NATIONAL QUALITY FORUM

+ + + + +

CARDIOVASCULAR STANDING COMMITTEE

+ + + + +

CARDIOVASCULAR MEASURE ENDORSEMENT PROJECT 2015

+ + + + +

WEDNESDAY
SEPTEMBER 9, 2015

+ + + + +

The Cardiovascular Standing Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 9:00 a.m., Mary George and Thomas Kottke, Co-Chairs, presiding.

PRESENT:

MARY GEORGE, MD, MSPH, FACS, FAHA, Co-Chair
Senior Medical Officer, Centers for Disease Control and Prevention, Division for Heart Disease and Stroke Prevention
THOMAS KOTTKE, MD, MSPH, Co-Chair, Medical Director for Population Health, Consulting Cardiologist, HealthPartners
SANA AL-KHATIB, MD, MHS, Associate Professor of Medicine, Duke University Medical Center
LINDA BRIGGS, DNP, Assistant Professor, George Washington University, School of Nursing
LESLIE CHO, MD, Section Head, Preventive Cardiology and Rehabilitation, Cleveland Clinic
JOSEPH CLEVELAND, MD, Professor of Cardiothoracic Surgery & Surgical Director for Adult Cardiac Transplantation/Mechanical Cardiac Assist Devices, University of Colorado Denver

MICHAEL CROUCH, MD, MSPH, FAAFP, Research Director and Quality Improvement Program Director, Memorial Family Medicine Residency Program and Associate Clinical Professor of Family Medicine, Texas A & M University School of Medicine

ELIZABETH DELONG, PhD, Professor and Chair, Department of Biostatistics and Bioinformatics, Duke University Medical Center

ELLEN HILLEGASS, PT, EdD, CCS, FAACVPR, FAPTA American Physical Therapy Association

JUDD HOLLANDER, MD, FACEP, Associate Dean, Strategic Health Initiatives, Sidney Kimmel Medical College, Professor, Vice Chair of Finance and Healthcare Enterprises, Department of Emergency Medicine, Thomas Jefferson University

THOMAS JAMES, MD, Chief Medical Officer, Baptist Health Plan and Baptist Health Community Care (via telephone)

JOEL MARRS, Pharm.D., FNLA, BCPS (AQ Cardiology),CLS Assistant Professor, Department of Clinical Pharmacy, Skaggs School of Pharmacy and Pharmaceutical Sciences, University of Colorado Anschutz Medical Campus, American Society of Health-System Pharmacists

GERARD R. MARTIN, MD, Senior Vice President, HLK, Medical Director, Global Services, Children's National Health System

KRISTI MITCHELL, MPH, Senior Vice President, Avalere Health, LLC

GEORGE PHILIPPIDES, MD, Chief of Cardiology, Newton-Wellesly Hospital

NICHOLAS RUGGIERO, MD FACP FACC FSCAI FSVM FCPP, Director of Structural Heart Disease and Non-Coronary Interventions, Thomas Jefferson University Hospital

JASON SPANGLER, MD, MPH, FACPM, Executive Director, Medical Policy, Amgen, Inc.

MLADEN VIDOVICH, MD, Chief of Cardiology, Jesse Brown VA Medical Center

NQF STAFF:

HELEN BURSTIN, MD, MPH, Chief Scientific Officer
JASON GOLDWATER, MA, MPA, Senior Director
ANN HAMMERSMITH, JD, General Counsel
DONNA HERRING, MPH, Project Analyst
LAURA IBRAGIMOVA, MPH, Project Analyst
KAREN JOHNSON, MS, PhD(c), Senior Director
MELISSA MARINELARENA, RN, MPA, Senior Director
LESLIE VICALE, MPH, Project Manager
ASHLIE WILBON, MS, MPH, FNP-C, Managing Director

C O N T E N T S

P-R-O-C-E-E-D-I-N-G-S

9:04 a.m.

MS. VICALE: Thank you everyone. We'd like to welcome you this morning. This is the National Quality Forum. My name is Leslie Vicale, and I'm the Project Manager for the Cardiovascular Project.

This is Phase 3 of the National Consensus Standards for the cardiovascular condition.

I'd like to welcome everyone to the meeting, like I said, and at this point I'd just like to ask the Co-Chairs to give some opening remarks.

CO-CHAIR GEORGE: Well, I just really want to welcome everyone. I know this is a lot of work and takes a lot of time and effort on your part to put into the work that we do, and I just really appreciate all the work that you do.

CO-CHAIR KOTTKE: Yeah, I'll just welcome everybody and sort of second what Mary said, and we'll get on with the show.

MS. VICALE: Thank you, Mary George and Tom Kottke.

I'd like to turn it over now -- I'd like to turn it over now to Ann Hammersmith, who is going to provide introductions and the disclosure of interest.

MS. HAMMERSMITH: Good morning, everyone. I am Ann Hammersmith. I'm NQF's General Counsel.

As Leslie said, we will combine introductions and disclosures because it's a little bit quicker and more efficient to do it that way.

Those of you who have been on the Committee have done this before. You've heard my little speech, but I'll give it again anyway.

Just to remind you, you all received a disclosure of interest form from us in which we asked you detailed information about your professional activities. We take that into consideration, of course, when we seat Committee members.

For those of you who have been on the Committee, this is the annual update, and you got a form, too, to fill out and update as needed.

So what we'll do now is we'll go around the table, we'll ask you to identify yourselves, tell us who you are with and if you have anything to disclose.

Please don't summarize your resume. We are only looking for you to disclose any activities that are directly related to the subject matter of the meeting today.

So if for some reason, even though this is the Cardiovascular Committee, you did a lot of work in dermatology, we don't want to hear about that. We just want to hear about heart stuff.

We're particularly interested in any research grants, speaking engagements that you may have engaged in, but only if they're directly related to the subject matter before the Committee.

I want to remind you that you sit as individuals. You are here because you are a subject matter expert. You don't represent your

employer, nor do you represent anyone who may have nominated you to serve on the Committee.

So with that, I always start with the Co-Chairs, and we'll go around the table.

CO-CHAIR KOTTKE: Tom Kottke, Consulting Cardiologist for HealthPartners Medical Group and Medical Director for Population for HealthPartners. I do sit on the NCQA CV Measurement Advisory Panel, so I'll recuse myself from the NCQA measures.

CO-CHAIR GEORGE: Mary George, I'm with the Division for Heart Disease and Stroke Prevention at CDC, where I'm the Deputy Associate Director for Science and Senior Medical Officer, and I have no disclosures.

MEMBER PHILIPPIDES: Good morning. I am George Philippides, Chief of Cardiology at Newton-Wellesly Hospital. I'm on the Founders Board of the American Heart Association.

MEMBER AL-KHATIB: Good morning. I am Sana Al-Khatib. I am an Associate Professor of Medicine at Duke University. I'm an

electrophysiologist, and I have no conflicts.

MEMBER CHO: Leslie Cho, Cleveland Clinic, Section Head for Prevention. I have nothing to disclose.

MEMBER MITCHELL: Kristi Mitchell, Senior Vice President of Avalere Health, and I have nothing to disclose except that company was recently acquired.

MEMBER HILLEGASS: Hi, Ellen Hillegass, Mercer University in Atlanta, Georgia. I have nothing to disclose.

MEMBER DELONG: Liz Delong, Duke University. Biostatistician, and I have nothing to disclose.

MEMBER CROUCH: Michael Crouch, family physician in Sugar Land, Texas. I was involved with measure number 0070's development, and I'll be recusing myself from evaluation on that one.

MEMBER VIDOVICH: Mladen Vidovich, I am an Associate Professor of Medicine, University of Illinois at Chicago, Chief of Cardiology at Jesse Brown VA, and Governor for the American

College of Cardiology.

MEMBER BRIGGS: Hi, I am Linda Briggs. I am a nurse practitioner, and I'm from George Washington University, and I have nothing to disclose.

MEMBER MARRS: I am Joel Marrs, an associate professor at the University of Colorado School of Pharmacy, and I have nothing to disclose.

MEMBER SPANGLER: I am Jason Spangler. I am Executive Director in U.S. Health Policy and Reimbursement at Amgen. Nothing to disclose.

MEMBER RUGGIERO: I am Nick Ruggiero, Director of Structural Heart Disease at Thomas Jefferson University Hospital in Philadelphia, and I have nothing to disclose.

