these projects, we've rated as moderate, although some people might think that they're lower because they need extensive training.

Feasibility doesn't stop a measure from going forward, but it's an indication that there are some questions. And certainly, in a maintenance review, would want more information from the developer on, you know, what's going on here, why aren't people picking it up.

I've certainly worked with surveys in my past where the survey was retired, because of the fact that nobody was picking up, because it was just too expensive to deploy.

So, those are considerations that you should be making when you're looking at measures and talking about measures for feasibility.

Next slide.

And then usability and use. Again, this is another one of those criteria that becomes more important during the maintenance process, but at the new measure process of which most of our measures are. And that review process, we want to know, are the measures accountable and transparent? Are performance measures being used or is there some plan used for accountability? Or, you know, at least publicly reporting. Or, are they being used for improvement and what's happening for improvement.

So, if organization comes – brings the measure forward and says, "Yes, we're using this measure for quality improvement, well, then, tell us how it's being done." And you know, give us the specific program, what's happening and what has been shown overtime. Typically, on a new measure, you may not have a lot of this data. But you should at least have plans.

And then, benefits outweigh the harms. Again, this is a question that is in the submission forms that we want some kind of discussion or consideration that – and this again becomes important in the maintenance review as, you know, what do you know after implementation of this measure and what has changed, or if nothing has changed, does the benefits continue to outweigh the harm of measurement or consequences of measurement.

Next slide.

I really already mentioned this, where for maintenance measures, you know, there's really – there's an increased emphasis on usability and use. New measures, all of these are important. They're part of your voting criteria. But they are not must-pass elements.

Next slide.

We, you know, I think – so this will be new to the new folks on the committee. But those of you who've been on the committee know that this project tends to have related and competing measure issues. And this is where we're looking for across measures and the criteria if measures are related, where you would have the same measure focus or same target population. And if they're related, you then assess if they're competing, which would be the same measure focus and the same target population.

For this project, we have – let's see, the core (Q) measures which are skilled nursing facility, customer or experience of care measures. Those are – those were identified as related to the AHRQ or the CAHPS SNF measures. But since the CAHPS SNF measures are no longer endorsed, we will not have to have the related or competing conversation. But that's typically what we're looking at, their experience of care measure so it's the same focus areas and the same target population.

So, technically, those could be competing measures. And then, what we'd ask you to do is drill down and say, "OK, yes, either exactly the same focus areas and the same target population. This is what we recommend for harmonization."

But with the CAHPS measures retiring, we don't have to deal with that. Where we will have to have these conversations are with the functional status measures, and that would be looking at to the FIM measures and specifically the mobility and self-care measures against ACA care measures, again, looking at self-care and mobility.

And we're going to prep a little bit more before the in-person meeting and talk to Lee, and Chris and Jim on how we can make sure that that conversation is

meaningful and useful to kind of get to an end result. Because while we want the committees to have the discussions on related and competing, discuss if there are harmonization opportunities, and discuss if we can choose the best in class or a superior measure. You might recall that we were unable to do so during the last project.

So, you know, we'll want to kind of tee that up a little bit more and have those discussions at the in-person meeting.

I will mention, though, is on the committee's SharePoint site, there is a related and competing document that goes over the related and competing issues that you might want to familiarize yourself with.

Suzanne, this is you.

Suzanne Theberge: Sorry, trying to come off mute. We are up to SharePoint, correct?

Sarah Sampsel: Well, we're on the evaluation process.

Suzanne Theberge: Oh, no. Sorry. I'm losing track here with my internet issues.

Yes, just to talk briefly about the evaluation process. The committee – as has been mentioned, the staff have prepared the preliminary analysis of the measure submission. And that, we ask you to use that as the starting point for your reflections.

We have assigned each of you a couple of measures to do your individual evaluation of – and we ask to you to complete the preliminary evaluation and survey for those measures.

And the – for the groups, the folks that are assigned to each measure, as – the people who are returning might recall, you will be leading the discussion at the in-person meeting, meaning that you'll be charged with kicking off the discussion and really taking – hoping the committee take the deep dive on that particular measure or two.

Next slide. And just a reminder, I think we went over this earlier. The entire committee will discuss and vote on a recommendation for each measure, again, each of the criteria once we get to the in-person meeting.

So, I think we are through the criteria portion. And we can stop and see if there are any question before Sarah talks about the SDS trial period.

Sarah Sampsel: Sherrie?

