NATIONAL QUALITY FORUM

Moderator: Cardiovascular Standing Committee
September 25, 2015
2:00 p.m. ET

OPERATOR: This is conference #: 84580494

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

Leslie Vicale: Good afternoon, everyone. This is Leslie Vicale, the project manager of the Cardiovascular Team here at NQF. I'd like to welcome everyone back today to resume discussions on a few of the cardiovascular Phase 3 measures that are under review and evaluation.

And I'd like to first start off by seeding the objectives of the call today. And today's purpose of the call is to continue the discussion and the vote for Measure 2763 to vote – discuss and vote the registry measures from AMA-PCPI, that's 0070, 0081, and 0083. Additionally, we will revisit one of the criterion for Measure 0965 from ACC.

Following all of the discussion and the voting, we will have a member and public comment time, and then we will time line.

Before I get started, I did want to note the rest of the cardiovascular team I have joining me here on the call, and we have our senior director, Melissa Marinelarena and Donna Herring.

Melissa, of course, is our senior director, and Donna is our project analyst.

Melissa Marinelarena: Good afternoon, everyone.

Donna Herring: Good afternoon.

Leslie Vicale: So, before – again before we get started, I did want to note that NQF staff will be facilitating these.

I wanted to go ahead and take roll of the committee that we have joining us. So I'm going to go ahead and read off those names and please do let us know if you are present on the call.

OK. Mary George?

Mary George: Present.

Leslie Vicale: Thanks, Mary.

Tom Kottke is not on the call today, and we've noted that. He's unavailable.

Sana Al-Khatib?

Sana Al-Khatib: I'm present. Good afternoon.

Leslie Vicale: Hi, Sana.

Carol Allred?

Linda Briggs?

Linda Briggs: I'm here.

Leslie Vicale: Hi, Linda.

Linda Briggs: Hi.

Leslie Vicale: Leslie Cho?

Joe Cleveland?

Michael Crouch?

Liz DeLong?

Ellen Hillegass previously noted that she is unavailable.

Judd Hollander?

Judd Hollander: Hi. I am here. Good afternoon, everyone.

Leslie Vicale: Hi. Good afternoon. Thanks, Judd.

Tom James?

Tom James: I am here.

Leslie Vicale: Hi, Tom.

Tom James: Hi.

Leslie Vicale: OK. Joel Marrs will be joining us after 3:30.

Gerard Martin?

Kristi Mitchell?

Kristi Mitchell: I'm here.

Leslie Vicale: Did I hear you, Kristi? Yes?

Kristi Mitchell: Yes.

Leslie Vicale: Great. George Philippides?

George Philippides: Here.

Leslie Vicale: Thanks, George.

Nicholas Ruggiero?

Jason Spangler?

Jason Spangler: Present.

Leslie Vicale: Thank you.

Henry Ting?

And Mladen Vidovich?

Mladen Vidovich: I'm present. Here.

Leslie Vicale: Thanks, Mladen.

Is there anyone else on the call that has not yet been called on or has joined since I started? OK.

OK. So before I go ahead and get started, I just wanted to note just some ground rules again that we've talked about during the in-person meeting, but I just wanted to remind everyone, of course, that you've come prepared, having reviewed these measures that we will discuss today beforehand. You'll base the evaluation and recommendations on the measures – on the measure of evaluation criteria and guidance.

We ask that you remain engaged in the discussion (inaudible) distractions. And we also ask that you put your phone on mute to reduce any background noise if you're not speaking.

We certainly ask that you attend this meeting to the fullest extent for the entire time that we're on the call. We do ask that you keep your comments concise and focused. Avoid any – dominating any discussion and allow others to contribute. And please indicate any agreement without repeating what's already been said. And we do appreciate your cooperation with all of that.

So, as you may know, I sent you and email just before this call with the voting instruction. We will be voting through a SurveyMonkey. And we ask that you – as we go through the voting questions today that you place your vote as they are asked. Because we are using the SurveyMonkey function, however, we will not be able to announce those votes in real-time. And so as I stated in

the email to the committee, we will keep that SurveyMonkey open and allow you to vote through next Friday, October 2nd Eastern Time.

Again, however, we do ask that you vote during the call while the information is being discussed and fresh in your memory as well as having that vote – the survey open until Friday. Any committee members that are unable to join will have the opportunity to review the transcripts and the recording and place their vote as well.

I like, at this time, to turn the call over to our senior director, Melissa, who will go ahead and facilitate the rest of the call.

Thanks, Melissa.

Melissa Marinelarena: Thank you, Leslie.

Hi, everyone. Thank you again for joining us. We're going to start off where we left off with 2763. This was with constant measure and it was Ischemic Vascular Disease Care. And I believe Mary Gordon is on the phone.

Mary, do you want to give us an update on the additional information that you provided that the committee had asked for. This was on validity testing. We had left off on that one because we lost quorum when we lost – we went under 12 committee members. And then there was questions about your validity testing.

So do you want to provide us with the update, Mary?

Mary Gordon: Sure, I'd be happy to do that. The sections that we had heard then in follow-up that we needed to provide some additional information on where it was available. Primarily, we're Section 2 and it was 2B as in boy and Section 2D as in dog. And we did add some information in 2B22, and that was for each level of testing checked above described the method of validity testing and what it has.

And we had – we had added some information about our process of selecting measures and the various committees that WCHQ has that are involved in the

selection of our measures. I don't know how much detail you want me to get into here, so maybe I'll just be – try to be brief but that we have a Measurement Advisory Committee that is comprised of physicians and some quality directors.

And the majority of those on our measurement advisory committee are also board members as well. So our process goes through a selection process using a policy that we have in place. I did attach that as a reference for – you know, to have as a reference. And those – there's a waiting criteria that goes into our Measurement Advisory Committee.

And then if it is approved by the measurement advisory committee, it goes to our board of directors. And the board of directors has a number of our member and other stakeholder representatives, and 17 of them are clinicians, five are directors or presidents of our member – some of our member organizations. And then we have four representatives from our Regional Quality Measuring and Reporting Association, which is the hospital association would be an example of one of those board members.

