NATIONAL QUALITY FORUM

Moderator: Care Coordination
January 12, 2017
2:00 p.m. ET

OPERATOR: This is Conference #: 93884709

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

Katie Streeter: Hi, good afternoon, everyone. This is Katie Streeter, senior project manager here at NQF. On behalf of our Care Coordination teams, thank you all for taking the time out of your busy schedule to attend the Care Coordination Standing Committee Orientation and Measure Evaluation Q&A Webinar today.

In here in the room with me is Peg Terry, senior director on this project; (May Nacion), project manager; and Yetunde Ogungbemi, project analyst. And this is the core team that you'll be working with throughout the remainder of the project.

The agenda for today's call mostly will be a refresher for you. On this call, we tend to give you an overview of the NQF Consensus Development Process or CDP and portfolio of measures. We'll also go over the major project activities in the timeline, orient you to the roles of the committee, the co-chairs and staff. Then we'll present the high-level introduction to our measure evaluation criteria.

Now, we'll pause and take a quick roll call to see who is on the call with us today.

Yetunde Ogungbemi: Hello, good afternoon.

Female: Hi, good afternoon.

Yetunde Ogungbemi: So I'm going to give a quick roll call. My name is Yetunde Ogungbemi, project analyst on the Care Coordination Standing Committee Project.

Do we have Don Casey?

Don Casey: Yes, good afternoon, everyone.

Yetunde Ogungbemi: Gerri Lamb? Rich Antonelli? Samira Beckwith? Colby Bearch? Ryan Coller?

Ryan Coller: Yes, good afternoon.

Yetunde Ogungbemi: Chris Dezii? Shari Erickson? Barbara Gage? Dawn Hohl?

Dawn Hohl: Here.

Yetunde Ogungbemi: Marcia James? Emma Kopleff? Brenda Leath?

Brenda Leath: Here.

Yetunde Ogungbemi: Russell Leftwich? Lorna Lynn?

Lorna Lynn: I'm here.

Yetunde Ogungbemi: Jean Malouin? Karen Michael? Terrance O'Malley?

Terrance O'Malley: Here.

Yetunde Ogungbemi: Charissa Pacella? Ellen Schultz? Beth Ann Swan? Jeff Wieferich?

Katie Streeter: Thanks, Yetunde. Did anyone else join whose name we did not call?

OK, thanks. So, we have just a few folks, I think about five or so committee members on the line today. We did have a great turnout on Tuesday when we had most of the other committee members attending including our new

members. There are five new members to the Standing Committee so we are now – or you are now a committee of 21 people. And we do thank Don and Gerri for agreeing to be co-chairs again.

And before we move on, we'd like to ask if you have any opening remarks.

Don Casey: Well, I want to just thank everyone for being here. I attended part of the call on Tuesday so I've heard a lot of the – some of the information, but welcome to the new members. We've been at this since 2005. And as you can guess, it's been a continuous work in progress. And I think this next phase of the work is going to be really important for us.

There's been a lot that's been done. And for those of you who are new to the committee, feel free to reach out to staff or Gerri and myself if you have any particular questions or ideas about things that have gone on before. This is really important and really challenging at the same time. So, we're looking forward to working with all of you and I appreciate the chance for also having a great staff to help support us.

So, I know our meetings are coming up next month. And this is going to be – you know, this is going to be real important. So, we are happy around this call and please feel free to participate.

Katie Streeter: Thanks so much, Don. So just talking with staff here, because we did cover quite in depth on Tuesday the overview of NQF, CDP and the roles when our new committee members were on the phone. And since most of you or actually all of you have been through this process before are familiar with NQF. And this is a – you're on your second term serving here. We are going to skip over the next few slides. And with the exception of one slide that I'd like to point out with you, that is the change in our CDP, our Consensus Development Process.

As you may notice on here, we've changed from eight steps for measure endorsement, we are now seven steps for measure endorsement. What has changed is there's no longer a board ratification step. So now, the measures – the recommendations that the committee makes will go through the first six

steps. The next step would be – the Consensus Standards Approval Committee, they will be responsible for ratification and endorsement. From there, the measures that are recommended will move into an appeals period. So that is one change in our process that we wanted to highlight for you.

So moving onto the Care Coordination portfolio. And I know most of you are quite involved in the off-cycle work last year that our colleague, Rachel Roiland, was leading. And in the fall, I know that you did have quite an in-depth discussion about the portfolio. So you are very familiar with it.

I want to assure you that our team has been working with Rachel to catch up on all of your discussions so that we can continue that work, perhaps, during our in-person meeting if there is time and really keep that momentum going.

For this particular phase of work, we do have five measures that were submitted for maintenance review. And we also have two new measures that were submitted to be reviewed for endorsement.

The two new measures come from the Collaboration for Advancing Pediatric Quality Measures. They are one of seven centers of excellence funded through AHRQ. And the measures have to do with asthma connection with primary care before and after an E.D. visit. So, that'll be a total of seven measures that you'll be reviewing in this phase of work.