MEMBER CLEVELAND: Good morning. Joe Cleveland, Professor of Surgery in the Division of Cardiothoracic Surgery at the University of Colorado. I am an adult cardiac surgeon. Nothing to disclose.

MEMBER MARTIN: I am Gerard Martin. I am the Medical Director of Global Service at

Children's National Medical Center here in Washington, D.C. I would note that I am a current member of the American Board of Internal Medicine sub-board Adult Congenital Heart Disease, and that's an exam-writing board, and also a past member of the Board of Trustees at the ACC. I do not think there are any conflicts for today's meeting.

MS. HAMMERSMITH: Okay, thank you. Is Dr. James on the phone? Tom James?

(No audible response.)

MS. HAMMERSMITH: Okay. Thank you all for making those disclosures. I was just told that Dr. James should be on the line.

MEMBER JAMES: Yes.

MS. HAMMERSMITH: Okay. Would you like to disclose?

MEMBER JAMES: The only disclosure that I have is changing from AmeriHealth Caritas to my current role as Chief Medical Officer at Baptist Health Plan in Kentucky.

MS. HAMMERSMITH: Okay, thank you.

MEMBER JAMES: I'll go on mute now.

MS. HAMMERSMITH: Okay, thank you.

Thanks for all of those disclosures, although most of you did not have anything to disclose.

Do you have any questions of me or of each other based on the disclosures?

(No audible response.)

MS. HAMMERSMITH: Okay, and I just want to remind you that during the meeting, if you think you have a conflict, please speak up. If you think someone else may have a conflict or is behaving in a biased manner, please speak up during the meeting.

If you are not comfortable doing that, please approach your Co-Chairs or NQF staff. We don't want you sitting there thinking ooh, I think I have a conflict, but I don't want to interrupt the proceedings, or I think Dr. Jones has a conflict or is behaving in a biased manner, but I'm not comfortable saying something.

In order to make this process work, we

rely on each of you to work with us on conflicts.

So with that, I will let you start your meeting.

MS. VICALE: Thank you very much, Ann.

And before we go any further, I wanted to cover a few housekeeping items and review the project staff.

For those of you in the room today, we welcome you again, Standing Committee, measure developers, and the public. You'll notice the restrooms are available if you exit the main conference area and go past the elevators towards the right.

As you'll see here on the slide, the breaks are noted: 10:45 is our first break, for 15 minutes; 1:00 p.m., we break for lunch; and 3:00 p.m., we break again for 15 minutes.

And for those of you joining us via webcast, you'll notice those slides will be up noting the breaks and the lunch.

For laptops and cell phones, the wifi network you'll notice here again on the slide is

-- username is guest and the password is nqfguest.

We do ask that you mute your cell phones during the meeting, and we do ask that you remain in the room for all of the meeting if that is possible.

One more note is also with the microphones. You --- as you've all done already, you press "Speak," and then you speak directly into the microphone. We just ask that you speak clearly so that that is captured in our transcripts and for anyone joining us on the phone and remotely.

So before we get started, I want to introduce the NQF staff that we have present with us today. I'd like to welcome Helen Burstin, our Chief Scientific Officer; I'd like to welcome Marcia Wilson, our Senior Vice President; Melissa Marinelarena is our Senior Director for the Cardiovascular Project; Ashlie Wilbon, our Managing Director for the Cardiovascular Project; again, I'm Leslie Vicale, the Project Manager, and we have Laura Ibragimova and Donna Herring, our Project Analysts for the Cardiovascular Project.

If you need any assistance, please don't hesitate to ask anyone on the project staff.

Okay. I'd like to begin by talking a little bit about the roles and responsibilities of the Standing Committee. And we do appreciate you all. This is like old hat. You've been here a few times in the past and have been on the Standing Committee through the first two phases of this Cardiovascular Project, and now that we're in Phase 3, many of you are familiar with your role on the Standing Committee.

As you know, the Standing Committee acts as a proxy for the NQF multi-stakeholder membership. Two- to three-year terms are standard. You work with staff to achieve the goals of the project, reviewing the measures, and as you know, the evaluation criteria is followed. For each criterion, you vote on whether the criterion is met and the rationale for the rating.

You make that recommendation to the NQF membership for endorsement, responding to comments submitted during the review period, respond to any

directions from the CSAC, and help to oversee the portfolio of the cardiovascular measures.

And the Standing Committee will oversee the entire portfolio. Some of the responsibilities include providing input on the relevant measurement framework, knowing the measures in the portfolio and the importance of those measures, considering all the issues of measure standardization, parsimony when assessing the portfolio, of course, identifying measure gaps is very important, becoming aware of other NQF measurement activities for the topic areas, and being open to external input on the portfolio, providing feedback on how the portfolio should evolve and considering the portfolio when evaluating individual measures.

I'll turn it over to Melissa Marinelarena for the next two slides.

MS. MARINELARENA: So we just wanted to provide a little bit of information on the Measures Applications Partnership because there was some information on some of the preliminary analysis

that you received about MAP recommendations on some of the measures.

And you may already be aware of this, but -- so just a little background on the MAP. It fulfills a statutory requirement to convene multi-stakeholder groups, and they identify the best available performance measures for use in specific applications. They also provide input to HHS on measures for use in public reporting, performance-based payment, and other programs, and they encourage the alignment of public and private sector performance measures.

And just to note that cardiovascular care is among the MAP's family of measures. It was a framework developed to promote alignment and gap-filling in NQF's priority areas.

And just so you know, the MAP recommendations were just, whether they were recommended or not recommended, it was for specific programs. It wasn't the measure overall, so it's very different from the CDP what we do here.

Next slide. This is just a slide, just

a sort of overall -- sorry -- so this is just overall showing how the CDP standing committees, the MAP work group committees, and the measure developers and measure users are related. We're working on integrating the CDP process and the MAP work groups and the measure developers a little bit more, and part of it is showing CDP what the MAP recommendations were.

A lot of the measures that go through MAP, they are being recommended with conditions that they come through NQF and get endorsed, so we want to show you those recommendations that they've made.

And then also working with the measure developers. Within the CDP process, we're used to have --- we're working closely with the measure developers, so MAP is also working on that on having a more transparent process with measure developers.

So we're working -- getting better at that, so just wanted to provide this background information for you.

Okay, Leslie?

MS. VICALE: Thank you, Melissa.

Now at this time, I'm going to provide a little bit of an overview of the cardiovascular portfolio. This will be brief. You'll notice the measures for the portfolio are listed on the slides, and I'll briefly run through the topic areas to note.

As you can see here, the National Quality Strategy, which all of you are very well-acquainted with, includes priorities of health and well-being, prevention and treatment of leading causes of mortality, person- and family-centered care, patient safety, effective communication and care coordination of affordable care.

The NQF's priority is promoting the most effective prevention and treatment practices for the leading causes of mortality, and this starts with cardiovascular disease.

So now moving on to the cardiovascular portfolio, which currently encompasses 24 endorsed

measures. And as you can see here on the slide, the conditions are listed, primary prevention and screening, coronary artery disease, and ischemic heart disease, acute myocardial infarction, cardiac catheterization, percutaneous coronary intervention, heart failure, rhythm disorders, ICD, cardiac imaging, and cardiac rehabilitation.

So on this slide, you'll notice a visual that shows where in Phase 1, the Cardiovascular Project team went and mapped back the measures to the primary prevention framework to the episodes of care.

And you'll see now we're in Phase 3, the post-acute rehabilitation phase.

Okay. And so, on this slide you'll notice these are the measures in the portfolio for the populations at risk for primary prevention, and you'll see here the asterisk actually notes the measures that are found in other NQF portfolios. You'll also see cardiac imaging measures listed on the slide.

On this slide, the population at risk

for secondary prevention.

And looking here, we have acute phase AMI measures, acute phase outcomes measures.

You all please feel free, if you'd like, to look at the slides more in depth as we are going through them to ensure that we are paying attention to time, but we will just cover the topic areas.

So this slide, you'll notice percutaneous coronary intervention measures.

Well, this one is a little bit smaller. It might be tougher to see. You'll notice the CABG measures, and these are found in the surgery portfolio.

The post-acute rehabilitation phase measures are found on this slide.

And for here, you'll notice the population at risk, the secondary prevention measures.

You'll also notice on the slides there are measures noted for reserve status.

And looking here, this is a similar diagram to what we had just seen before for the

coronary artery disease AMI measures, and these are for the heart failure measures, again mapped back to that primary prevention framework for the episode of care of heart failure.