So we – and then from that, our measure for the detailed work once it's approved by those two committees, we have an Ambulatory Care Specifications Workgroup that is comprised of various members. It's basically a voluntary group and it's usually data analyst and quality folks who are on. It's a call that we have weekly and so they kind of flesh out the detail with the codes and all of that type of thing and the ease of getting the data electronically and things like that.

So we add an information about that. We – and just that it was very – this measure, in particular, was very unanimously approved by our Measurement Committee and our board of directors, and the fact that it is actively – a measure that's actively in use by our membership for, you know, quality improvement and things like that.

So that we added to Section 2B2. I – then we added – while we didn't – OK, Section 2B5, this was additional testing, if available. We had the reliability testing in 1B and we did not have additional testing to add to 2B5.

And just as a side note, we did include the person at WCHQ who is more of our statistical guru, Matt Gigot. He isn't present today because he's off on a paternity leave, but we did come back.

When we got this additional information that was requested, Matt did come on some calls and, you know, kind of walked through all of this with us so that we did have his input as well. But at any rate, at this – for this point in the measure, it is a newer measure. We did not have any additional testing beyond what was in 1B. And Section 2B6, there is only one set of specs, so and I just added that verbiage that it's just one set.

And the missing data analysis and minimizing bias, which is 2B7, that again was asked to be provided if it was available and we did not have that available. I think I noted that in the document.

And then the empirical analysis to support composite construction approach, what we added – we did not empirical analysis of this as a new measure. We do understand that if it were endorsed that in any follow-up like measure maintenance or anything like that, we understand that the empirical analysis would be required at that time. But for now we just – what we were able to provide was the statistical results of each of the components and so – of each of the component measures that comprised the all-or-none measure. And so we did include those results.

And I believe that is kind of covers the things that we had added.

Melissa Marinelarena: Thank you, Mary. This is Melissa. And what you provided for 2D is required, and that is sufficient.

Mary Gordon: OK, great. Wonderful. Thank you.

Sana, do you want to take up the discussion now?

Sana Al-Khatib: Yes, absolutely, I'd be happy to. Also I think there is a description of this measure for the different members here, so I'm not going to delve into what we have already discussed. I'll just pick it up some validity.

And the – you know, we heard that that they provided some additional information. You may recall when we had our face-to-face meeting, we couldn't really comment on validity because we had some missing information. What we have in front of us today is the face validity that you can find on page 8 of the Measure worksheet that you heard about as well earlier, you know.

So I mean, based on what they said, I think, you know, and the criteria that NQF uses, we certainly allow face validity as the sole test, if you will, of validity. However, I don't think that we could give this anything more than moderate in terms of voting. So as I said, you know, they definitely provided some, I think, convincing data on the face validity aspects of it, but we don't have any additional information or testing regarding validity sizing. The maximum that we can give would be moderate.

Melissa Marinelarena: Great. Thank you, Sana.

So if you have the survey in front of you, unless anybody else has any questions or wants to discuss this further on validity, you can go ahead and vote on the survey, on the SurveyMonkey. And while we're doing that, we will get ready to discuss – we'll go ahead and discuss – read the questions, sorry. We should read the questions.

So we're going to vote on validity, which includes the specification consistent with the evidence, testing and the threats addressed, exclusions, risk adjustment, stratification, meaningful differences, comparability, multiple specifications, and missing data. And your options are High, Moderate, Low, and Insufficient.

And again we don't have this in real-time. We'll get back to everybody with the results. But if you could submit that, if you're online, that would be great.

And now you want to move on to…

Karen Johnson: This is Karen Johnson from NQF. I did want to just point out that face validity is accepted at NQF. It is an option if empirical validity is not

available. However, if you look closely at our criteria, the way through it now is – and I'll just read this is directly for you. Face validity is a measure score that the quality indicator may be adequate if accomplished through a systematic and transparent process by identified experts and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

So, as I was listening and looking at the developer's information, it definitely sounds like they had experts helping with the – maybe the development as we as thinking about some of the data elements and stuff. I was a little unclear and maybe the developer can help with this – with this piece. And your experts actually discussed whether the results themselves would be able to distinguish good from poor quality. That part I couldn't quite get from what you said.

Female: I'm not entirely sure if I understood, but do you mean that those that we had like on our Measurement Advisory Committee, those experts.

Karen Johnson: Yes, yes, yes.

Female: Well, yes, they – and I'm not sure if I'm answering your question or not. But definitely, you know, many of those are – as I said, they're clinician members of WCHQ and then as well as other stakeholders, which are some of our – like business affiliates, purchasers, things like that. And those measures are, you know, very widely used. They're publicly reported and then they are also used by really probably the majority of our membership for internal purposes. So I don't know if that's really explaining – I don't know if I'm answering your question, I guess, though.

Sana Al-Khatib: I think the question that she has is whether this panel of experts – and I actually assume that the case. So, Karen, thank you for asking the question.

If this panel of experts actually like they have looked at all the different aspects of the Measure specification, you know, exclusions, what have you.

Melissa Marinelarena: Oh, yes.

Sana Al-Khatib: And so based on the different elements that the validity in terms of, you know, making sure that you are going to correct – to get the correct data from the specifications that you use if the validity will be good – you know, validity. That's what I understood Karen's question.

But, Karen, certainly I don't want to put any words in your mouth.

Karen Johnson: That's pretty close, Sana. Thank you. Yes, it's probably going even just a little bit further in saying that they have some input or some commentary on the specifications, and it's going a little bit further than that in saying, once the measure is safe and you actually get results, can those results actually differentiate for quality providers from good quality providers?

Sana Al-Khatib: Yes, I would say that very affirmatively, yes.

Mark Kaufman: And this is Mark Kaufman. I'm a physician advisor for WCHQ also on the call. And, you know, many of the organizations use these at the individual provider level. I know the clinic was (inaudible) used to be chief medical officer. We use all the WCHQ measures for physician compensation purposes as well as corporate goals.

So the Measurement Advisory Committee really goes over these very closely. There's every six-month scorecard that's passed out with relative performance. It's all transparent. And the member organizations at WCHQ really take the results and the validity quite seriously, and there's a lot of internal incorporation of those goals into the individual organizations.

Karen Johnson: OK. Thank you.

Melissa Marinelarena: Are there any other questions?

OK. Sana, do you have any comments on the empirical analysis to support the composite?