And just quickly, this is a slide I'm sure you're familiar with, you obtained it last fall. But the highlighted here, the measures that are coming back to you for maintenance review.

So, these measures were endorsed in phases one and two. And you'll see here three measures are bolded and those are the measures that you'll be seeing in this phase.

And then here, the other measures that were submitted for maintenance review, from phases one and two.

Phase three measure sets were recommended for endorsement and are remained endorsed. At this time, we are not due for maintenance review so you won't be seeing them during this phase three work – of work.

High-level snapshot of our timeline here, as you can see this week, we have our committee orientation and Q.A. calls. Towards the end of the month, we will be sending you measures, the seven measures, we'll be posting them on SharePoint. And in this information will be staff preliminary analysis which is something new. And Peg will kind of go into more detail about what that is later in the call.

When you review the measures, we will also be asking you to complete evaluation surveys on SharePoint. We'll also be giving you more detailed instruction about that soon.

The workgroup calls, we will be dividing the committee into two workgroups. They're scheduled for February 6th and February 7th. If you know now that you are not able to attend one of those days or have a preference on which day that you would like to attend, please send our team a note and we'll be sure to place you on that appropriate workgroup.

We will be also assigning lead discussants for each measure similar to last time. So be on the lookout for that information coming your way soon.

The in-person meeting, something of importance that we wanted to note is that we have switched from a two-day in-person meeting to a one-day in-person meeting. So right now, it's scheduled for Wednesday, February 22nd. The meetings department from NQF will be contacting you probably within the next week or so to make your travel arrangements for the meeting.

We have a post-meeting call on the calendar scheduled for March 7th. We may or may not need that meeting depending on any follow up that may be necessary from the in-person meeting. So we'll leave that on the calendar for now.

Your recommendations will be posted for public and member comment from March 30th through April 28th. And then on May 16th, the committee will

need to review and respond to all public and member comments that were received.

We'll have the typical 15-day member voting period in early June. Then your recommendations will be presented to the CSAC during an in-person meeting on July 11th or 12th. And then the project will close with the last step of the 30-day appeals period.

So before we dive into a refresher of the measure evaluation criteria, we're also be highlighting a few changes that have been made since the last review in showing your staff preliminary analysis example.

Are there any questions about what has been covered so far?

Don Casey: Katie, this is Don. Who on our call has not been through this process before? Could you identify yourself?

Ryan Coller: Yes.

Don Casey: We should know.

Ryan Coller: Ryan Coller. This is my first time, I'm a new member.

Don Casey: Great. Anyone else?

Well, that's good. I think, Katie, if we don't – if we didn't get the other four new members, I didn't remember the list from last time. But if we didn't get the new four members, I would happy to – I would be happy to help you in an abbreviated call, because I think those of you who've been on the committee know that this is a lot of work. We get a lot of dense information, hand it to the newcomer who has not done this in other parts of NQF. It can be daunting in the beginning. And that's why we developed these many teams to help us guide our discussion to help get through this.

But, I think the important thing is, once you learn the process, I think – which is one of the goals of today, I think it won't be – it will be – it won't be easier in the sense that it'd be less time consuming. But I think you'll understand it better, it's kind of a learn skill, and I know Katie is going to go through a lot

of the details about how this goes. So, it may be daunting to you, and certainly afterwards if you have any questions for us, please reach out to us.

Ryan Coller: Thank you.

Katie Streeter: Thanks, Don. And we do know that the four other new members were present on Tuesday, but I think that's a great idea, perhaps, our team will reach out to all five new members and ask if they like any follow up or if there's anything else we can do to orient them to the process.

Don Casey: Sounds great.

Katie Streeter: OK. So now, I'm going to turn it over to Peg who will be speaking to our measure evaluation criteria overview.

Peg Terry: Thank you, Katie. And I'm just going to dive right into this. And measures are reviewed against all evaluation criteria that are current at the time of the review. NQF guidance was updated in 2016. The criteria has undergone significant changes over the last few years with more emphasis on feasibility and usability, and less as an emphasis on scientific acceptability for maintenance measures.

Some of the changes include the SDS review, the PRO-PM composite measures, all or none, and we do have two composite measures that we'll review in this during this time, an additional guidance on outcome measures for performance gaps and new guidance on population health and access measures. This new guidance is in place to assist in the changing measurement landscape and to raise the bar on changing the process evaluating measures against the criteria.

Because measures have been endorsed previously does not mean they're automatically accepted – expected to reach the current – to meet the current criteria.

So, next slide. How do we decide what is good enough for accountability purposes? Standard criteria is known to all, developers know what to expect, end users know that a measure has been evaluated a certain way.

So, NQF endorses measures for accountability applications including public reporting, payment programs, accreditation, as well as for quality improvement. Standardized evaluation criteria. Criteria has evolved overtime in response to stakeholder feedback. The quality measurement enterprise is constantly growing and evolving, greater experience, lessons learned, expanding demands for measures, the criteria evolve to reflect the ongoing needs of stakeholders.