As you can see here, the population at risk in the measures listed here, evaluation of ongoing management measures for this listed, acute phase and hospitalization, atrial fibrillation and ICD measures, and as you can see here, the NQF-endorsed measures for cardiac catheterization are listed.

On this slide, you'll notice that NQF has one measure that is endorsed for hypertension, and that is 0018, Controlling High Blood Pressure.

Currently, the additional high blood pressure measures are condition-specific.

MS. WILBON: Leslie, I will just pause right there, just add really quickly, the 0018 measure, as you know, is going to be the measure we're reviewing for the ad hoc review today, so it is just -- just to highlight, it is our only hypertension measure in the portfolio that is just

about -- a broad hypertension measure, so it's going to be part of the discussion as we get into the ad hoc review, and just something to keep in mind in terms of thinking about the portfolio as a whole and the implications for the review of that measure, so we'll come back to that in more detail.

MS. VICALE: Thanks, Ashlie.

So you'll see here the NQF-endorsed measures for cardiovascular cost and resource use.

And finally, looking at the gaps in the portfolio, I will discuss these further, this includes care coordination measures, advance care planning, and advance directive measures for patients with heart failure, risk-adjusted and risk-stratified outcomes measures, patient-centered composite measures, new and innovative measures.

And these gaps were identified in the previous Cardiovascular Phase 2 Project.

So now I'll turn it over to our Senior Director, Melissa Marinelarena, and Karen Johnson to go through the measure evaluation process.

MS. MARINELARENA: Thanks, Ashlie.

So I just want to do a quick review of the measure evaluation process.

So this is just where we're at now. We're in the standards review process of -- the standards review step of the CDP. Next slide?

So here is the NQF measure evaluation criteria, which I'm sure you are all very familiar with at this point. The conditions for consideration, importance to mission report, scientific acceptability of measure properties, feasibility, use and usability, and harmonization and selection of best in class.

Evidence is sub-criterion 1(a). This is where we're going to be looking at 40018. For outcome measures, we want a rationale, which often includes the evidence for how the outcome is influenced by health care processes or structures.

And the requirements for evidence is the same for process and intermediate outcome measures, which we do have, I believe, a couple of intermediate outcome measures in this project, and

for that we're looking at the quality, quantity, and consistency of the body of evidence. And for that, it includes empiric studies, which is not expert opinion, and systematic review and grading of evidence.

Here's the algorithm which we have copies, I believe, for everybody as well if we want to walk through this.

So the first thing we look at is does the measure assess performance on health outcome? If it's a health outcome, then we look at the relationship between the measured health outcome and at least one health action, and then we decide whether it passes or not.

If it's not a health outcome, then we look at the intermediate clinical outcome process or structure, and is it based on a systematic review and grading of the body of the empirical evidence? If it is, then you look at the summary of the quantity, quality, and consistency of the body of evidence.

And then from there, is it -- I can't

read it from here -- sorry, let me pull it up here. Okay. Is the systematic review a scientific investigation that focuses on a specific question? If it's not, then -- sorry, I can't read it from here. I can't read it from here.

(Pause.)

MS. MARINELARENA: Okay. Anyway, they are all in front of you. I'm sorry. It's in front of you. Go ahead, next slide.

We go on to reliability and validity. And the page numbers here correspond with the page numbers in the Standing Committee guide, which is on the Standing Committee page.

Reliability is a must-pass for each measure. So is validity. And the measure, to the extent to which the measure as specified -- it has to be as specified, so it must be tested as specified.

Next slide. This is just a diagram of reliability and validity. Again, this diagram is also on page 41.

Measure testing. Empirical analysis to

demonstrate the reliability and validity of the measure as specified, including analysis of issues that pose threats to validity of conclusions of quality of care such as exclusions, we're looking at risk adjustment, stratification for outcome and resource use measures, methods to identify differences in performance, and comparability of data sources and methods.

Reliability testing, we look at the measure score or the data elements, and then you decide whether either of those methods were appropriate for the measure.

And there is an algorithm for this as well. Was the empirical reliability testing conducted using statistical tests with the measure as specified? If it's not, then you decide was empirical validity testing of patient-level data conducted? If it's not, then it's rated as insufficient. If it was, then you use the rating from validity testing of patient-level data elements, which is the next algorithm.

If the answer is no to empirical

reliability testing conducted, they only use descriptive statistics or they describe the process for data management cleaning or computer programming, or the testing does not match the measure specifications, again, that is all known, you go on to validity.

If it does, then you ask was reliability testing conducted with computed performance measure scores for each measured entity? If it's yes, then you move on to step 5, was the method described inappropriate for assessing the proportion of variability due to real differences among measured entities?

So this is where you're looking for signal-to-noise analysis such as the Adams or RAND tutorial or random split half correlation. There are other accepted methods, but these are two more of the common ones.

If those are in the -- part of the analysis, then you move on to yes, and you can either rate it as high or moderate depending on which in box 6 that they fall in.

If it's no -- I can't see the bottom of that -- if it's only one overall score for all patients in a sample used for testing provided data. Sorry, okay, next slide.

Empiric testing, again, you're looking at the measure score or the data element. The measure score assesses the hypothesized relationship of the measure results to some other concepts, assesses the correctness of conclusions about quality, and the data element assesses the correctness of the data elements compared to the gold standard.

There is also phase validity, which is subjective determination by experts that the measure appears to reflect quality of care, and both of these are acceptable.

This is algorithm number three found on page 49 for evaluating validity. The first question we ask, are measure specifications consistent with the evidence provided in support of the measure?

If they're not, then you rate it as low.

If they are, you move on to box two, where you ask were all potential threats to validity that are relevant to the measure empirically assessed? If they're not, it is rated as insufficient.

And then here you're looking at the exclusions, the need for risk adjustment, able to identify statistically significant, and meaningful differences in performance, multiple sets of specifications, and missing data or non-responses.

If they are all assessed, then you're looking at -- you're looking for empirical validity testing conducted using the measure as specified in appropriate statistical tests.

And if that was done, was phase validity systematically assessed by recognized experts to determine agreement on whether the computed performance measure score from measure as specified can be used to distinguish good and poor quality?

If the answer is yes, you move on to box 5 and ask do the results indicate substantial

agreement that the performance measure score from the measure as specified can be used to distinguish quality, and potential threats to validity are not a problem or are adequately addressed to resolve -- so results are not biased?

If the answer is yes, then it is moderate validity, and if it is no, you rate it as low.

If you go back to box 4, and the answer was no, it focused on data element accuracy only, fallibility, feasibility, or other topics, then it's rated as insufficient.

If empirical testing was conducted as specified, then you look at validity testing conducted with -- I can't see that bottom, go to the next slide -- okay.

So threats to validity. You're looking at conceptual threats, unreliability, patients inappropriately excluded from measurement, differences in patient mix for outcome or resource use measures, measure scores that are generated with multiple data sources, and

systematic missing or incorrect data.

Criterion three, feasibility: the extent to which the required data are readily available, retrievable without undue burden, and can be implemented for performance measurement.

Criterion number four, usability: the extent to which the potential audiences are using or could use performance results for both accountability and performance improvement to achieve the goal of high quality efficient health care for individuals and populations.

Related or competing measures: if a measure meets the four criteria and there are endorsed new or related measures or competing measures, the measures are compared to address harmonization and/or selection of the best measure.

We are not going to discuss related or competing measures during the meeting today or tomorrow. We are going to do it -- if we get through all the measures during the meeting, we'll do it on the first call. If not, it's scheduled

for the second call.

So this is probably the most important slide: voting on endorsement criteria. So again, importance to measure and report is must-pass. We're voting on the evidence in the gap sub-criterion.

Scientific acceptability of measure properties is must-pass. Vote on reliability and validity. Feasibility, you vote on feasibility. It is not a must-pass. And usability and use, you're voting on usability and use criteria.

Okay. Then I will pass it back to Leslie.

MS. VICALE: Do I see a question?

MEMBER DELONG: Yeah, I guess I'm a little confused about putting competing measures at the end, and maybe I should have brought this up earlier, but if there is a competing measure that actually covers the territory and is valid, why would we then debate for quite a while about a measure that may not even match the competing measure?

DR. BURSTIN: Let me take that.

So the idea was before we even asked you to compare and contrast, you want to make sure they're even both reasonable, that they've both passed a certain bar. But you could easily do it the other way as well. This has just been our decision, and again, if it -- you're Standing Committees, if you think that doesn't make sense going forward, we're always happy to reconsider that.