Sana Al-Khatib: Yes. I mean, I only have a couple of comments there, as was stated by the developer, innovated not do an empirical analysis to support the composite construction, but they did provide, you know, some information regarding the

four individual components for the measure at 17 different clinics. And the people who are looking at the Measure worksheet can see the data in front of them.

So I don't – I don't know that they – you know, they did – they clearly didn't do the analysis, but I wouldn't take that as a negative because I do think that as a negative because I do think that it's probably easy at least critically from the clinical perspective to make the argument that the component measures add value to the composite. And, you know, from reading everything, the waiting rule, I mean, they're giving every one of these components the same weight as the others.

And honestly if we try to – try to give them different ways, I don't know that we can get anybody to agree on the waiting, and this is something that we, as a group, have discussed before. So, I mean, overall I think it's acceptable.

Melissa Marinelarena: Thank you. Does anybody else have any other questions or comments?

OK. So for this, this would be number three in the survey – in the monkey – SurveyMonkey, so we'd ask you to vote on composite analysis including 2D composite empirical analysis, support composite construction, and demonstrate that the component measures stay at the quality of construct, add value (inaudible) to extend possible aggregation and weighting fit quality construct, simplicity to extent possible. And for this, your options are to vote High, Moderate, Low, or Insufficient.

And then we go on to feasibility.

Sana Al-Khatib: From the feasibility perspective, I actually think this is feasible. However, I do have to disagree with the developer on something that they stated under feasibility where they said that data for this measure may be derived from administrative claim.

I certainly can see, you know, people using electronic health record information registry data, things like that to do it. But in terms of administrative claims data, you know, we really don't have that level of detail in terms of blood pressure, you know, measurement. And I really question the

accuracy of any smoking-related, you know, diagnosis or things like that in the administrative claims data. So I disagree that administrative claims data can be used for this measure in any accurate way, but I certainly can see us using electronic and others using electronic health records and registry data to report on the measure sizing that's feasible.

Mary Gordon: Can I comment, Melissa? This is Mary. Just on the question on administrative claims. I think that might be – that might really have been my misunderstanding of what to check there in that respect.

I – when I was reading what type of data was used for testing, I think I was looking at it from all sources. I completely agree that you could not do these four component measures and this all-or-none measure strictly from the administrative claims. I completely agree with that.

We do – for instance, our denominator in our measure is typically derived from like billing or claims data where we're looking at, you know, the office visit CPTs and the service date. And so when I was checking this, I guess, I was – I was kind of looking at it as more of any sources that might be involved, but I – but if – it sounds like I probably misunderstood that so I would agree that administrative claims alone could not be used.

Sana Al-Khatib: Yes. So thank you for the clarification. I mean, I think it was probably the wording that confused me a bit because it says data for this measure may be derived from administrative claims or electronic health records and perhaps maybe rewording that by saying maybe a combination of different sources, you know, including administrative claims and electronic health records, what have you for the different elements that are needed would have been probably better wording to make it clear. But I think we're on the same page.

Mary Gordon: I completely – I completely agree with you, yes.

Melissa Marinelarena: Does anybody else have any other questions? Comments?

Donna Herring: All right. Well, then we're going to move on to voting for feasibility, which includes data generated during care, electronic sources, and data collection

can be implemented. Your choices are High, Moderate, Low, and Insufficient. We'll just here for a second to give you a chance to vote.

OK If you want to move on to usability and use.

Sana Al-Khatib: Yes, I mean, I think the developer presented a very convincing argument there under usability, and you're saying that the measure is being used for quality improvement with benchmarking as well as public reporting. And they talk about their WCHQ publicly reported performance information on group practices. So they actually list different ways by which this measure has been used and reported.

I also do want to commend them for the fact that they said that, you know, 2014 is really the first year this measure was published so they don't have the ability to give us information about the improvement, but that they do intend to collect the data. And I have information on improvement for future reference, so I do commend them for that.

Melissa Marinelarena: Thank you. Any comments, questions on usability?

Donna Herring: All right. We then we'll move to voting for usability and use, which includes accountability and transparency, improvement, progress demonstrated, and benefits outweigh evidence of unintended negative consequences. Your choices are High, Moderate, Low, and Insufficient. And again, we'll pause a moment here.

All right. If you want to move on to overall suitability.

Sana Al-Khatib: So before actually we do this piece, and I think that's for each of us to do that, I do remember that a couple of the people who are on the call today had to leave a bit early when we had the discussion about this measure.

So, for this last question, would it not be perhaps helpful or appropriate to, you know, maybe give a quick summary of how people voted on the first few aspects of this measure?

Melissa Marinelarena: Yes, that's a good point, Sana.

Donna Herring: All right. So if we go back to 2763, for evidence, we had 100 percent of people vote High. For opportunity for improvement, we had 100 percent High. For the composite logic, we had 75 percent High and 25 percent Moderate. For reliability, 50 percent High and 50 percent Moderate. And then for validity is where we lost quorum.

Sana Al-Khatib: OK. Thank you for doing that.

Donna Herring: You're welcome.

Melissa Marinelarena: Are there any other last questions or comments for 2763?

Donna Herring: If not, we'll move on to voting for overall suitability. Does the measure meet NQF criteria for endorsement? Your choices are Yes or No.

Melissa Marinelarena: OK. Thank you, Sana, and thank you, Mary.

Sana Al-Khatib: My pleasure.

Mary Johnson: Thank you. Thank you very much.

Melissa Marinelarena: OK. So the next measures that we are going to be voting on are the AMA-PCPI Measures 0070, 0081, 0083, and these are the registry measures. You voted on the eMeasures in the meeting and so now we're going to be doing the registry side. And we're not going to revote evidence because evidence is the same, but we're going to vote on the rest of the criteria.

So the first one is coronary artery disease, beta-blocker therapy prior myocardial infarction of left ventricular systolic dysfunction, and this is the registry measure.

Leslie or Judd, are you on the phone?

Leslie Cho: Yes, I'm here.

Melissa Marinelarena: OK, OK, perfect. So, is Sam on the phone? Do you want to give an introduction of the measures again, Sam?