So, the page numbers onto the next slide and the page numbers that you'll see on many of these slides reference the committee guidebook. So as you want to look later, you can go back and you can kind of look at the guidance and really make sure you understand it.

The criteria are in a specific order, there is a hierarchy, there is a logic to looking at them in a specific order. The first will be importance to measure and report, followed by reliability and validity-scientific acceptability to meet measure properties. Criteria one and two are must-pass criteria.

Importance to measure and report, the goal here is to measure those aspects with the greatest potential of driving improvements, if not important, the other criteria are less meaningful. It is must-pass as I've said.

Reliability and validity-scientific acceptability of the measure properties, goal is to make valid conclusions about quality, if not reliable and valid, there's a risk of improper interpretation. That's also a must-pass.

Feasibility, the goal is to ideally cause as little burden as possible. If not feasible, consider alternative approaches.

And usability and use, goal is to use decisions related to accountability improvement. If not useful, probably do not care if feasible.

And also, we'll be looking at a comparison, we'll be comparing related and competing measures in this portfolio.

So importance to measure, caution that importance to measure and report does not speak to if the topic is important. The process of care for this topic area is probably very important, everything we do in health care is important. But in terms of having the right measures, not everything needs to be measured.

A committee must consider if this aspect of care should be measured. Is extending resources and developing a fairly considerable infrastructure to collect and report on data for a measure seemed reasonable and necessary? Does the value and importance of the information we're obtaining offset the burden of measurement?

So under importance to measure, it is the extent – and report, it's the extent to which specific measure focus is evidence-based, very important and important to making significant gains in health care quality, whether it's a variation and overall less-than optimal performance.

So, under each of the major criteria, we have sub-criterion. The sub-criteria here is evidence, which we just talked about, which is the major focus – the measure focus is evidence-based.

1B, or the second sub-criteria, the opportunity for improvement, demonstration of quality problems and the opportunity to improve, data demonstrating considerable variation or overall less-than optimal performance in the quality of care across providers and/or disparities in care across population groups.

And 1C or the third sub-criteria, is the – for composite measures, is the quality construct and rationale. Is there a quality construct and rationale that you can (discuss)?

So, under the first one, evidence, and we'll be talking about this and we'll be referring to this in the algorithm of evidence. For outcome measures, the way that outcome measures are evaluated is, that the outcome measures are apparently important and really the reason and most important information that people want to know about health care delivery.

Patients want to know what happened and providers and professionals should want to know how well they're doing. But the bulk of measures in this project or process, so for outcome measures, we're really looking in a rationale for how the outcome is influenced by either a health care process or structures, or an intervention.

For process, intermediate outcomes, and structure measures, what we want to see is the quantity, the quality and consistency of the body of evidence underlying the measure, which should demonstrate that the measure focuses on those aspects of care known to influence desired outcomes. Are there empirical studies? Expert opinion is not evidence. Is there – are there systematic reviews and grading of evidence? Clinical practice guidelines, variable in approach to evidence review.

Developers are asked many questions around the issue of quantity, quality and consistency of evidence. You're going to see that that is a theme as you're looking at your process measures.

So, what we have here is really a quick look at the algorithm. And the algorithm is setup to have a standardized approach. And as you – as we walk through this – we'll show you just a little bit more when we have our example.

Outcome measures are evaluated differently than structured process, and intermediate outcome measures. And this algorithm really provides that guidance for the staff as well as for committee members as they're evaluating the measures.

So again, as you look – and I'm not going to go through this, first of all, it's hard to see. And second of all, you'll get a better look when you're doing this yourself. But basically, the algorithm is a way to standardize our process for us and for you.

Under criterion one, importance to measure and report, the criteria emphasis is different for new and maintenance measures. For new measures, evidence, as we talked about, for process measures, we want to know the quantity, quality and consistency of evidence. For out – and for outcome measures, we want to

– there needs to be link for the process – for process measures with outcome. So, what are the (lengths) between the process of care and the outcome.

Under new measures, gap is important and it's the opportunity for improvement. Is there an opportunity for improvement? Is there a variation in the quality of care across providers? And for maintenance measures, a little bit of a different emphasis. So, the decreased emphasis on maintenance measures, we do require the measure developer to attest evidence is unchanged for the last evaluation. Standing Committee will have to affirm that there's no change in evidence.

If change is in evidence, the committee will evaluate as they would a new measure. And for the area of gap, there is an increased evidence. So, we should see new information on the performance of these measures. And so data on current performance, gap in care and variation will be the same as for a new measure.

Criterion two, reliability and validity-scientific acceptability of measure properties. Reliability are not all-or-nothing properties. They are a matter of degree. Reliability and validity are not static, they can vary with different conditions of using the measure. In order to be valid, a measure must be reliable. But reliability does not guarantee validity.

Empirical evidence reliability, the measure testing is expected. Reliability and validity are demonstrated for the measure specified, not the measure concept.

And measure specifications are addressed under both, reliability and validity. And under reliability, you're looking for precise specifications including exclusions. Under validity, you look under the validity, you're looking for – that specifications are consistent with evidence. Flexible testing options rather than prescriptive are used for reliability and validity.