But the idea would be you don't want to compare two measures unless you know they actually have both met criteria.

MS. WILBON: I'll just add that sometimes when the Committee goes through and evaluates each measure individually, one of the measures will drop off just on evaluating the measures on its own merits, so then at the end, you don't even have another measure to compare it to anyway.

So if we get to the end and evaluate each measure individually, we see kind of what's left

at the end. Sometimes measures lose endorsement through the process. Sometimes there's other issues that you identify just by applying the criteria. And then we kind of see what's left at the end and then apply the related and competing criteria.

So that's how we ended up with that process.

MEMBER MARTIN: So I have a question for Leslie.

Your slides were very helpful seeing how the measures span the primary prevention, secondary prevention, and hospital care, and I know there's some for transition as well.

Still kind of -- I guess I can't say I'm still a rookie at this since this is the second time, but is there a look at this from NQF as a strategy where you overall say you're after kind of promoting health, this is the number one killer, cardiovascular disease, to look at that breadth and depth of measures to see where there are gaps in the continuum of care from primary prevention to

-- so that there is actually a strategy? Or are we always reacting to what the measure developers propose?

MS. VICALE: I think I'm actually going to let Helen --

DR. BURSTIN: Yeah, so --

MS. VICALE: -- address this.

DR. BURSTIN: So early on in this Committee, and it's probably time to take a look at it, we did give a view across those phases of the continuum of where the measures fit and where the gaps are. I think this is just more of a quick review.

But again, as the standing committees are really getting up and running, this could be what you think you would like to work on off cycle or when you don't have as many measures as well.

So there are clearly gaps in some of the areas, and we'd love to try to fill them or encourage developers to fill them, and I mean you can see in some areas we have a lot of measures, and perhaps we don't need quite that many.

So we should revisit that, and we could bring that back for the Committee.

MS. VICALE: Any other questions before we move on?

(No audible response.)

MS. VICALE: Okay, thank you.

So we're going to take a look at the measure evaluation process improvements briefly.

And this allows the Committee members to focus their time and expertise on other important issues surrounding the measures. And you'll see expanded opportunities for a robust discussion. This ensures the application of the criteria. It's grounded with the evaluation criteria. More consistent and transparent deliberation process. It enhances stakeholder proficiency with the technical evaluation of the measures and provides technical or statistical translation of scientific acceptability criterion.

And what you'll see here you're very familiar with. However, there are some additional

bullet points to note here.

The measure worksheet contains the preliminary analysis. There is an eMeasure technical review for eMeasures only, and we have four eMeasures in this phase of the project.

Something new that you'll see here is a sociodemographic trial review, and as you're all aware, this is the new trial, and the Cardiovascular Phase 3 Project is the first project to be part of this trial.

The Committee's pre-evaluation comments are included in these measure worksheets as well as any public and member comments. This is all part of the measure information form that is submitted by the measure developers, so we combine the preliminary analysis, the public and member comments, and the Committee evaluation comments all together with the original submission from the developer, again with the evidence and testing attachments, any spreadsheets, and any additional documents that the measure developers have submitted for their measures.

And we'll just review really quickly the measure discussion guide. You all in your handouts today have the measure discussion script.

So going through this, again, you're very well aware of this process. The Co-Chairs will introduce the measure, with the title, developer, description, and then the developer will be allowed to provide a brief overview of their measures for two to three minutes.

And we will be mindful of time. If time is, you know, starting to go on, and we need to keep mind of the time, we will have some cards, yellow and red cards, to just notify you if that -- that time is getting a little long.

The NQF staff will provide an overview of the criterion, and then the lead discussants will be allowed to review the input that's relevant to the criterion, again very briefly, one to two minutes.

This summarizes the relevant Committee and NQF staff preliminary reviews that were part of those measure worksheets as well as the relevant

public comments.

The Co-Chairs, the only discussants, will lead the Committee discussion, and then you will vote on the criterion, each criterion, and the staff reviews the voting results. We will read them aloud for the record, and the Co-Chairs will provide a summary of the Committee discussion for the criterion again for the record.

And as you can see here, we have allotted about 20 to 30 minutes per measure. Again, we don't have a white card, however, we do have yellow and red cards just to note whether time should be kept in mind, and if we are starting to get a little long.

As you all are aware, you've been given the voting clickers, and you will rate the criteria via these handheld devices. The criteria will be displayed on the screen with the numbered response option. You'll have 20 seconds to enter your vote, and the votes are anonymous. The results will be displayed on the screen, and we will announce -- the staff will announce the votes, and the

Co-Chairs -- or the Co-Chairs will announce those votes.

Again, achieving consensus, which is what we're all here to do today: quorum for the Committee is 66 percent. The pass or recommended result for the votes is greater than 60 percent, which is 60.1 yes votes for the quorum. This is a sum of the high, moderate, and insufficient with evidence exception, and consensus not reached would be the gray zone, and that's 40 to 60 percent of the yes votes, inclusive of 40 and 60 percent of quorum.

The do not pass, not recommended vote would be less than 40 percent of the yes votes of a quorum.

MS. WILBON: I will just add that staff is going to be, you know, keeping track of -- of people -- of the votes.

We just ask that if we're getting to a point in the discussion where we're getting ready to vote that folks try not to go out and take calls, and you know, we're going to -- we need to make sure

that we have a quorum for all the votes, which means we need at least 12 people to vote on every measure.

So, you know, we don't want to have to chase you down, but if you can, just kind of keep track of what's going on in the meeting. If you have to step out, we completely understand, you know, to do -- go to the bathroom or take a call, but if you can, try to make sure that you're here for the vote so that we can make sure we can keep moving and make sure we meet quorum for all the measures, that would be great. Thank you.

MS. VICALE: And on this slide, you'll notice just some ground rules for today's meeting. So we just ask that you be prepared having reviewed the measures beforehand, which I am sure you all have done, I know you all have done.

And these, the evaluation recommendations of the measure, on the measure evaluation criteria and guidance, please remain engaged in the discussion without distractions. Please minimize any sidebar conversations.

We ask that you do stay in the room to

attend the meeting at all times except during the breaks. Again, keep your comments concise and focused. We do have a lot of measures to review. We have one ad hoc discussion as well as 23 measures, so we do have quite a bit in the agenda.

And we ask that you avoid dominating a discussion, and please do be considerate of your fellow Committee members and allowing others to contribute to the conversation.

Indicate agreement without repeating what has already been said. Again, that is, you know, to be mindful of the time constraints that we do have for the agenda.

I'll turn it over to Melissa and Ashlie.

MS. MARINELARENA: Thanks, Leslie.

So just a quick overview of what happened with 0018. As you know, we had a call a couple of weeks ago. We did not reach quorum. We had 10 votes. We needed a minimum of 12.

And these were the results: for the do the changes made to the measure 0018 meet NQF evidence criteria, including the quality,

quantity, and consistency of the evidence, we got 40 percent yes, 60 percent no.

Does the Committee recommend the revised measure for continued endorsement? 60 percent yes, 40 percent no.

So we wanted to have a discussion again. So Mary and George are going to be the lead discussants, and because we didn't have a quorum, we are going to vote again on the evidence, and we are going to vote on the measure --

MS. WILBON: We will be very clear when we get ready to vote, but as you can see from the results, they don't quite align, right?

So you would expect that if most people felt that the evidence did not -- was not sufficient, that the measure -- the revised measure would not be recommended for endorsement.

So we just want to be very clear when we vote that when we vote on the -- when we vote on the -- the measure recommendation, that we're voting on continued endorsement of the revised measure, not the old measure.

So as we get to the discussion, we'll just -- we'll -- staff will just try to be very clear about what we're voting on, so versus the -- you know, keeping the old measure versus continued endorsement for the new measure, and we'll talk a little bit about the implications for that obviously with the developers here, thank you for joining us.

So we'll just, you know, keep that in mind. When we get to voting, we'll talk through that and make sure everyone understands exactly what we're voting on to make sure there is consistency and a good full discussion on what the issues are, so thank you.

CO-CHAIR GEORGE: So before we start, I'll just remind you, if you want to speak, if you can just turn your name card up on its end, it will help us try to keep things in order.

And I don't think I really need to introduce this measure again. I think we'll move on to brief comments from the developers.

MS. BARTON: Hello. I'm Mary Barton,

Vice President for Performance Measurement at NCQA, and my colleague Dan Roman is going to introduce our measure.

MR. ROMAN: Sure. So I'm Dan Roman. This is our controlling high blood pressure measure.