Sam Tierney: You know, I think that I could broadly say I could broadly say and I know there was a part of bit of discussion about these measures. I guess, by way of introduction I could say that the three measures that we submitted were developed by workgroups that were convened by the ACC, AHA, and PCPI to identify and define quality outcomes for patients with CAD and HF.

The measures were originally developed in 2003 and then there were formal updates in 2005 and 2011. We developed the measures in alignment with the latest guideline recommendations focusing, in particular, on the evidence-based therapies that have been proven to have a significant impact on morbidity and mortality, and that still address areas that remain in need of performance improvement. So, I guess, at a high level, that's a short and sweet introduction.

Thank you.

Melissa Marinelarena: Great. Thank you. OK. Leslie and Judd, do you want to start?

Leslie Cho: I'm happy to lead. We voted for the evidence in this case and the evidence is extremely high. And I think it passed on the evidence protocol. In terms of the gap, there still continues to be gap in terms of patients not getting on this standard, well-recognized therapy so there definitely continues to be a gap.

So I think where we started out in our voting is, according to my (inaudible) performance gap, and it – and the reason why this performance – the performance gap that they provide in the measure is not the eMeasure, but the registry measure. And there continues to be a gap year-to-year (inaudible) patients around 30 percent of patients not getting beta-blocker therapy, so the performance gap is high.

Judd Hollander: And I would agree with that. This is Judd.

Donna Herring: Are there any questions or comments regarding the performance gap? If not, we'll move to vote.

For performance gap, you're voting on whether the data demonstrated a considerable variation or overall less than optimal performance across providers and/or population groups. Your choices are High, Moderate, Low, and Insufficient.

At this time, we'll move on to reliability.

Leslie Cho: So, in terms of (inaudible) measures have actually been used for a while and the data comes to us from PQRS as well as the PINNACLE Registry data, and so the reliability is also quite high. And they've been around for a long time.

The reason why it's different between eMeasures and the registry measures is that in eMeasures we didn't have any data. But for registry, there is actually a good reliability testing.

Judd Hollander: I have nothing to say as to that. I think that's exactly the way I feel about it as well.

Mary George: This is Mary George. So what kind of reliability testing was done?

Leslie Cho: So the PQRS, the GPRO database was queried.

Female: Yes.

Leslie Cho: And the sample shows actually that it's very reliable and the quality rating events were point – was 0.92, so very good reliability, very high reliability.

Kristi Mitchell: Leslie, this is Kristi. Did they provide data based upon the PQRS website, the web interface as well as it relates to reliability?

Leslie Cho: It's the GPRO database. Is that the web?

Kristi Mitchell: So, there's GPRO Registry, but then there's CMS PQRS – excuse me, EHR web…

Leslie Cho: No, so…

Kristi Mitchell: … interface.

Leslie Cho: So, the eMeasure testing is – I thought we were not going to talk about it.

Kristi Mitchell: It's not the eMeasure testing.

Leslie Cho: I don't know. Maybe the developers can answer.

Sam Tierney: Yes, hi, this is Sam Tierney. Thanks for the question.

So, the analysis that was done was a signal-to-noise analysis and it comes from the GPRO Registry not…

Kristi Mitchell: So…

Sam Tierney: … the EHR web interface (registration).

Kristi Mitchell: OK. So – OK. But for the other – for at least 83 it is – it's both, correct?

Sam Tierney: For 83, it is both, yes.

Kristi Mitchell: Thank you, thank you.

Melissa Marinelarena: Are there any…

Judd Hollander: Yes, so the – so the only comment I'll add to it and I don't think it necessarily will affect the voting, the (inaudible) is going to be a 0.19 was that the average number of reporting events (inaudible) 61. So when we went down to the minimum reporting events, it dropped to 0.65, which is a little lower than we like. But I think the overall that it's very good, I don't think that's a huge issue (inaudible).

Melissa Marinelarena: Are there any other comments or questions?

OK. We'll move on to the vote.

Donna Herring: You are voting on reliability, including precise specifications and testing. Your choices are High, Moderate, Low, and Insufficient.

Melissa Marinelarena: OK. Thank you.

Donna Herring: At this time, we move on to validity.

Leslie Cho: So, the validity testing was done. This is a validity testing part, right? Yes, the validity testing part. So the validity testing was done and actually the testing result show that 92 percent of respondents agreed with the testing. And then of the threat – in terms of threat to validity, I think the – there was a total of 4,000 patient exception noted in terms of why they were not put on beta-blockers. And the overall exception is around 3.9. I mean, I think that's a reasonable – I think that's very reasonable. I think in terms of validity, this measure passes a validity test.

Judd Hollander: So, I'm going to – I'm going to agree with that. I was told that in the future for measures that has been around for a long time that we get an analysis of missing data that we don't have here. We should actually be able to figure out and we should be informed as a committee on what happens in cases of missing data. So I don't throw that out to impact the voting, but really for something that hopefully questionable for in future assessment of measures that have been around for a while.

Donna Herring: Does anyone have any additional comments or questions regarding validity? If not, we'll vote.

Validity including specification consistent with evidence testing and threats addressed, exclusions, risk adjustment, and stratification, meaningful differences, comparability, multiple specifications, and missing data. Your choices are High, Moderate, Low, and Insufficient.

Moving on to feasibility.

Leslie Cho: The feasibility has been around for a long time. It's very feasible. I'm using the PINNACLE database and the PQRS database, so it's very feasible.

Melissa Marinelarena: Before we go any further, we just wanted to double-check that the committee has been able to place their votes in the SurveyMonkey, and that they haven't experienced any issue.

Female: Yes, I have no issues.

Male: I'm doing fine at this time.

George Philippides: I'm having a hard time finding the next button to go to the second metric.

Melissa Marinelarena: OK, George. Yes, we're checking the survey on our side and we're not seeing any trouble. Perhaps you might be able to refresh your page.

George Philippides: Will do. Thank you.

Melissa Marinelarena: OK, thanks.

Donna Herring: So just to bring it back, does anyone have any questions regarding feasibility or comments? OK. We'll vote.

Feasibility including data generated during care, electronic resources, data collection, and data collection can be implemented. Your choices are High, Moderate, Low, and Insufficient.

OK. We'll move on to usability and use.