Specific thresholds are not set, results can be within acceptable norms. And you'll get a sense of that as we walk a little bit through the measure that we are going to show you today.

Insufficient evidence cannot be evaluated or considered for endorsement. Reliability can be tested at the data element level with the measure score level.

I want to just take a minute and go through this really quick snapshot of a little bit about reliability and validity. In the first target, all the measures are quite similar but they don't do a very good job of hitting the target. This portrays the measure that is reliable but not valid.

And the second target, measurements aren't very close to each other or the center of the target. This portrays the measurement is neither reliable nor valid.

And in the third target, all of the measures are close to each other and to the center of the target. This portrays the measure that is both valid and reliable.

Note that in order to be valid, a measure must be reliable, but reliability does not guarantee validity. I think I will keep saying that throughout the presentation.

Measure testing, excuse me, empirical analysis is used to demonstrate the reliability and validity of the measure as specified, including analysis of issues that pose threats to the validity of the conclusions about quality of care such as exclusions, risk adjustments, stratification of – for outcome and resource use measures, methods to identify differences in performance and comparability of data sources and methods.

Again, let me emphasize that these – that there are examples of how a developer may actually test. So, under – so here, under reliability, first we are going to talk about the measure – reliability of the measure score refers to the proportion of variation in measure score to systematic differences across the measured entities in relation to random variation or noise, the precision of the measure. And an example is, in a statistical analysis of sources of variation in performance measure scores, we talk about signal-to-noise.

In reliability of data elements to the repeatability – refers to the repeatability and reproducibility of the data and uses patient-level data. An example of a test, the way we test for that is inter-rater reliability.

Consider whether testing used an appropriate method and included adequate representation of providers and patients and whether results are within accepted norms.

So here's another algorithm of reliability. And basically, it gives you a real standardized approach to looking at reliability and determining whether the measure is reliable and whether you can continue to proceed to review this measure.

I just want to point out one thing here. So, if the – was the empirical reliability conducted using statistical test with the measure specified, and if that's not – if that doesn't happen, the next – we're on box two. So as you look at this, and you'll see our example, each of these sections are boxes and we refer to these in terms of our evaluation of measures.

When you look at the box two that you move to the box three, where it says, was empirical validity testing of patient-level data conducted. If it wasn't – if it was, and you can move to the validity testing because it's patient-level data and when you test for patient-level data, you do – you use it in validity and you can also use the same results in reliability.

So, I'm going to move forward and talk a little bit about validity testing.

Some key points, empirical testing. We have testing at the measure score and testing at the data element level. A measure score assesses a hypothesized relationship of the measure results to some other concept, assesses the correctness of conclusions about quality.

Data element assesses the correctness of the data compared to a gold standard. And we'll talk about what some examples of those are in a minute. And face validity is the subjective determination by experts that the measure appears to reflect quality of care.

So here is your reliability – I mean, your validity algorithm. Sort of a walk through some of the steps you need to take or we will take and then you will

review and have your own review of the measure going to the algorithm, just gives you a sense of the standard that we're using to assess for validity.

So, threats to validity, how to consider the potential threats to validity? There are numerous threats to validity. Conceptual, measure focus is not a relevant outcome of health care or strongly linked to a relevant outcome. Unreliability, generally, an unreliable measure cannot be valid. I think I've already said that several times.

Patients inappropriately excluded from measurement. Differences in case mix for outcome and resource use measures. And measure scores that are generated without – with multiple data sources and methods, and systematic missing or incorrect data, intentional or unintentional.

Scientific acceptability, for new measures, measure specifications are precise with all information needed to implement the measure. And for reliability and – that's for reliability and validity under maintenance measures. No difference, you need to have specific – requires updated specifications. So, we get to the actual testing of reliability and validity. If prior testing is adequate, no need for additional testing at maintenance with certain exceptions. This is important because you'll be looking at a lot of several maintenance measures in this portfolio.

Data source, level of analysis or setting, must address the questions for the SDS trial period that we're still in.

Feasibility. Feasibility is the extent to which the required data are readily available, retrievable without under burden and can be implemented for performance improvement. There are three sub-categories – sub-criterion, I should say. The first is clinical, for clinical measures, the required data elements are routinely generated and used during care delivery such as blood pressure, lab test, diagnosis, medication order.

Second, the second B of three, the required data elements are available in electronic records or other electronic sources of data. If required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

And 3C is, demonstrates that the data collection strategy, source, timing, frequency sampling, patient confidentiality, cost-associated with fees, licensing for proprietary measures, can be implemented, already in operational use or testing demonstrates that it is ready to be put into operational use.

Well-known and more seasoned measures tend to have a feasible standard data collection strategy with newer measures, committee members must pass what is the developers' path.

So, usability and use. Feasibility is not a must-pass criteria. So, we're onto the fourth criteria, usability and use. So, extent to which potential audiences are using or could use performance results for accountability and performance improvement to achieve the goals of high-quality efficient health care.