Sherrie Kaplan: Hello?

Sarah Sampsel: We can hear you.

Sherrie Kaplan: Oh, OK. Sorry, I just took it off of speaker.

So, just to make sure I'm clear because I'm one of those deep divers who's going to do the – one of the start-off presentations at least for one of the measures. So, in the reliability terms, it's not adequate or it's insufficient, necessary but insufficient to demonstrate reliability, for example, in functional status. You know, very – a lot of these measures had been around for 34 years. They're very reliable to patient level, but now NQF is asking also that reliability be compared at the unit level, whatever is being compared, facilities or clinician – whatever is being compared that the measures developers provide some evidence that the – of reliability at the unit level. Is that right?

Sarah Sampsel: Correct.

Sherrie Kaplan: Thanks.

Sarah Sampsel: Yes, and I'll just – you know, we'll mention there, and I brought it up before. So, there are six functional status measures from UDSMR and, you know, even though they already have three measures that came through NQF and are endorsed. The six measures when they submitted their data, they provided some of the, you know, kind of some information about variation at the facility level, but not the level of reliability testing that we are seeing from other developers.

And so we were concerned that, you know, should they not provide that data, knowing they have an incredible amount of data that, you know, they should have the opportunity to provide that, and in fact, that we would want to see that at the unit level. Otherwise, it could be an adverse determination on their recommendation. And so, UDSMR has been working, you know, very diligently in order to get that data to us and we expect it next week.

So, those of you who are assigned the UDSMR measures, I think those are 2769 and then 74 through 78. That data will be coming next week and you'll see notes in the preliminary analysis from the staffing. We're looking for that data. You can look at everything else at this point, but we expect that data next week.

OK. Let's move on. Keep going.

So, as some of you are aware, we have been having some ongoing discussions within NQF and then appointed a social demographic – or socioeconomic status panel to consider how outcome performance measure should be adjusted for SDS and SES and other related demographic factors.

You know, you may also be aware there are two divergent perspectives on adjustment. Some folks will say that adjusting for sociodemographic factors masked disparities. Others say, it's necessary to avoid making incorrect inferences in the context of comparative performance assessment.

You know, I will be very clear, you know, I came out of NCQA, we didn't adjust. So, you know, for me, this is a learning experience as well. But based on the expert panel, they recommended and the Board approved the three-year trial period during which the adjustments of SDS factors will no longer be prohibited. So basically, prior to this trial period, we suggested that SDS factors not be included in outcomes measures. But now, during this period, there are some options involved. And that period actually started in April 2015. So actually mid point to the last project.

Next page.

So, basically, what we're asking you to do is assess each measure individually to determine if SDS adjustment is appropriate. We're not saying that all outcomes should be adjusted but, you know, we would like you to consider it and make sure that if the developer is adjusting that their analysis and rationale for the adjustment is appropriate, and if they are not adjusting that we were – we're looking for a conceptual basis, so our rationale as to why they aren't adjusting.

We also need the conceptual basis and empirical evidence for when they are adjusting. So we're looking for a lot more information.

We also, you know, know and recognize that the SDS adjustment must be constrained by the – maybe constrained by data limitations and data collection burden. And so, you know, we're kind of still going through this process and some developers are working on this, there are others doing, you know, kind of coming up with and, you know, kind of in their analysis helping us learn on different ways to apply this criteria.

Next slide.

So, basically, the scope of this is newly submitted measures. All measures submitted after April 15th of last year are part of the trial periods, that includes this project. The Standing Committee may consider whether such measures are appropriately adjusted as part of their evaluation. And then, with previously endorse measures, the – if they're undergoing maintenance review, they're considered fair game.

And you know, what I'll say here is that in this project, we only have one maintenance review, it's not risk adjusted so it won't be part of it but all of the new measures. These are some of the considerations because I believe all of the new measures are outcomes measures.

Beth, I'll get to you for a quick – in a second.

Next slide.

So, basically, and I think I had skipped ahead before, but we want you to continue to evaluate the measure as a whole including the appropriateness of the risk adjustment approach.

We want you to use the validity criteria to evaluate the appropriateness of the SDS factors, and there are questions specific to the committee in that section, as well as clinical factors used in the risk adjustment model. There are a number of times where a developer have said, "No, we're not risk adjusting this", and – but they tell us why. And we ask you to assess it, you know, based on your expertise, is this appropriate? You know, this is rationale work for you, or do you need more information.