Just a little bit of history on the measure: this measure was developed in 1999, and it was first endorsed in 2009. Aside from its use in NCQA's accreditation and recognition programs, the measure is used in Medicare Advantage star ratings, PQRS, and meaningful use.

The measure, as you can see, focuses on patients with hypertension age 18 to 85 and assesses whether or not their blood pressure is adequately controlled.

In 2014, we updated the definition of adequate control to align with the recommendations from the 2014 Evidence-Based Guideline for the Management of High Blood Pressure in Adults, a report from the panel members appointed to the Eighth Joint National Committee.

The change included relaxing the blood pressure threshold for patients 60 years and older from less than 140/90 to less than 150/90 in the general population, and because we relaxed that threshold, we also needed to specify that the blood pressure target for all patients with diabetes is less than 140/90 because we didn't want to have anyone who has diabetes 60 years and older be treated. Previously, there was not any specification to call out that diabetics should be treated to that different goal.

The change was discussed with our Cardiovascular Measurement Advisory Panel, our Diabetes Measurement Advisory Panel, and our Geriatrics Panel, which I would like to note in particular was very supportive of this change.

Ultimately, we took this change to our committee for performance measurement. All the panels approved this. We took it to public comment. The majority of our comments were also in favor of making the change and aligning with the JNC recommendations.

We submitted that change during our 2014 update cycle, and I believe that is what prompted this ad hoc review. Thank you.

CO-CHAIR GEORGE: Thank you.

And I'm going to just review the comments that I had prepared for the call that we had a couple of weeks ago, and then George will give his review as well.

In reviewing the evidence for this measure, it was a little bit tricky because there are now not just one idea that we're voting on, there's actually three different things incorporated in here.

And the evidence for -- in the review for the 18 to 59 year olds, and for the older diabetics, was based on expert opinion. The evidence for the 60 to 85 year olds was based on a systematic review of six clinical trials, and those six trials, two were specifically in patients 80 and older, and another one 70 to 85.

One was a Japanese study which may not be representative of the U.S. population, and a

couple of them were mentioned as possibly being underpowered.

The developers did provide their own QQC, which was the -- basically the systematic review from the panel.

And I think it really is -- is very tricky in trying to say how are we going to do this when two of the recommendations are based on expert opinion, which would lead us to insufficient with exception, and yet there is a systematic review, empirical review for the other.

I will say, and I'll just be very brief, that the review did not look at or consider any of the harms, and this was recognized in the developer's submission, but they didn't look at any of the harms that might result from the changes that are being proposed.

So I really -- I really think it's up to each Committee member to decide how they evaluate that evidence.

George?

MEMBER PHILIPPIDES: I agree. Those

are the same issues that I came across.

For those who haven't looked at this recently, I think on page 7 is a very brief look at recommendations 1, 2 I'll guess, 3, and 5, and those are the four recommendations out of the 9 that JNC 8 came up with upon which this particular measurement is based.

And I won't get into this -- look into too many weeds here, but for the over the age 60 population, they give the recommendation a fairly high grade because as you mentioned, there are some RCTs that sort of get at this particular question.

But then there is a corollary recommendation, as you mentioned, saying that if you're over the age of 60, and you're down below 140/90 after treatment, that that is okay. There is no sort of hard evidence to say that it's okay. They didn't look into the harms of it. But I guess they didn't want to sort of undo what had been done before or ding people, I guess, who had previously gotten their over-60-year-old patients to that threshold.

Again, whether there is harm in going down that low, it's never really explicitly stated unless I am mistaken, Mary.

Then for the next recommendation, they start talking about the younger than 60 age group, and that's even more difficult because my understanding from looking at this a couple times is there aren't great RCTs focused specifically on systolic hypertension targets and thresholds for treatment. There is some evidence as far as diastolic treatment.

And then it gets even more difficult when you get down into the younger of the young, the 18 to 29 year olds, where there is really nothing that I could find, and therefore it sort of rises only to the threshold of expert opinion, again relying on a group of people in a room coming up with their sort of best guess on this.

So as you go through the four recommendations, you can see, as you mentioned, there is a spectrum of RCTs that are focused on the recommendations, and then a combination of expert

opinion.

I think based on the good faith effort to look at the trials that have been done, take sub-groups and see well there was no harm done here, there was a trend toward benefit here, but nothing that's focused, certainly not data that's as rigorous as what we usually uphold here.

Having said that, I think -- I don't want to give my bias. Let me stop there.

So I think if people look at page 7 and take a look at the four recommendations and what the evidence is, it's sort of a good starting point for our discussion.

CO-CHAIR GEORGE: I think that having to come up with your best guess is -- because we do have some gaps in the RCTs, there's one SPRINT trial that is ongoing that will look at this in a 55 and older population, looking at specific blood pressure goals, but we don't have the results from that yet.

MEMBER PHILIPPIDES: I thought it was interesting to find out, I don't think that I knew

this, that, correct me if I'm wrong, there is no one specific RCT focus looking at diabetic thresholds in any kind of rigorous way. We've gotten to this through lots of different trials sort of cobbled together, am I correct? I hadn't known that.

CO-CHAIR GEORGE: Any discussion?

(No audible response.)

CO-CHAIR GEORGE: If there's no discussion, then we'll move on to public comment, is that right?

MS. WILBON: Maybe if anyone on the Committee would be comfortable sharing what prompted them to vote maybe the way they did the first time, just to kind of maybe stimulate a little discussion and help some others maybe consider some other viewpoints on the issue.

MEMBER DELONG: I am sorry, I should probably know this, but are we evaluating whether the change should be made? Okay. Not whether the measure itself originally -- well, we're voting on whether the original measure should stay versus the

change? No. What are we voting on?

MS. WILBON: So right now, we're just discussing the evidence, so whether or not based on our criteria you feel like the evidence is sufficient.

So based on the new evidence that they have submitted, do you believe that it is sufficient to support the measure concept and what they're choosing to measure?

DR. BURSTIN: And just to make it maybe a bit more crystal clear, an ad hoc review means there is one element that has been changed that needs to be evaluated.

In this instance, a measure change was made based on evidence, so in an ad hoc review, we only ask you to evaluate the one criterion that is affected that led to the change in the measure.

So in this instance, we're only asking you to consider the quality, quantity, and consistency of evidence as you would for any measure, really all three of those, to consider whether the evidence that supports the update that

NCQA has made to the measure is sufficient, and then as a second point, we will ask you to decide whether you would like to uphold endorsement of the modified measure.

So happy to answer questions. It's a little contorted.

CO-CHAIR GEORGE: Linda?

MEMBER BRIGGS: Hi. I guess my consternation over this particular measure had to do with the interpretation of whether it was an outcome measure, or this, as it talks about, it's an intermediate outcome measure.

Because if you look at our guidance in terms of the clinical evidence, if we consider it to be an outcome measure, then we're not looking at the standard of evidence the way we would a process measure.

So we basically go past and go to yes, and in that case, if we consider this to be an outcome measure. At least that's my interpretation.

DR. BURSTIN: We have always

considered this an intermediate outcome measure. It's always been classified as such. We're not changing that classification. And intermediate outcome measures, like process measures, require quantity, quality, and consistency of evidence assessment.

CO-CHAIR GEORGE: Other comments?

MEMBER PHILIPPIDES: This is just another question: if we don't pass this measure, what stays in that vacuum? The old measure, or there's no hypertension measure?

DR. BURSTIN: NCQA has made this change to the measure, so there would not be another hypertension measure available at this time. Yeah, it's kind of a tough situation.

Well, this measure as changed. We do have some disease-specific hypertension measures. You're going to see some of them, I think, shortly. But nothing -- no broad overall blood pressure measure, controlled measure.

CO-CHAIR GEORGE: Leslie?

MEMBER CHO: So I voted yes for both.

I thought that this measure was actually a good measure because if you look at the previous measures, previously what was passed -- you know, let me just back up.

This is based on the new guidelines for hypertension, which has been very controversial, as everyone agrees in this room.

But the -- if you look at actually the guidelines and the prior guidelines, I mean, this is a rigorous assessment of blood pressure trials that are currently out there.

And if we -- I feel like, yeah, there is some room, we don't know about certain age groups, but that is just a gap in the evidence, but the totality of what is out there currently I think supports these current blood pressure guidelines.

And so that is why I thought, instead of having no measure, I thought having this particular measure with this blood pressure guideline is beneficial.