So, on the sub-criteria under usability and use is the accountability and transparency. Performance results are used in at least one accountability program within three years after initial endorsement and are publicly reported within six years after initial endorsement.

The second sub-criteria is improvement. Progress towards achieving the goal of high-quality efficient health care for individuals or populations is demonstrated. And when we look at the example, you'll get a sense of maybe a measure that's not showing a great deal of improvement.

Benefit outweighs the harms. The benefits of the performance measure in facilitating progress toward achieving high-quality efficient health care for individuals or populations outweigh evidence of unintended negative consequences.

Last, vetting the – last sub-criteria is vetting by those being measured and others, those being measured have given results and assistance in interpreting results. Those being measured in others have been given the opportunity for feedback and the feedback has been considered by developers.

So, questions to ask, has the measure been in use for a while? Is it working? Is it driving to improve our measures? Are things improving? Are we going in the right direction? And do the measure – the benefits outweigh a harm?

So, feasibility and usability, I just want to go over the difference between newer maintenance measures again. For feasibility and new measures, measure feasibility including e-assessment needs to be done and needs to be included.

Under new measures for usability and use, use in accountability programs that I just mentioned, applications and publicly reporting. And really what are the impact in unintended consequences and usability.

Under maintenance measures, there's really no different implementation issues may be more prominent. And in the area of usability and use, it's really an increased emphasis here, much greater focus on measure use and usefulness including both impact and unintended consequences.

So, related and competing. We really want to know what we can do to reduce the chaos and foster harmonization and make measures – and make decisions about closely related and competing measures.

If at the Standing Committee you recommend a measure for endorsement, you may want to have to decide – you may have to decide whether they are related and competing measures and you may also have recommendations about how this should be handled.

So, what is a related measure? If a measure meets the four criteria and there are – and they are endorsed, new measure – and new related measures have the same measure focus or the same target population. For competed – competing measures, they have both the same measure focus and the same target population. The measures are compared to address harmonization and/or selection of the best measure.

What you'll see for the first time this year, even if you've been – excuse me, even if you participated in prior review of measures as a Standing Committee,

there's a preliminary analysis aspect where – and this is in place to assist the committee in evaluation of each measure against criteria.

Here, the NQF staff will prepare a preliminary analysis of the measure submission and offer preliminary ratings for each of the criteria. There would be – this will be used as a starting point for the committee and for discussion and evaluation. So, the staff will do this and on the measure, I'm going to show you in a minute.

You can see how that would look, just be aware that it is just the starting point. Each committee member will be assigned to subset of measures for an in-depth evaluation. Those who are assigned measures will lead the discussion of this measure with the entire committee.

During the workgroup calls for this – for each committee, the evaluation process that assist committee members with their first – why do we have workgroup calls, that assist the committee members with their first evaluation. Committee members and measures will be divided into groups for preliminary cost to discuss measures and share initial insight. And ensures an initial familiarity with measures, (allows practice with) NQF criteria and processes, and gives early feedback to developers of committee questions and concerns, very important.

Measure evaluation and recommendations at the in-person meeting, the entire committee will discuss and rate each measure against the evaluation criteria and make recommendations for endorsement.

I wanted to just mention a change that we have at NQF and it's called Recommendation for Endorsement and Endorsement Plus. Endorsement Plus is a new designation here at NQF.

So, the committee votes on whether to recommend a measure for NQF endorsement. Staff will inform the committee when a measure has met the criteria for possible endorsement plus designation. And some of that criterion includes, it meets evidence criteria without exception, there is good results on reliability testing of the measure score. Good results on empirical validity of

the measure score, not just face validity, and well-vetted in real-world settings by those being measured and others.

Committee votes on recommending the endorsement plus designation, indicating that the measure exceeds NQF criteria in key areas, excuse me.

So, what I want to do is just take a quick look at a current measure. And then we'll stop and have some questions. Here we are.

(Off-Mic)

Peg Terry: Great. So, this is a measure that actually NQF staff have already reviewed and done a, what we call, P.A. or preliminary analysis on. This is a measure that is – measure title, (essentially) to get a look at what this brief measure information would look like. And the title is Chronic Stable Coronary Artery Disease Antiplatelet Therapy. There's a steward, the American College of Cardiology. There's a brief description of the measure and a rationale.

There is a numerator. Patients are prescribed aspirin or Plavix, which is the brand name for the measure within 12-month period. And underneath that, there is information on the denominator statement, the denominator exclusions, and it's really a good description of actually the measure.

Then you go down a little further and you get – you see where there's a description of the type of measure. So this happens to be a process measure, the data source. It is an electronic – from electronic clinical data and the registry is where they get the data for this measure, for evaluation of this measure, the level of analysis at the clinician-individual level. And this is not an eMeasure.

So, most of this information that you're looking on – at now, it comes directly from the developer. So, as you go down further, you kind of get a sense of the beginning of the preliminary analysis. And this is evidence.

So, for this particular measure, and as you look at it, you get – you see that the developer states that there's a systematic review of the evidence specific to this measure. Quantity – quality, quantity and consistency of evidence

provided, and evidence is graded. All of that is checked off by the actual developer.