Again, we've completed the preliminary analysis. We've identified areas where we asks you to focus on this, and really rely on you as the expert to help us understand how this plays out.

Next slide.

So, these are the questions, is there a conceptual relationship between the SDS factor and the measure focus, what are the patient level SDS variables that were available and analyzed? And did they do so? Was the empirical evidence and does it show that the factor has a significant and unique effect on the outcome in question, otherwise, that should be part of the rationale. And then, does the reliability and validity testing match the final measure specifications?

Next slide.

And this kind of – just kind of goes over what we've already talked about, but these are kind of some of the additional questions that you should be looking for in the forms. The conceptual relationship, if the factors are present at the start of care, therefore, you know, can you identify them and measure them and then is – or the SDS factors caused by the care that's being evaluated.

Next slide.

So, what you should do is review the patient level of sociodemographic variables that were available and analyzed if they were provide. We also want to know if those variables were available and analyzed and that they align. And are they generally accessible.

So, again, you know, a lot of this goes back to reliability and validity and we're just asking you to go beyond the measure specifications. And also consider if the measure is being suggested for risk adjustment that all of the variables were also analyzed to understand how they play out and how they impact the measure.

Next slide.

I think – so I'm going to skip because I feel this is redundant, that this information is in the preliminary analysis and what I actually already talked about.

So, again, there's – and there's a link here and the slides are available, but we did look during the preliminary analysis and ensure that if their stratification risk adjustment, et cetera, that the appropriate information at reliability and validity level was provided mostly in validity. If there are SDS variables that there's information required that was provided in order to stratify and that's all part of the measure logic questioning. And in addition, if the developer says the measure is not risk adjusted, there is that possibility of saying, "OK, there's a rationale provided and we'd be asking you to say, does that rationale makes sense?"

Next slide.

OK. So, questions, and I'm just going to be honest with you that I may not be able to answer all of your SDS questions because we haven't really dealt with it a whole lot on these projects. But, certainly, could get back to you and identify where we might need to do more education at the in-person meeting.

So, Beth?

Beth Averbeck: Yes, thank you. Good overview. One of the questions I had is, are we – was the committee considering when do we segment the results by similar populations versus risk adjusting, is that part of the consideration specifically around risk adjusting?

Sarah Sampsel: I think what you'll see – I think there are a couple measures where they're doing stratification instead of risk adjustment and that would be part of the rationale that they would say – and typically what I've seen is the developer says, "You know, we're not applying a specific or any type of standardized risk adjustment approach, but we do suggest you stratify results by, you know, X, Y or Z." So, yes, you would be looking at that as well.

Beth Averbeck: OK. Thank you.

Sarah Sampsel: Any other questions?

So, realize that's kind of a quick dive, want to always refer you back to your committee guidebooks, you know, with the caveat knowing they're being updated as well as the NQF criteria are being updated, but I think we touched upon the high points there.

We will also, on our question and answer call, which we'll be talking about in the next steps in couple of minutes, walking through what a measure evaluation looks like. We will not be using one of the measures in this project. But that's what we do on our Q&A call is talk a little bit more and dive in a little bit deeper.

You know, that's not a mandatory call, at the same time, if you want to refresh your course on how we go through the meeting process, et cetera, that's why that call is so important.

So, I'll go ahead, and if there are no additional questions, we'll go ahead and move on.

Suzanne Theberge: OK. So, I wanted to take a few minutes and talk about SharePoint. We know a number of you had some issues accessing SharePoint over the last couple of days. There was an upgrade to one of our systems over the weekend

and I think that caused some trouble but, it's our understanding that is been mostly fixed and that you should all be able to access SharePoint now. So if you can't get at the measure forms, please let us know and we'll see what we can do with. It should be working again, and we apologize.

So, what you should be looking at on the screen is a screenshot of another project homepage. That this is what – if you have another chance to look at it yet, this is what the (PSCC) SharePoint case looks like …

Sarah Sampsel: Hold on – Suzanne, hold on.

Suzanne Theberge: Yes.

Sarah Sampsel: Will you move to the screenshot? Thank you.

Go ahead, Suzanne.

Suzanne Theberge: OK. Thanks.

So, up top, you got the general documents and that would include things like the committee guidebook that Sarah has referred to, the criteria guidance and the algorithms, things like that, will all be up there.

In the body of the page, you've got all the measure submission for measure information forms, and below that, there would be the meeting and call documents. So we'll list out each meeting and we'll post the agendas in the slides and whatever else you might need for a particular meeting down there.