Leslie Cho: So, currently it is being used in the PINNACLE Registry, in the Meaningful Use, Stage 2, and in the PQRS. And so I think for usability and use, it's very high as well.

Tom James: This is Tom James. I got a question about that. Is the use in PQRS (inaudible) of the PINNACLE Registry or is it done independently? In other words, is this a measure which really requires access to PINNACLE?

Kendra Hanley: This is Kendra from the PCPI. It does not require access to PINNACLE, but PINNACLE is also a registry that reports the data through the PQRS program.

Tom James: OK, sure.

Kendra Hanley: So – yes. Sure.

Judd Hollander: This is Judd. I'll just add to that and really it's a little bit more of a question because according to the usability voting template there it says for the improvement, progress demonstrated.

And when we look over, you know, the gap and the data that's given, it's really all over the map in the last four years and there is not a trend towards improvement. So I think although it's easy to use it, there is no demonstrated progress made and that's one of the things you asked us about.

Kristi Mitchell: This is Kristi. I have a follow-on question, comment as well regarding the accountability and transparency sub-bullets which speaks to if a measure has been endorsed which this one has been since '09, I think, that public reporting should be within six years. Is this still applicable to this measure, if my assumption is correct about the 2009 endorsement?

Leslie Vicale: Yes, that's correct. Although this is not – just a reminder, this is not must – a must-have criterion, but yes, that is our guidance. If they're to be used for accountability purposes within three years and then public reporting after six years.

Sam Tierney: So this is Sam Tierney over at PCPI. Could I address those questions?

Kristi Mitchell: Yes.

Melissa Mariñelarena: Go ahead, Sam.

Sam Tierney: OK, great, thanks. So as it relates to demonstrated improvement, I definitely can appreciate the comment and looking at the data, as you said, it is all over the map. One of the challenges in the data from the PQRS program is that the rate of participating professionals have changed from year to year and that PQRS is in of itself testing, and so it had started as a program that rewarded providers for participating and it has moved into a penalty phase. So the participation PQRS has increased over the years quite a bit.

So the physicians that we're looking at when we look at their performance over the years and PQRS has probably changed quite a bit, so it's not like we're looking at a consistent population for whom we would expect to see

improvement and hope to see improvement over time. The population that you're looking at when you look at those percentages has changed quite a bit from year to year, if that makes any sense.

And then the other question about public reporting, so CMS, as you all may know, is moving towards public reporting via the Physician Compare program. And based on the recent rule, it does mean that this measure will be publicly reported in the next maybe year or so based on the rule, but it is somewhat out of our control related to the implementation of the measure and the government program.

Secondarily, and I am not sure if my colleagues from ACC are on the phone but the measures also being used in the PINNACLE Registry and the PINNACLE Registry is a QCDR and there is a public reporting requirement as part of the QCDR implementation in the PQRS program.

So I do believe the measure would be publicly reported via that mechanism, and certainly, it does appear based on the signals in the proposed rule that the measure should be publicly reported in the next year or so through the Physician Compare website.

Melissa Mariñelarena: Thank you, Sam. Are there any other questions or comments?

Donna Herring: All right. At this time, we will vote on usability and use which includes accountability and transparency, improvement and progress demonstrated and that the benefits outweigh, evidence of unintended negative consequences. Your choices are high, moderate, low and insufficient.

All right. We'll move on to overall suitability.

Kristi Mitchell: So I think I would vote yes on the overall suitability of this measure. I think that the evidence is very good, is very high, as well as the usability and the feasibility and the validity, so – and we continue to see gaps, so I would vote yes on the measure.

Donna Herring: And just to review, during the in-person meeting, we voted that the evidence was 100 percent high. Are there any other questions or comments regarding

overall suitability? OK, we'll vote. Overall suitability, does the measure meeting NQF's criteria for endorsement? Your choices are yes or no.

Leslie Vicale: OK, thank you, everyone. And we will move onto Measure 0081.

Donna Herring: OK, so 0081 is heart failure, an ACE inhibitor or ARB therapies for LVSD, and this is the registry measure, so I don't believe it's on the phone yet but Mladen I know is on. Mladen, do you want to start with the performance gap?

Mladen Vidovich: Yes. So the performance gap is there. It seems that it's prescribed in about 80 percent of eligible patients. And if you look further down in the sheet that they have provided, it looks about 85 percent in 2010, 80 percent in 2011, 82 percent in 2012, 83.5 percent.

So clearly, not everybody is getting it but it seems it's stuck at the same – at the same level for several years. There's just a little bit of a question, can this go any higher than this where this measure is showing.

There is also a question about disparities. But actually interestingly, although there are some data in the literature to show this from the TRICARE, they are actually – they have not been able to analyze this and report this.

So despite that actually we know this is happening, (inaudible) they're not providing this. So I'd say it's somewhere, I would recommend so between moderate and high because I don't think you can go much, much, better than this but then again, it would be helpful to have some of the disparity data available.

Sana Al-Khatib: Can I ask you a question? Does the measure allow for contraindications, you know, patients who couldn't tolerate the medication, the medication has contraindicated for some reason?

Mladen Vidovich: Yes, it does, and let me – let me just…

Sana Al-Khatib: Yes.

Mladen Vidovich: Let me just…

Sana Al-Khatib: Because if that's the case then I completely agree with you that, you know, it's not likely – the percentage is not likely to get much better.

Mladen Vidovich: Let me – let me just go down like, you know, to the numerator and denominator section so I can – I can accurately answer your, you know, question. I mean just I'm scrolling down. I think it does (inaudible) right now.

Sana Al-Khatib: Yes.

Mladen Vidovich: Maybe somebody – a developer can help me out here.

Sam Tierney: Yes, this is Sam. I'd be happy to answer that.

Mladen Vidovich: OK, then thank you.

Sam Tierney: Thank you. So, yes, we do have as part of the measure medical reasons for not prescribing an ACE inhibitor or ARB therapy which would include contraindications. And the rates that are presented are after the exceptions have been accounted for. So that's the performance, again, after the exceptions are removed from the denominator. So because of the inclusion of exceptions in the measure, you could expect to see higher rates, but it's still not there.

Mladen Vidovich: OK. Thank you.

Sam Tierney: Thanks.

Sana Al-Khatib: Thank you.