So, as you look here, it's a summary of prior review in 2014. So this is not a new measure. In fact, when you look on the first sheet on the page before, and you know how to go back there, it just gives you, say, a sense that this was first endorsed, this measure in 2009, so it has several times that has been endorsed.

So here, you see that there's a summary of the prior review. And in the summary, what is provided is information on the use of guidelines through the ACCF and AHA, which provides some evidence and its graded evidence. So, it's graded at the – as class one and its level of evidence is an A for aspirin and for Plavix, the level of evidence is a B. And so, the evidence is really good here.

Then you go down and you see that the – the developer said – the developer attest that there have been no changes in the evidence since the measure was last evaluated. But then, they go in and provide a little update. But we know from looking at this that the evidence, the updated evidence, actually does not conflict, there's no conflict with the prior evidence. So they're just updating what evidence they have in the past.

And then, there are questions that the staff have developed here for the committee. And then below that, you'll see that this – the guidance from the evaluation algorithm, remember those algorithms and this is what the staff does. They go to the algorithm and they basically rate the measures to the boxes and they put the information that's available. And at the end, they come up with a possible rating, and here, the evidence is rated as a high. That's evidence.

Then, the next part is opportunity for improvement.

So, gap in care opportunity for improvement. So you go to the next page and there's lot of tables that sort of give you a sense of how this measure is done. And the measure is from a registry, which is from the ACCF or AHA.

So, as you look at this evidence, you can see – so it doesn't – the first table is about performance data. You can see that – if you look at the mean for 2013 and '14, you see that the performance has really not changed very much. Then they have other data, descriptive data about disparities. But then there's really no analysis, it just provide that descriptive data.

So when you go down to the bottom, here, you look at preliminary rating for this. You see that the rating is moderate. And the rating is moderate probably because there is really no improvement over a two-year period. You can go down and look at – and then there are questions the staff have come up to ask the committee to consider as they're evaluating this measure.

Next is reliability. And in this area, this is a lot of – as you can see, the specifications are written here. As we said, they must be whether to newer measure that's already been endorsed, the specifications are updated and there are questions regarding that, and there's testing. So in this particular measure, there is testing at the data element level and the performance level.

And so here, if you go down further and it says summary of testing is both measure score and data element testing, they tell you the kind of testing that was done, data element, reliability testing, how they're comparing manual extractions from an electronic health record as opposed to automatic from an electronic health record.

The score level of testing is done using a statistical model called the beta-binomial, signal-to-noise and then they have results. So you always want to see results. You want to see some numbers that tell you how this is going, and you do have results. You have a percent agreement on the manual versus automatic EHR for the data element and you have score level of testing. And so you have some numbers from that as well.

And so you can go down below and you can see, again, there are questions for the committee and there's guidance from the reliability algorithm and the staff really provides this walkthrough with the boxes to show you how they came up with their rating, which is a high here.

Validity testing. So validity testing is sort of run – it's sort of the same. You – the specification is consistent with evidence, which we have written here. And when you're looking at validity testing, you're looking also at exclusions and threats to validity. And there's a lot of detail on the exclusions and threats to validity.

As you continue to go down here – a little bit further. The way this – I'm sorry, the way this validity was tested with face validity, they have an expert panel. And the question – I was looking for the question the expert panel was asked. But basically the expert panel is asked to validate that the information is provided an accurate reflection of quality and can be used to distinguish good and poor quality. And they have a percentage and they have the number of people that participated.

Then we can talk about threats to validity as well as spelled out here. Questions provided by the staff. We know it's not risk adjusted. They don't usually risk adjust, process measures, and questions for the committee.

Also, we're looking at missing data. That's important. They do talk about missing data. They don't really do much of an evaluation. And so as you go down here further, validity is based on face validity. And so, face validity can only be at the moderate level. So, we have a moderate rating on that.

And then feasibility is the same kind of walkthrough as I just said before using the feasibility criteria. I do want to just point out one thing now on usability, just on page eight.

So this is a measure that's been out for a while. And it's – the developers had data on how the measure is doing and we saw some of that already from the registry, the PINNACLE Registry.

But, we do know that, also, this feedback provided from what is called the Measure Applications Partnership here at NQF. And there was some sense that they really didn't support the clinical workgroups that our consumers and workgroup members did not support the measure for the Physician Compare and Value-Based Payment Modifier Program because the measure does not adequately address current needs of the program.

And then the feeling here that the measure – also in – that was 2014, 2015, again, did not support the measure for the Medicare Savings – Shared Savings Program as the Optimal Vascular Care measure, contains the measures, the component, the composite. So both measures have redone this. So two reasons they didn't support this measure.

The preliminary rating for usability was a low. I must say that this is not a must-pass measure – must-pass criteria, but I wanted to point that out to you so you see – kind of get a sense of how that works.

I think I'm going to – this – I'm thinking I'm going to stop there, a lot of detail, a lot of detail. Any questions? Any comments?