Over on the left hand side bar, there's – you'll see committee home up at the top and then, the surveys are about halfway down the committee preliminary evaluation survey. You can click into that and then select, there'll be a menu to select which measure that you're completing the survey for.

Next slide.

That I think is a measure documents page. What you'll see is just the measure number, title and description. And then when you click on the title, it will open up. You can see on this, there's a little plus sign, if you click on that, it

will open up another folder and that will give you the measure worksheet with the preliminary analysis. And any associated measure files that we were unable to add into the measure worksheet.

So, if the developer submitted something as a PDF or as an Excel files, it would be posted separately.

And finally, just to talk you through next slide, we can talk to for a second, what's in the measure worksheets. It will start off with the preliminary analysis, that's the staff review. And you'll find links throughout that that will jump you to various points in the measure submission or will take you to the other files that were submitted.

Below that, there will be pre-evaluation comments that were submitted by members of the public or NQF members that will show up in a purple box and then – then we'll have all the information submitted by the developer in order, first the evidence and then the gap and then the scientific acceptability and so on and forth.

So that's how those worksheets are laid out. If you have any trouble …

Sarah Sampsel: OK, Suzanne, can I – Suzanne, can I comment a couple of things so people are aware of?

Suzanne Theberge: Yes.

Sarah Sampsel: So, when you open a measure worksheet, you know, it starts – there's a brief overview that the developers filled out, which is really the description, numerator, denominator, what kind of measure it is.

And then when you get into the preliminary analysis section as when you'll see the jumps or the links, you know, to the evidence form, to the testing form, to other attachment and documents, and anytime you see a link, and this is something I never knew, you can link – you can go ahead and click on it, go to that link. And if you want to go back to exactly where you were, you hit Alt in your left arrow.

And maybe everybody knows this, but I never knew this, and you'll go back to where you left off and it's really saved me a lot of time.

The other thing that I'll mention there and I mentioned it in the very beginning is you will see staff analysis now. And I know we did staff analysis during the last phase of the project, but you're also going to see some staff preliminary ratings.

Again, you don't have to agree with everything that we said. It's just how we followed the algorithm through whenever we followed in a certain way that we think somebody may or may not disagree with, we have provided a rationale as to how we got to where we got.

But, again, it's just kind of to generate discussion, to give you all a leg up, and it's really up for you all to follow the algorithm based on your interpretation and discussion when we get to the in-person meeting.

Sorry, Suzanne, go ahead.

Suzanne Theberge: No, I was just going to finish up that portion by saying, let us know if you're having any trouble with accessing SharePoint or anything like that. Please do let us know.

And I think the last item on our agenda is next steps and timeline. And Kirsten, I will turn it over to you.

Kirsten Reed: Great. Thanks, everyone.

So, couple of things coming up, one week from today, we will have our measure evaluation Q&A call. This call will really be used to answer any questions that may have come up during your review of the measures. And then we'll also go over a preliminary analysis more in depth to kind of help you along.

In addition to that, your surveys of the two measures that you were assigned are also going be due one week from today on May 26th.

Then we'll have our in-person meeting on June 6th and 7th. As a reminder, you should have received a link to RSVP from our meetings team. And then it was also in – I sent a link in the e-mail that I sent to you guys last week. So if you have not received that, please let me know and I will resend it so we can get a final headcount for the in-person.

On June 15th, we are scheduled for a post-meeting webinar. This is really used in case we can't get to everything during the in-person meeting. So, if everything is completed, this won't take place. But for now, please leave it on your calendars.

Following that, we will have a public and member comment period from July 14th through August 12th, in which the public will have the opportunity to comment on the draft report which will summarize your discussions and ratings of the criteria for each measure during the in-person.

And then, following that, we'll have our post-comment call. This call, we will go over the comments that we received and further discuss how to respond to each of those.

So originally, this was scheduled for August 31st, but we realized that we have a couple of internal conflicts and that it's right before Labor Day. So it's not really an ideal time.

So what we're going to do is (Desi) will be reaching out to you in the next few days with the Doodle poll for the new dates and times, in hopes that we can all kind of agree on one that works.

Right now, we're kind of looking at September 7th, 8th or 9th, but again, we'll be in touch with that.

So Sarah, Suzanne, did I miss anything?