Melissa Mariñelarena: Are there any other questions? OK, we'll move to a vote.

Donna Herring: For performance gap, the data demonstrated considering – considerable variation or overall less than optimal performance across providers and/or population groups. This raises our high, moderate, low and insufficient?

Melissa Mariñelarena: OK. We'll move on to reliability.

Mladen Vidovich: So this was tested in the PQRS GPRO database and they got a 0.83 score. So I think that's really within the level of what we've voted before on similar measures. I think that it has adequate reliability to be accepted.

Melissa Mariñelarena: Does anyone else have any additional questions or comments regarding reliability?

Sam Tierney: This is Sam, if I could just add to the previous comments, so the minimum of number of events, the reliability was 0.83, but if the average, which was 31.5, then reliability was 0.94.

Melissa Mariñelarena: Thanks, Sam. Are there any other comments or questions? So we are to the vote?

Donna Herring: Reliability including precise specifications and testing, your choices are high, moderate, low and insufficient.

OK, we'll move on to validity.

Mladen Vidovich: The validity was done and actually had a little bit – a little bit of a question that I didn't quite understand. It seems that there is some difference between the registry testing and the eMeasure testing, and they were not the same.

So I would like to ask the developer what this truly means because it says – just to read this, they found the exception rate of 17 percent, but then the developer is unable to determine the type of exception reported in the PQRS registry. So why is there – why is there such a problem here?

Kendra Hanley: This is Kendra from the PCPI. The way that CMS – how he data reported does not break the exceptions into medical patients or system reason, they just receive the data meaning the patient had a valid exception overall. So that's what we meant by saying we couldn't distinguish the type of exception because we have those three categories, but that is more a function of how the data is recorded to CMS.

Mladen Vidovich: Yes, because otherwise, it seems that the mean performance rate is 0.8, so it seems it works here just like the previous beta blocker measure, I just had that question.

Melissa Mariñelarena: Are there any additional questions or comments regarding validity? OK, we'll vote.

Donna Herring: Validity including specifications consistent with evidence, testing and threats addressed, exclusion, risk adjustment and stratification, meaningful differences, comparability, multiple specifications and missing data, your choices are high, moderate, low and insufficient.

We'll move on to feasibility.

Mladen Vidovich: So this measure is feasible and it's feasible both as an eMeasure and – in an eMeasure and I think there's not much to say extra about this.

Melissa Mariñelarena: Are there any other comments or questions?

Donna Herring: OK, we'll vote, feasibility including data generated during care, electronic sources and data collection can be implemented. Your choices are high, moderate, low and insufficient.

We'll move on to usability and use.

Mladen Vidovich: It is a – it is a publicly reported measure from multiple registries, PINNACLE, Meaningful Use and PQRS. So it's currently new and I think we aligned this dataset as high.

Melissa Mariñelarena: Are there any comments or questions? OK, we'll move on to the votes.

Donna Herring: Usability and use including accountability and transparency, improvement, progress demonstrated and benefits outweigh evidence of unintended negative consequences. Your choices are high, moderate, low and insufficient.

OK, we'll move on to overall suitability.

Mladen Vidovich: I think this is suitable measure. I think just, as the committee, we discussed in our live meeting, is some measures seemed to be reaching their peak target levels, and I think this is one of those. You know, it seems to be stuck at 80 percent. I don't know how much more we're going to get them this but again, that remains to be determined. It is suitable but my overall gestalt will be at some point if this will plateau, so I would say yes at this time.

Melissa Mariñelarena: Are there any other comments or questions before we vote on the overall suitability of the measure? Hearing nothing, we'll move to a vote.

Donna Herring: Overall suitability, does the measure meet NQF criteria for endorsement? Your choices are yes or no.

Melissa Mariñelarena: OK, thank you very much. So we will move on to Measure 0083.

Donna Herring: OK, so 0083 was the heart failure beta blocker therapy for LVSD. This is the registry measure and this is Mary and Kristi. Do you like to start the discussion with performance gap?

Mary George: This is Mary, and I'll go ahead. The performance gap provided with the PQRS data from 2010 through 2013 which are the performance bouncing around over those years between 75 percent and 85 percent without sustained improvement over the four years that were presented. And that they said that this was consistent with results from the improved HF registry.

Disparities were not reported and were not available at this time. They quoted the same TRICARE study that was previously mentioned and found that some African-Americans were less likely to receive treatment.

Melissa Mariñelarena: Are there any questions or comments?

Donna Herring: OK, we'll move to a vote. Performance gap, data demonstrated considerable variation or overall less than optimal performance across providers and/or population groups. Your choices are high, moderate, low and insufficient.

All right. We'll move on to reliability.

Mary George: So the numerator specifications are provided for both inpatients and outpatients. Prescribed is defined as either a prescription given to the patient or on the medication list. Denominators are documented. The LVEF less than 40 percent or documentation of moderate or severe dysfunction. And the data elements seemed to be well specified with ICD-9 or ICD-10-CM codes and CPT codes.

In terms of the reliability testing, it was done with the data by (Domio) model using a signal-to-noise analysis.

They had 684 physicians with at least 10 records in all data elements available for their testing. That was 39 percent of the physicians were included in this testing. The average number of events reported by – per physician was about 34 and a total of more than 23,000 events for the testing.

The testing for those with a minimum number of events which was 10, the signal-to-noise ratio was 26, and for those reporting at the average number of events, it was 0.96. There – so that would seem to indicate high to moderate rating.

Kristi Mitchell: I have a question, this is Kristi, for the developers. And this is again, about the presentation of the PQRS EHR Web interface and how doing a reliability testing using that data source is related to the registry measure or not, it's just wasn't clear in the write-up.

Sam Tierney: So this is Sam again. So it's not, I would actually say that EHR – the data that is the EHR Web interface data is actually more like testing data from an electronic source. So it would be more specific to, like the eMeasure testing that was discussed at the in-person meeting. And this is a different modality, different process altogether related to their registry.

Kristi Mitchell: Thank you.

Melissa Mariñelarena: Are there any other questions or comments regarding reliability? Ok, we'll move to a vote.

Donna Herring: Reliability including precise specifications and testing, your choices are high, moderate, low and insufficient.

All right. We'll move on to validity.