Don Casey: Katie, Don, thank you for this elegant overview. I have a couple of observations relative to your other slides, and then I want to talk about this example quickly, because I think it is important.

I don't – having sat in on your call last – on Tuesday as well, I noticed that – I can't remember who was, someone said expert opinion is not considered evidence, and I wanted to just say that I don't think that's a fair statement. It's a type of evidence. It isn't subject to the types of rigor that we got, for example, for the one you just used, where there's tons of pharmacologic randomized controlled trial data showing that there is, you know, clear improvement in survival overtime by the use of appropriate antiplatelet agents for STEMI.

And by the way, I was – I am on that guideline writing group so I know this issue. But, for example – and I think this is relevant to the Care Coordination Steering Committee such measures, for example, of weight for heart failure. An electrocardio – it's important to do an electrocardiogram to make – critically important to make a diagnosis of S.T. elevation, M.I. or it's important to give discharge instructions for patients with total hip replacements.

These are the types of measures that I'm talking about that have been used in the past that will never have the scrutiny of randomized controlled trials and

are, you know, sort of common sense, you know, interventions that occur in daily practice and are based upon expert opinion. No one is not going to do, for example, an EKG on someone that doesn't have at least the suspicion of an S.T. elevation, M.I.

And I think that when we get into Care Coordination as you'll see to the, you know, to the committee members, we're not going to get probably the type of information that Katie just reviewed, I'm sorry, around, you know, the dual antiplatelet therapy, because it just doesn't exist in that format.

So, I think for purposes of showing you how the decisions are made, it's useful. But I think we will be challenged when we get to seeing some of these reports because of the way the nature of the measure says it were.

Often times too, I've seen people present – measure developers present what they refer to as literature reviews, which in my mind is not, you know, in the hierarchy that comparable to say a formal systematic review.

And some of these measures may have direct or indirect evidence supporting them out of systematic reviews, but we have to be careful when we see the term literature review because it generally means that it's of a lower quality caliber of analysis.

And then, with respect to validity, then I think you talked about this, and I apologize I think I called you Katie, but it is Meg, right?

Peg Terry: Peg.

Don Casey: Peg, excuse me.

Peg Terry: Yes, that's OK.

Don Casey: The, you know, when we get to validity, I think that initial measures provide evidence supporting the use of the measure such as this one. You know, it was good at the start to at least evaluate whether people were getting dual antiplatelet therapy post STEMI because we knew that that had a major impact on mortality, if we're – if not being taken.

But I think we have to get – and I see a little bit of uncertainty here. Evidence of usefulness sometimes gets confused with evidence of impact. And I think you talked about it fairly nicely, that is, in this case, of the dual antiplatelet therapy seeing progress.

Now, I could argue based upon newer information, let's say, with the Quality Payment Program that's now supplanting PQRS, where measures are actually ranked in decile format. But, looking at the low end of – or the spectrum for that measure which is below 65 percent means that there's significant number of physicians who have a big opportunity for improvement, you know, whether it gets into the Medicare Shared Savings Program, and that was a different subject.

But I think you see what I mean, that looking especially from measures that we are reviewing for maintenance or for reevaluation, it will be important, it will be important to really look at the second validity question, which I think sits in 4B as I recall.

So, those are kind of my generic comments. Thank you.

Peg Terry: Well, thank you very much. Obviously, you've been paying close attention. And I appreciate your perspective, and I think it's very good when you sat on this committee before. So, thank you again.

Any other questions from anybody else?

Just remember, this is the first overview. And if you're really not familiar, really ensure that the committee handbook as your beginning to look at any measures that are assigned to you, and would – for a workgroup call.

OK. If there are no other questions …

Don Casey: Peg, I just wanted to say that I do think those of you who've been involved with maybe one cycle, have a sensitivity to this, but not enough to know that NQF staff is continuously looking at these criteria and continuously working on them, and continuously getting feedback. So this is always an evolution.

And we may learn some things in this process that might help them inform the additional – they're not revisions, but just for refinements to this process. Because it's come a long way and it's really, really been improved at each iteration. So, want to thank NQF for being very sensitive to that.

Peg Terry: Well, thank you. Thank you very much.

So, excuse me, I have a bit of a cold. So, I'm going to turn this over to Yetunde, excuse me.

Yetunde Ogungbemi: Hello. So, I'm going to give a SharePoint overview for those of you who are new. Obviously, this is something that you would like to pay attention to. But for those who have been with us before, it'll be just a quick refresh of SharePoint and how – where things are located.

So, on your screen here, you can see the Care Coordination SharePoint Committee homepage. It looks pretty much the same, but you will not have access to all of these things up here because on staff, that's the only reason it looks a little different.

So, directing you to the committee homepage, the general documents is the first thing that you will see here. And here, it's located just general documents for the committee.

And anything that is CDP specific will be located here. And then some other things like your roster or algorithms will also be located here like Peg reviewed before.

So, specifically, they are the framework domains from the first and second phases of work, the committee guidebook that you should review that's been updated within the past year. Standing Committee policy for CDP groups, your committee roster for this phase of work, your evidence in testing algorithms are also here. And those are the algorithms that Peg showed you before when she was reviewing the measure evaluation criteria.