Sarah Sampsel: No, you got it all. I would just say that, you know, regarding the measures that you're assigned for evaluation with the exception of those six UDSMR measures, all of the information is out there, the SharePoint site and has been since late last week or a couple – or released early this week.

The UDSMR measures, what we'll just ask is that you review as much as you can and understand we may not have any reliability scores until the in-person meeting.

We will compile all the information from those surveys and you'll actually – and we'll re-release all of those before the in-person meeting so you can see each others' scores, as well as feedback prior to the meeting as well in the event that you have a chance to pull that up before the in-person meeting to informed discussions.

Kirsten Reed: So Sarah, those last UDSMR measures, those are all available, so at this point, all 13 are up and ready.

Sarah Sampsel: Right, they're there but the reliability testing data is not there.

Kirsten Reed: Oh, right, right, OK, sorry.

Suzanne Theberge: And I'll just add the reason that we do the surveys prior to the actual meeting is so that, you know, you get a head start on looking at the measures where you're the lead discussant and we can get your comments on those measures and put those back into the worksheet, so that, you know, as you finish looking at the remaining worksheet in a week prior to the meeting, those comments will be there.

We'll get those comments, put into the worksheets the day after the survey is closed. So they should be available, like, next – end of the day next Thursday.

Sarah Sampsel: OK. Any final questions?

Any final comments? Lee, Chris?

Lenard Parisi: This is Len Parisi. I'm having trouble still accessing the assignments, the link that was sent in the memo brings you to an Excel spreadsheet, that's prompting me for a password and a log in also. And I'm not able to get it. Can someone just e-mail that to me?

Sarah Sampsel: Yes, we can get that to you, Len.

Lisa Gale Suter: Yes, this is Lisa Suter. I'm having the same problem, I can log in to SharePoint but certain files are restricted, and I can't use the same password to get into those.

Female: Yes, so am I.

Sarah Sampsel: And have you tried today like this afternoon?

Lisa Gale Suter: Trying right now.

Lenard Parisi: Me, too.

Suzanne Theberge: OK. They put a work around in place to get around their – two of our systems are having trouble talking to each other. It sounds like the work around has not been fully effective. We will just put all that into an e-mail, the body of an e-mail and send it around the committee, your lead discussant assignments. And as you come across other files, I guess just let us know by e-mail, you know, what you can't access and we'll get it you somehow and we'll speak with our tech team as well and find out what's going on (anyway).

Linda Melillo: So none of the measure worksheets of the measures that I've been assigned, I can access, so I can only access the supplementary documents and not all of those, and none of the main measure worksheets that's on any of the measures that I've been assigned.

Suzanne Theberge: OK …

Linda Melillo: Also they're password protected.

Suzanne Theberge: It sounds like the work around doesn't work things.

Lisa Gale Suter: OK. Thanks.

Suzanne Theberge: Unfortunately. I am so sorry.

We are going to follow up with our tech team as soon as we get off this call and we'll – we will try to figure out what's going on and try to get that (sent) out as soon as we can.

And unfortunately – yes, the volume of files is still large that we found knowing them isn't really workable. They either don't go through or they get caught by spam filters or it crashes people's mail system, so that's why we can't just e-mail you and everything. But, we will try to find the technical work around as soon as we can.

Sarah Sampsel: So can I – so – OK, so I heard Len, who else was having problems?

Lisa Gale Suter: This is Lisa Suter.

Sarah Sampsel: Lisa.

Lisa Gale Suter: And I also sent an e-mail to the info@qualityforum, too.

Sarah Sampsel: So it's everybody whose name starts with L.

All right, let us look into that and if – you know, even if – and if others aren't having problem – if you are having problems, go ahead and e-mail us. But maybe we could even e-mail individual documents to Len, Lisa and Linda right now, but we'll work on that.

And I just want to say and should really recognize the rest of the project team because a lot of our teams have really had a hard time adapting to these preliminary analyses and they're not getting documents out until right before meetings, and we are really proud of ourselves, we got them out, you know, well over two weeks before the in-person meeting and technology isn't being a friend. So, we're sorry about that.

But if there's nothing else, then we thank you for your time. And again, the Q&A meeting next week is an opportunity to ask questions as you're going through evaluations. And we'll do kind of a quick run through of a preliminary analysis and how it'll work in the meeting based on a measure that's – from another project.

Male: Thank you.

Female: Thank you.

Female: Great, thank you.

Female: Bye-bye.

Suzanne Theberge: Thanks, everyone.

Female: Thank you.

Female: Bye.

END