Mary George: So face validity with testing, you're then using 12 experts who are asked to respond as to whether the measure would provide an accurate reflection of quality and to distinguish between good and poor quality using a 5-point Likert scale. Of the 12 responses, eight were agree and the other four were strongly agree.

Their validity testing on the measure itself showed a mean performance rate of 0.7, median 0.93, mode 1.0, range 0 to 100, and IQR was 0.42 to 1 from the registry data.

In terms of (perhaps the) validity, their total exclusions averaged 4.9 percent overall with an average per physician exclusion or exception being 1.8 among the 684 physicians.

We previously discussed the exclusion in all of these PCPI measures which include medical reasons, patient reasons and system level reasons, but it appears as though these exceptions aren't being used in a very wide number of cases.

The measure is not risk-adjusted or risk-stratified. And there was no missing data analysis because their testing was done using records that had complete data.

Melissa Mariñelarena: Are there any questions or comments regarding the validity? Hearing nothing, we'll move to a vote.

Donna Herring: Validity including specifications consistent with evidence testing and threats addressed, exclusions, risk adjustment and stratification, meaningful differences and comparability for multiple specifications and missing data. Your choices are high, moderate, low and insufficient.

OK, we'll move on to feasibility.

Mary George: The data elements are available when electronic data and medical records that seems to be quite feasible.

Melissa Mariñelarena: Are there any questions or comments regarding feasibility? OK, we'll vote.

Donna Herring: Feasibility including data during care, electronic sources and data collection can be implemented. Your choices are high, moderate, low and insufficient. OK, we'll move on to usability and use.

Mary George: So this is currently used in PQRS Meaningful Use Stage 2, QI and benchmarking in the PINNACLE Registry and previous comments related to public reporting and PQRS would probably acquire this as well.

Melissa Mariñelarena: Are there any questions regarding usability and use? OK, we'll vote.

Donna Herring: Usability and use including accountability and transparency and improvement, progress demonstrated and benefits outweigh evidence of unintended negative consequences. Your choices are high, moderate, low and insufficient.

OK, we'll move on to overall suitability.

Mary George: Just to remind the committee that our previous vote on the evidence was 88 percent high, 13 percent moderate.

Melissa Mariñelarena: Are there any remaining questions or comments before we vote on the overall suitability of the measure? Hearing nothing, we'll move to a vote.

Donna Herring: Overall suitability, does the measure meet NQF's criteria for endorsement? Your choices are yes or no.

OK, thank you, everyone. That concludes the measures for the AMA-PCPI registry measures. I'll turn it back to Melissa,

Melissa Mariñelarena: OK. So the last thing that we wanted to bring to your attention was ACC's Measure 0965, discharge medications, ACE, ARB, beta blockers and eligible ICD implant.

Patients – this measure passed on other criterion. The thing was that for 2D, the composite concept, it didn't actually have the testing – the testing of the components. So we asked for that and we provided it to Joseph and Judd to take a look at that.

So we just wanted the committee to review that and ask you if you wanted to do a vote on 2B after you review the results if you're comfortable with the data that's provided. And just as a reminder – do you have the voting results for 2D? We'll tell you what the results were, but we just want you to look at that.

And if you're in the measure worksheet, it's on page 77, and it's (in round) statement, its distribution of the composite measure and its components for ACE, ARB and beta blockers.

Joe or Judd, do you want to summarize the findings? Judd, I believe that you're on the line. Did you have anything to add about the additional information? Judd, are you available just to ensure that if there are any additional comments or not.

Judd Hollander: Yes, yes, I'm sorry. I have construction going on outside my house so the (inaudible) much noise. So I wasn't connected for a minute but…

Melissa Mariñelarena: OK.

Judd Hollander: Yes. So I looked at this. I'm OK with those pages. In fact, I'm totally OK with this anyway, but I don't understand the table and I was hoping that a developer could just answer to some questions.

But for those of you that – those that have the table in front of you, it is both the composite measure, a volume and a value as well as for the ACE and ARB and then one for the beta blockers.

What I see when I'm looking at the volume going across, if I just pick the median as an example, but the same is more or less true in all of that, I have a

volume for the composite measure of 168 but I only have 134 for the ACE and only – and 292 for the beta blockers.

So I don't actually understand how I could have a higher volume from the same dataset for people that can be assessed for the composite measures and then pick either of the individual measures.

Melissa Mariñelarena: (Fidelity), are you on the phone or somebody else from ACC?

Female: Jensen.

Jensen Chiu: Yes. This is Jensen Chiu.

Melissa Mariñelarena: Hi, Jensen.

Jensen Chiu: ACC staff. Am I able to – can everybody hear me or no?

Melissa Mariñelarena: Yes.

Female: Yes, we can hear you.

Female: Go right ahead, Jensen.

Melissa Mariñelarena: Thank you.

Jensen Chiu: OK, OK. Yes, I think the comment Dr. Hollander brought is a good comment on the table. I just want to bring one step back and I think the reason why we brought the data forward as well is because I think what we've reviewed originally was just the composite data.

So this obviously I think Dr. Hollander is correct but I just want to like whole steering committee make them aware that, you know, this is reflective of individual components of the measure because I believe either it wasn't attached or it was – it wasn't attached in the earlier document you, guys, have see.

So I think the question at hand really is, are the individual scores really, are they consistent to the overall construct and (inaudible) performance rates.

I think, you know, one of the questions is, does the rate – do the rates improve overtime. I knew that came up earlier and that definitely shows that in some of the other materials. But the second question I think at hand right now here is looking at the volume and value.

So the volume is basically – this is in-hospital measure, not like the other measures that have been shown earlier to the outpatient. This is inpatient measure for the electrical physiologist colleagues.

So the N would be like 1606 would be the hospitals. Pretty much the majority of hospitals in the U.S. are in the ICD registry. That's what this measure is for. So that's the N right there.

And in terms of why the – you know, the volatility of value and volume are different, let's just do the volume first. So the volume, as you can see, the number is pretty high, a lot of hospitals are between the 90 percent up through – you know, through the 99 percent. A lot of the – a lot of the hospitals are in that.

Now, every – due to the exceptions and I know we're not supposed to talking about all that but some of the specifications of the exceptions are partly the reason why numbers come out the way they do and that some people that are included in one of the measures, when you get to the composite, they're not necessarily included, so that's why they kind of drop off to some extent.