The newest measure evaluation criteria guidance is also located here as well as an NQF glossary.

Further down is a section for measure documents. And here, will be located – or I should say – pardon me, is located. What will be located here is the measure document sets that house all seven measures and any important information pertaining to those.

So, anything that has to accompany the documents or the measure (PAs) and worksheets from the developer will be housed here. And further down just a little bit is the meeting and call document section.

So every meeting that we have here will be located – every meeting that we have for the Standing Committee will be located here and all pertinent information that you need to review will be located in this section here.

So, we have the orientation agenda and slide deck from day one. And then, information is right here as well.

And the survey that Katie spoke about earlier today or maybe Peg, I'm sorry, will be located right here.

And in this survey or in this link, you will find all of the measures and questions that you need to review in order to respond to the survey. And here, this is the committee home link.

All of the dates for the meetings that you will have during the phase of work are located in the committee calendar. And the committee links will be anything that your chairs or fellow committee members deem important that you know, just reference materials or pleasure reading, if you will.

Your committee roster is located behind this link. So if you want to get in contact with your other committee members, and I'm continuing to update this with e-mails. So, this is a great resource if you want to get in contact with your fellow committee members.

And last but not least, staff contact. So here are the four people that we introduced you to at the beginning of the call, and that is us four. We are

located in the NQF offices and you can reach us at any time at our project inbox or the e-mail addresses listed here.

Does anyone have any questions?

OK. If not, I will move into our next steps and we can adjourn.

So, next steps, we have concluded our measure evaluation orientation and Q&A calls, that was this week. Thank you for everyone who participated and joined us for these calls.

Next, we will have our workgroup calls in which we will review measures that are under review for the phase of work. And that's early February.

And our next big meeting is on Wednesday, February 22nd, investor in-person meeting which is here at NQF offices in D.C. And you will receive travel memo from our meetings department within the next one to two weeks.

And here is our projects contact information. Again, you can reach any of the Care Coordination staff at this e-mail here, carecoordination@qualityforum.org. You can call us, or you can access the projects page which is the public page at this link. And the SharePoint site is also here as well.

I sent a follow-up e-mail on Tuesday after our first call with information on how to access that. And you should be receiving information from nominations@qualityforum.org with that information to log in to SharePoint if you do not remember that contact – or that access information.

So, if no one has any questions, I will turn it back over to my colleague, Katie, to adjourn.

Female: May I just ask a quick question, please? And you probably already have talked about this. I know we're going to get further information. But the in-person meeting, you envision that as an all day, is it like 8:00 to 6:00, or I don't know.

Yetunde Ogungbemi: We envision it happening from like 8:30 to 5:00 pm?

Female: OK.

Yetunde Ogungbemi: Yes. So because we only have seven measures, we shortened it to a one-day meeting so we didn't have to have you here for two days.

Female: OK. Thank you.

Yetunde Ogungbemi: You're welcome.

Katie Streeter: Thanks, Yetunde. Any other questions from committee members?

And as I mentioned earlier, we will be sending you detailed information soon on kind of what to expect when we post the measure information and worksheets to SharePoint before you start your preliminary evaluations. So we'll send instructions for that. We'll send you the workgroup assignments and what measures that you'll be assigned to via primary discussant for.

And I think – we just want to remind you that our project team is always here for you to guide you through this process. And if you have any questions about our criteria or anything, accessing the documents on SharePoint, please don't hesitate to reach out to any member of our team. And we're more than happy to walk you through that.

Don Casey: This is Don. I think Gerri on Tuesday's call also made – mentioned of potentially providing some additional background files on SharePoint. And forgive me if you covered that. But, you know, that's kind of helped the newer members of the committee trace some of the history so that they can get a little more context if they wanted about what we've done.

Because I think oftentimes, and this is not a criticism nor a worry, the new members get on the – into the discussions and start bringing up the very same types of things we've been wrestling with for two or three iterations which is perfectly natural. But we don't want to dismiss them.

But we certainly – and we certainly don't want to stifle conversation. But we also want to take advantage of any prior documents that help them see what the types of things we've been discussing with Care Coordination which is, as

I pointed out, nuance compared to some of the other committees just because of the complexity of the challenge.

So, I think you had said you are going to talk to Gerri about trying to do that with a couple of documents as I recall.

Katie Streeter: Yes. Thanks, Don. I think that'll be really helpful. And we'll follow up with you and Gerri about that and place them – place the historical information on SharePoint. So, thanks for reminding us.

OK. Well, on behalf of our team here, thank you so much for taking time out of your busy schedules. And we can't stress enough how thankful we are to have committee members like you who are pretty much volunteering your time to help us get this important work done.

So, thank you so much. And we will be following up with you soon with next steps.

Female: Thank you.

Male: Thank you, ladies, so much. Have a good day.

Katie Streeter: You too.

Operator: Ladies and gentlemen, this …

Female: Bye-bye.

Operator: … conclude today's webcast. You may now disconnect.

END