CO-CHAIR GEORGE: Joe?

MEMBER CLEVELAND: I guess I'm hearing

the discussion too. I am a little struck that kind of drifting up at 30,000 feet and kind of looking at the portfolio measures, it would be kind of silly for us to be talking about very detailed, you know, are people getting medicines after tertiary interventions and not have a blood pressure measurement.

I just think that to Leslie's point, it may not be absolutely stellar in terms of getting in the weeds and seeing where the evidence is even though some of it is controversial, but I guess if we're having a blood pressure medicine in the portfolio, it's reasonable I think that one could look at the entire totality, you know, the entire portfolio package as an argument for this.

CO-CHAIR GEORGE: I will just remind the Committee that we are expecting new guidelines again, the AHA/ACC guidelines, which are due out in 2016. Who knows what they will say?

Sana?

MEMBER AL-KHATIB: I completely agree with the point that you said about the upcoming

guidelines. In fact, I voted against this measure because I really would like to see what the guidelines will say, and although I actually raised that question during the call when we had the call, I asked why not wait for another year until we -- because it's going to be really confusing.

You know, let's say we support this, and then the guidelines come out with something totally different delaying it by a year. Is that going to like have major adverse outcomes? I don't think so.

MEMBER PHILIPPIDES: Just to play devil's advocate, I don't know that this is the case, but it feels like there is always a guideline that's one or two years away down the road, and sometimes when we wait for the next set of guidelines, we just wait indefinitely.

I don't know that to be the case, but that is my concern.

CO-CHAIR GEORGE: Ellen? Sorry, Leslie.

MEMBER HILLEGASS: What I'd like to say

is that I did vote no originally, and I have actually been swayed by looking at the fact that there -- when we saw the overview of all the different measures, and I see that this is the only blood pressure measure, and part of me says okay, the evidence isn't that strong with expert opinion, but does it do harm?

And so we don't have the strongest evidence on a couple of these gappy age groups, but I don't think that this measure does harm for the lack of evidence that we have in some of those age groups, and if this is our only measure, I too do not want to wait for other guidelines. I think maybe this is something that we should consider.

CO-CHAIR GEORGE: Leslie?

MEMBER CHO: Based on how ACC/AHA guidelines have been recently regarding the lipids guideline even, they have basically focused on large randomized control studies and have ignored registry or epidemiological things.

Just, you know, if you look at the guidelines, the lipid guidelines, you know, it's

all based on randomized control study.

Unless something dramatic happens in the next year at the AHA or this year at AHA about blood pressure guidelines, it's -- my guess is that it will be very similar to what we have here.

So I don't think we should have no blood pressure guideline in the NQF portfolio for a whole year. I just don't think that's good practice.

CO-CHAIR GEORGE: Joel?

MEMBER MARRS: I guess my overall concern was that -- and more of a question to NCQA -- was why was diabetes isolated out versus the reference to JNC 8 also isolated out CKD, and that didn't necessarily funnel into this?

MR. ROMAN: So we considered both. So as I kind of said in the beginning, diabetes was culled out specifically just because we changed -- we added in this -- the threshold for the adults 60 years and older, so now we have less than 140/90 for 59 and younger, and less than 150/90 for 60 and older for the general population.

Well, the recommendation for diabetics

is less than 140/90 for everyone. So we had to add that in.

We considered the chronic kidney disease. It was culled out.

The measure uses administrative claims to identify the patients for the denominator. The definition that is described for CKD in the JNC 8 is not something that you can capture in claims. It's not just a code for kidney disease that would qualify. It's kidney disease with certain lab values that are in certain ranges, and it was too complicated.

And so we did discuss it with our expert panels, and they said that it was not appropriate to include in the measure at this time. It's just not defined well enough.

CO-CHAIR GEORGE: Any other final comments?

(No audible response.)

CO-CHAIR GEORGE: Right. Should we move on to the public comments?

MS. WILBON: Is there anyone in the

room who would like to comment? You can form a line at the microphone.

MEMBER JAMES: How about phone persons?

MS. WILBON: Oh, we were going to see if anyone in the room had a comment first, but --

MEMBER JAMES: Thank you.

MS. WILBON: -- those --

MEMBER JAMES: This is Tom James --

MS. WILBON: -- rising to the occasion, so -- oh, okay, so I think we'll open it up to those on the phone. Oh, okay, sorry Tom.

MEMBER JAMES: Okay.

I agree with a lot of the comments that were made there. I just think that -- and speaking as a primary care physician as well as somebody in the insurance industry, we've got a big gap in care that goes on relative to hypertension. NCQA has been making a significant effort in the -- in bringing this forward, and physicians and other health care providers tend to respond when there is a measure, so I think that the gap in care issue

has to be considered rather than a lot of the nuances.

I am in favor of this.

MS. VICALE: Okay, so at this time, we will move to public comment.

However, we are a little early. So are there any other comments from anyone in the room?

(No audible response.)

MS. VICALE: Okay. No public comments.

(Pause.)

MS. VICALE: Okay. There are no public comments from the room, so we'll ask the operator to open the lines for public comment, and we just ask the operator to keep those lines open just a few minutes longer since we are running a little early today.

THE OPERATOR: Okay. And at this time, if you would like to make a public comment, please press star, then the number 1.

(Pause.)

DR. BURSTIN: Operator, could we

please see if Dr. Sanchez is on the line?

DR. SANCHEZ: I am on the line.

DR. BURSTIN: Oh, hi, Eduardo, go ahead.

DR. SANCHEZ: Hi. My name is Eduardo Sanchez. I serve as the Chief Medical Officer for Prevention at the American Heart Association.

I am a family physician by trade, have done primary care, so the comment about gaps in care resonates. But perhaps I come to a different conclusion, and the AHA comes to a different conclusion.

We support keeping the definition of adequate blood pressure control as blood pressure below 140/90 millimeters of mercury.

The AHA feels that the 2013 JAMA recommendations are not JNC 8 and should not be considered a guideline. They were -- while they were initiated by NHLBI initially and commissioned by NHLBI initially, the recommendations have neither been sanctioned nor endorsed by NHLBI.

They have not been endorsed by the AHA,

the ACC, or a number of other international organizations whose endorsement was sought.

It was authored by a subset of the panel members that were initially appointed to the at that time eighth JNC, and although there was near-unanimous agreement on almost all the recommendations, the panel didn't reach unanimous consensus on the recommendation to raise the target blood pressure for those older than 60 years who didn't have diabetes or chronic kidney disease.

And as already mentioned, the ACC/AHA hypertension guideline is under development, expected to be released next summer, and I will say that the Guideline Committee is comprised of a more diverse membership than perhaps past ACC/AHA Guideline Committees, including the American College of Preventative Medicine.

AHA also feels that increasing the target blood pressure to less than 150 for the over 60 population would adversely affect public health and could increase health disparities and undermine decades of progress on reducing CVD and

stroke.

This is where the gaps in care issue resonates for me. Despite the JNC goals and national performance measures over the past decade or so, still only half of patients with hypertension in the United States have their blood pressure controlled as defined today by the AHA.

We worry that this change will reduce the intensity of any hypertensive treatment in a large population at risk for cardiovascular disease and worry that a higher target would apply to some of the groups at highest cardiovascular risk, such as African Americans, hypertensive patients with multiple CVD risk factors other than diabetes or chronic kidney disease, and those with clinical CVD, potentially worsening disparities in care and outcomes.

The AHA believes that the evidence for a change in target blood pressure is insufficient. Recommending less aggressive targets in this or other high-risk populations requires stronger justification in our opinion than the paper cites.

Evidence cited for the higher target blood pressure is also inconsistent with the evidence cited for recommending a systolic blood pressure of less than 140 in those younger than 60 and those 60 or older with diabetes or chronic kidney disease.

There is little RCT evidence of risk or benefit in treating persons younger than 60 to this target except in those with diastolic hypertension.

And then last thing, regarding the evidence, the best RCT evidence available for an SBT target of less than 140 is in those younger -- older than 60 years, older than 60.

And as it relates to others, three recent guidelines from other countries and other places, Europe, UK, and Canada, that reviewed similar evidence, concluded that the appropriate cut point for age-related systolic blood pressure is 80 years or older. The systolic hypertension in the elderly program trial showed benefit of treating hypertension to a systolic blood pressure goal between 140 and 145.

Japanese trials in older individuals cited to support the higher systolic blood pressure were underpowered, and we agree that a systolic blood pressure goal of less than 150 for frail persons aged 80 or older is a reasonable alternate approach to address concerns that elderly persons are at higher risk for adverse events.