Judd Hollander: Right, that's what I'm seeing. I'm seeing the opposite. I'm seeing that the quality of the composites exceeds the volume of either one of the individual measures. It seems to me that that should be impossible. It has to be that the volume that is measured on the composite has to be less than the lowest of the individual measures.

Jensen Chiu: I figured what you're saying, yes, I'm taking a look at that right now.

Judd Hollander: And that seems to be across, you know, every percentile so they question the accuracy of the data or whether the volumes are accurately labeled.

On the other hand, the value and how well it works seems to work the way it should be. So what I'm seeing is that the composite measure is slightly lower than the lowest of the individual measures. So that made sense but the volume doesn't make sense to me.

Melissa Mariñelarena: Jensen, is there anything else you'd like to add?

Jensen Chiu: I think of this and I would need to take a backup from our statistician to confirm why that's the case in the – on the volume, why that's the case. It could potentially be just an error that was done because I do see the point Dr. Hollander is bringing up, why the – why the volume for the composite is slightly off.

Judd Hollander: Yes. So I think I mean I would make a recommendation that we have to defer on a further vote because we'd be voting on data that we don't understand the accuracy of.

Tom James: It's Tom James. Judd's comments make sense to me.

Judd Hollander: But on the other hand, I would propose that where we get the accurate data, we thought we don't need to do the phone call because that's just not – so, you know, we test this already.

And I think based on the data that I'm seeing based on the values, it makes sense and there are some justifications for the composite measures. So I think it's consistent with our initial vote but I just think for record keeping, you know, we can't finalize the vote based on (inaudible) probably additional.

Melissa Mariñelarena: OK, good, thank you, everyone. So – did I hear someone else wanted to make a comment? OK. OK, so at this time, based on the information that's been provided, we would allow the community to go ahead and revote today on this criterion.

And during the public comment period, comments can be submitted and the committee will have an opportunity to consider any comments or information that is provided during that time and can make any other determinations

during the post-comment call. So we would go ahead and ask the committee to revote on this component today.

(Lisa): Yes. And this is (Lisa) from NQF. I think on one committee decision is on this measure today, if it doesn't go forward, if the committee doesn't recommend it, ACC and Jensen, if you could please gather information, the information that the committee is seeking for the measure and provide it at the post-comment call, that would be great. That's part of our process as well.

Jensen Chiu: OK, we'll certainly do so.

Melissa Mariñelarena: Thank you. So at this time, are there any other comments or questions from the committee?

Tom James: This is Tom James again. If we don't vote in favor that it will come back, correct?

Leslie Vicale: Yes. So ACC can put in a formal reconsideration request and we can reconsider that. We can put it in during the commenting period and during the post-comment call, the committee can reconsider it, vote to reconsider and vote again based on the information they provide.

Tom James: Great.

Melissa Mariñelarena: OK. We will – we'll go ahead and move to a vote on this criterion.

Donna Herring: All right. Composite analysis including empirical analysis, support composite construction and demonstrate, component measures, state quality construct, add value, personal need to extent possible, aggregation and weighting, state quality construct, simplicity to extent possible. The new choices are high, moderate, low and insufficient.

Leslie Vicale: OK. So that concludes the discussion and the review of the measures that were on the agenda for today. As you all know, the SurveyMonkey that the committee have been sent will be open until next Friday. We ask that you complete that survey and place your votes no later than 5 P.M. close of business next Friday.

In addition, NQF will provide a summary to the committee members who are unable to join the call today, or for those of you that would like to look back, and we will be providing the transcript recording information so you can also listen back to assist any voting that you don't need to provide.

And at this time, we would like to move to the NQF member and public comments. So I would like to ask the operator to go ahead and open up the lines for any comments.

Operator: Thank you. If you would like to make a comment, please press star then the number 1 on your telephone keypad. We'll pause for just a moment.

We do have a public comment from Collette Pitzen.

Leslie Vicale: Yes, go ahead.

Collette Pitzen: Great. Hi. This is Collette Pitzen. I'm a measure developer with Minnesota Community Measurement. And I just wanted to bring awareness to the fact that there is already an existing Optimal Vascular Care Measure number 0076. That's been endorsed since 2009. The last maintenance review was in June of 2012 and the measure is scheduled for the committee's review in your next phase of work.

Measure 2763 that was presented today is a competing and duplicate measure, and we would hope that the committee would consider the issue of a competing and duplicate measure prior to its recommendation for endorsement. Thank you.

Operator: And there are no further comments at this time.

Leslie Vicale: Thank you very much. And we would like to note that follow-up to that comment that the next committee post-meeting call which will be held on October 9th from 2 to 4 P.M. Eastern Time, the agenda for that call is to discuss any related and competing measures. So please note, we will have that discussion at that.

I would also like to note that the public and member comment period was a little bit earlier that previously noted on the agenda during today's call. And I wanted to remind everyone that the formal public and member comment period will begin on October 16th and run through November 15th. And this is when the draft report will be posted for that public comment.

So again, we'd like to thank everyone for joining us today and I wanted to recap some of the next steps and the timeline for the CV – the cardiovascular stage 3 measure endorsement projects. And as you can see here, like I had just noted on our post-meeting call, the second call would be October 9th for the related and competing discussion. And again, the draft report will be posted on October 16th for 30 days for public comments.

We will have a standing committee call to review in response to those comments on December 7th from to 1 to 2 PM Eastern Time. The draft report will be posted for NQF member votes from December 18th through January 3rd, that, I should say, 2016.

The CSAC review and approval will be from January 6th to the January 26th of 2016, followed by endorsement by the board, January 27th to February 5th of 2016. And the appeal period will be open from February 8th until March 9th of 2016.

Again, we'd like to thank everyone for joining us today and we'd like to thank the committee members who have joined us and for placing your votes through the SurveyMonkey link. If there are any questions or concerns, please do not hesitate to contact the NQF staff.

So at this time, we'd like to adjourn the meeting for the day. Thank you again, and we hope that you all enjoy the rest of your afternoon. Thank you.

Male: Thanks.

Female: Thanks.

Female: Thank you.

Male: Thank you.

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

END