In summary, the ACC/AHA strongly disagrees with the recommendation to change the current -- current control of blood pressure measure. The systolic blood pressure target of 140 milligrams of Mercury or lower in all those age 60 or older should be maintained until there is greater certainty of the risks and benefits of a higher target, and we urge NQF to defer approval of the revised NCQA BP control measure pending release of the updated ACC/AHA hypertension guideline.

Thank you very much.

MS. VICALE: Thank you. Are there any other public comments on the line?

THE OPERATOR: We have a public comment

from Janet Wright.

DR. WRIGHT: Hey, good morning, everyone. Janet Wright, I am Executive Director of Million Hearts, co-led by CDC and CMS.

We submitted some comments via the -- from CMS about this, but I am also carrying forward opinions from CDC.

We agree with the comments that Eduardo made. We are concerned (a) about the controversy that still exists -- who knew at this point in time that there should be such a dustup over hypertension targets? But we feel there is still residual controversy and no agreement among experts in the field.

We are concerned about the harm, particularly for people of color, for women, as stated in the minority report, again, as Eduardo mentioned.

And we are concerned that there is not sufficient evidence of harm of the lower target for older individuals.

But the third point that I would like

to make is that we know that practitioners make decisions based on their own experience and individual characteristics of a patient, so that whatever target we all decide is right, we -- we expect and depend on health care professionals to fine tune based on patient tolerance, patient preference, and their own clinical judgment.

Giving them guidance through the interpretation of the evidence is the favor, the gift that we do for them. We understand there is enormous unrest or -- or controversy around the interpretation of the body of evidence. We recognize that the group convening to write JNC 8 was constrained in a way by being required to only use randomized control trials as the basis for their decision-making, and we know that they were able to issue a few recommendations on important questions, not a comprehensive approach to hypertension management.

We look forward to that comprehensive national guideline from the multi-stakeholder group in the summer and know that that will turn

into performance measures which will then be brought to NQF.

So in -- in summary, we would prefer maintaining the target of less than 140/90 as it has been used for so long and has been associated with a decline in morbidity and mortality from cardiovascular disease.

We look forward to those comprehensive national guidelines which may necessitate a retooling of the performance measure that is currently embedded across federal and private programs.

Just to remind, this current less than 140 over 90 is in PQRS, is in the EHR Incentive Programs or meaningful use, it's in the Medicare Shared Savings Program, Pioneer ACO, in CMMI's Comprehensive Primary Care Initiative, and also in the new Cardiovascular Risk Reduction Model out of the Innovation Center.

So for all these reasons, CMS and CDC are in favor of maintaining the current specifications of NQF 0018.

Thank you.

MS. VICALE: Are there any other public commenters on the line?

THE OPERATOR: At this time, there are no public comments.

MS. VICALE: Thank you.

MEMBER DELONG: Can I ask a question?

You said that this is a measure that has been used, but it is based on claims data, right? So how does somebody that does not support the use of claims data for a measure vote on something like this?

I mean, the -- the change is not my issue, so I don't know how I feel about the change.

DR. BURSTIN: You have to just vote on what's before you today. It's not a question of it being claims, it's a question of really the evidence. You've heard the discussion of the evidence. You have to vote just based on your sense of the, again, not to be a broken record, but this is what we do, it's the quality, quantity, and consistency of the evidence.

MR. ROMAN: Can I just clarify though, too?

So it is based on claims for the denominator, so that's how we identify patients initially. The -- the numerator for it does requires medical record review, where we actually look to see.

I just want to make sure we're clear that it's a hybrid measure.

CO-CHAIR GEORGE: Kristi?

MEMBER MITCHELL: Just a point of clarification.

So if we vote against the evidence, just hypothetically, would the measure revert to its old 140/90, or will it be pulled, or lose its current endorsement? I am sorry, I am just making sure I know what I'm voting on.

CO-CHAIR GEORGE: We will vote on the evidence, and then we will vote on the measure.

MEMBER MITCHELL: So that's not my -- my full question, but we're going to vote on the evidence, I get that. What I need to understand

is, hypothetically, if we vote against the evidence, what happens to the measure 0018? Will there no longer be a hypertension measure in the CV portfolio?

DR. BURSTIN: Yeah, I mean I think the key thing is you have to focus on the question at hand. I mean I understand the externalities here of the question of it being in the portfolio, but for now, focus on the evidence.

And I guess that's really a question for NCQA. I mean, at this point, they have modified their measure, so the existing 0018 from your perspective is not something you're maintaining, correct? Right.

CO-CHAIR GEORGE: It would be really up to the developers whether they wanted to come back with a different measure.

MEMBER CHO: So just a clarification: this is for patients without CAD blood pressure goal. Because you have other measures for CAD patients, correct?

CO-CHAIR GEORGE: This measure, for

the 60 to 85, is everybody that's 60 --

MEMBER CHO: Regardless of --

CO-CHAIR GEORGE: Regardless.

MEMBER CHO: -- whether they have CAD or not?

CO-CHAIR GEORGE: Right.

MEMBER CHO: Okay.

CO-CHAIR GEORGE: The exclusions are only pregnancy and end stage renal disease.

MEMBER CHO: Okay, because just so that everybody is clear, the recent publication by the ACC/AHA guidelines on patients with CAD and hypertension, they quote the evidence for CAD patients blood pressure less than 130/90 as two B C, which is pretty low, right?

So two B C. And so we're talking about evidence, and we're talking about randomized control setting, we're talking about that this measure is, you know, not good, and whatnot, but the guidelines that the ACC/AHA recently revised, whatever they put out on March of 2015, even for their CAD patients, the evidence was two B C, which

is pretty low.

Yes, 130/80, yes. I know, but everybody is talking about the new upcoming guidelines, how the new upcoming guidelines will be better, will be more granular, and what -- but actually, you know --

CO-CHAIR GEORGE: Sana?

MEMBER AL-KHATIB: So I mean I completely agree that we do need a performance measure on hypertension, on blood pressure control, but I also agree that we need to be thoughtful in how we do that.

You know, I certainly agree with all the comments that remained about the lack of evidence, so we need more evidence.

And I -- my concern is if we are willing to be more lenient, if you will, or demanding less of an evidence for this measure, and our approach that has been I think consistent across all the other measures that we have been demanding more evidence, I think we won't be consistent if we were to allow this measure to pass.

And -- and I agree that, you know, maybe we only have expert opinion to rely on, but we all heard, and I'm actually aware that they are currently working on that performance measure guideline document, so hopefully that won't be in the, you know, too far future.

I think we'll hopefully get a lot of information from experts in the field, as was stated on the line, that there is very good representation of different entities who are, you know, working on this document, so if we're going to be relying on expert opinion, let's make sure that we are relying on the best expert opinion that exists.

MS. VICALE: We have a public comment in the room.

DR. OFILI: Thank you so much.

Good morning. My name is Elizabeth Ofili. I am a cardiologist from Morehouse School of Medicine.

And I just thought I would offer a comment from my vantage point. I see many African

American patients with high blood pressure, and a lot of the times, that's the main reason they come in to see me even though I'm a cardiologist.

And I know for us in the community, we have been and remain concerned about the most recent JNC guidelines because of the reasons that have been shared in the room in terms of, you know, potentially lowering the bar of intensity of therapy in individuals who are still high risk even though they don't have either CAD or renal disease.

So I just wanted to give a comment in agreement with Dr. Sanchez based on some of what we're hearing from the field, and also to add that even as we speak, there are trials now that are looking at comparison of different levels of blood pressure, and I think there will be that evidence as well to help guide this discussion.

Thank you.

CO-CHAIR GEORGE: Last-minute comments before we vote?

I'll remind you the first vote is on the evidence only.

MS. IBRAGIMOVA: Okay. Just some quick tips before we get into the voting.

In the beginning of the meeting, you all received clickers, and you have the options to vote on the slides, so you can click 1 for yes and 2 for no.

And yes, you would point directly to me because the fob is here.

So the question is do the changes made to measure 0018, Controlling High Blood Pressure, meet the NQF evidence criterion, including the quality, quantity, and consistency of evidence? 1 yes, 2 no.

(Pause.)

PARTICIPANT: So it looks like it's only captured -- how many people do we have in the room?

MS. IBRAGIMOVA: There should be 17 votes, including Tom over the phone.

PARTICIPANT: Okay. We have 14.

MS. IBRAGIMOVA: We can try a re-vote. You can -- it will capture your second -- yeah. Oh,

Tom, okay.

(Pause.)

MS. IBRAGIMOVA: The results are 25 percent yes, 75 percent no.

CO-CHAIR GEORGE: All right. We will go ahead and vote on endorsement of the measure.

MS. IBRAGIMOVA: So the question is does the Committee recommend this revised measure for continued endorsement? 1 yes, 2 no.

MS. WILBON: And to clarify, this is on the revised measure, not the old measure. We're voting on whether or not you want to continue endorsement for the new measure with the updated guidelines that they submitted and discussed today, okay?

(Pause.)

MS. IBRAGIMOVA: Tom, would you mind submitting your vote?

MEMBER JAMES: I just sent it, yes, an email. I am just getting my computer up so I can put things in, but I sent it by email.

MS. IBRAGIMOVA: Thanks, Tom, we

received it.

MEMBER JAMES: Okay.

MS. IBRAGIMOVA: So the results are 31 percent yes, 69 percent no.

DR. BURSTIN: Just to -- just like all your other measures, this will go out for public comment, so as part of the report that goes out, you'll have one more opportunity obviously following public comment for having further discussion. Thank you.

CO-CHAIR GEORGE: I think we are now at a break.

MS. VICALE: We'll have a 15 minute break, and we'll return to the room at 10:45. Thank you.

(Whereupon, the meeting went off the record at 10:30 a.m. and resumed at 10:45 a.m.)

CO-CHAIR GEORGE: So, we're going to go ahead with the rest of our packed agenda for this morning starting with Number 0068. Discussants are Ellen and Jason, I believe. And we'll ask the measure developers to give us a brief introduction.

MR. ROMAN: Hello. This is our Ischemic Vascular Disease use of aspirin or another antiplatelet measure.

The measure was first developed in 1999, and first endorsed in 2009. It is used in NCQA's Heart, Stroke Recognition Program. It's also in PQRS and in Meaningful Use. The measure focuses on patients with established cardiovascular disease age 18-years and older, and assesses whether or not they are using aspirin or another antiplatelet.

The changes, so I think you'll note that the title for this measure and anywhere you see antithrombotic formerly has been changed to antiplatelet, so again we have removed antithrombotic and replaced it with antiplatelet. This was to better align with the true intent of the measure and to decrease confusion. We received questions kind of about anticoagulants because we had antithrombotic in the measure, even though the measures only ever included antiplatelet or aspirin.

The other change that we implemented for this submission to NQF was to add an exclusion for anticoagulant therapy, and this was done at the recommendation of our Cardiovascular Measurement Advisory Panel. Basically, recognizing the complexity of the situation where anticoagulant therapy might be indicated, so the patients for whom that is indicated are removed from the population for this measure. It's not to say that there's not a case for it; it's just for this measure we are focusing on antiplatelet therapy.

I think those are all the changes for the measure. Yes, that's all. Thank you.

MEMBER SPANGLER: I can go first. I was going to lead this one, and Ellen was going to lead the next one.

So, Dan described the measure already so I was just going to go into the evidence. The developers did provide a QQC, and from four separate guidelines, multiple guidelines, statements there was, I think, high quality from the guidelines and systematic review, as well. The

evidence does apply directly to the process and to patient outcomes, so I gave this a high rating for evidence.

I don't know if there's anything, Ellen, you wanted to add.

MEMBER HILLEGAS: And I'd like to include that part of the other reason why we thought this was high evidence is the number in the systematic review included 287 studies, and 135,000 patients, so it was pretty intensive and it was all 1a.

CO-CHAIR GEORGE: We'll move on to discussion of the evidence. Comments on the evidence? If not, we'll move on to voting on the evidence.

MEMBER PHILIPPIDES: One quick question. This might not be the right time. Was there a performance gap showed for this particular metric? Is that part of this vote, whether there's a performance gap?

MEMBER SPANGLER: I think that's the next discussion topic. Right?

CO-CHAIR GEORGE: Yes, the next --- we'll vote on the evidence, and then we vote on the opportunity for improvement.

MEMBER PHILIPPIDES: I'll ask then. Thank you.

MS. IBRAGIMOVA: No, we're not loading it. Just one second. Okay. So, importance to measure and report, 1a evidence structure, process, intermediate or outcome. One, high; two, moderate; three, low; four, insufficient. You can begin voting. Tom, can you please submit your vote via the chat?

MEMBER JAMES: Will do.

MS. IBRAGIMOVA: Thank you. We're still missing two more votes. We can try a revote.

MS. VICALE: Tom is ---

MS. IBRAGIMOVA: Okay. So, we're missing one more. Tom, we still haven't received your vote via chat.

MEMBER JAMES: I sent it by email. I'm going to have to redo some things on this computer to let it function. It's too slow.

MS. VICALE: Tom, have you sent the email?

MEMBER JAMES: Yes, I did. I can resend it.

MS. IBRAGIMOVA: Okay. So, the results are 93 percent high, 7 percent moderate, zero percent low, zero percent insufficient.

CO-CHAIR GEORGE: We will move on to the opportunity for improvement and any data on disparities.

MEMBER SPANGLER: So, the developers presented information from their own Heart Stroke Recognition Program, as well as PQRS, and there does show some performance gap. One thing that's really interesting is there's a drop-off in performance after a certain year period. And I have their data in explanation for that. It also corresponds that as volume goes up, the numbers they look at, the performance gap increases, the number of people are adhering to the measure goes down. So, I don't know if there's an explanation for that, but that's the only thing that's a little

strange. And going on the more recent numbers, there's definitely a performance gap. And I'll let the developers, if they want to address that.

But, also, as we talk about disparities, they didn't have anything specific around disparities, but they did note a couple of studies that showed some racial and socioeconomic disparity.

MR. ROMAN: So, with regard to the performance from NCQA's Heart Stroke Recognition Program, the volume increase, I think as we have more clinicians joining the program, we're going to see a change. The program is self-report, so there are some --- I think some nuances to the program that kind of make it unclear sometimes why there are changes. We do think there's opportunity for improvement, though, shown. But it is --- I think there's --- part of it is the program is clinician will select 30 patients to report their data on, so there's some self-selection bias, which I think skews some of the results sometimes.

MEMBER SPANGLER: Dan, is there

anything -- I mean, have you guys talked to CMS at all about the drop from 2011 to 2012?

MS. BARTON: No, I think the PQRS data that you're referring to is really emblematic of the PQRS Program which started out as a relatively small program, entirely voluntary, and as CMS has continued to signal that it will be less voluntary in the future, and there is a great swell of people who are reporting it, not surprising that that's going to include some people who don't see this as a priority, or who don't really have systems in their practice to implement these kind of things. The true believers signed up first, I think is the way it looks.

MEMBER SPANGLER: Makes sense. Thanks, Mary.

MEMBER HILLEGAS: And I wanted to include that it was also reported that there are some specific diseases that may not be receiving appropriate care, and they mentioned PAD. So, another disparity in care, not just socioeconomic and black versus white, or racial.

CO-CHAIR GEORGE: Discussion on the opportunities for improvement?

I had a question for the developers about the drop-off in performance, and whether that might be related to change in possibly the exclusions; because during the time period since the last time this was looked at, we had the introduction of the Anti-Xa drugs for stroke prevention, in particular. So, there's a whole population of people there that might have been on aspirin before, and now are on something else, and so that there might have been something there, but I was just wondering if you saw the change related to the Xas? No?

MR. ROMAN: The change --- the addition of the --- sorry. The addition of the exclusion anticoagulants has not been included in the PQRS version of the measure; that we just made this change in time for the submission for NQF, so the change would not have any effect on the performance rates for PQRS.

MEMBER BRIGGS: So, there might be

people that were counted in the denominator who were on Anti-Xas. Is that, basically, what you're saying?

MS. BARTON: I'm not entirely familiar with the universe of how providers report PQRS measures, so it would be speculation for me to suggest that we know the answer to that.

MEMBER BRIGGS: I was just curious.

CO-CHAIR GEORGE: Mladen?

(Off mic comment.)

MS. BARTON: Again, the use of the measure in our own Heart Stroke Recognition Program is pretty clear. Clinicians sample their own patients who are eligible for the measure and it would --- they are incentivized to do well, and so they, I think, generally pick good samples to help themselves do well.

In the PQRS side, the measure --- again, I'm not the implementer. CMS is the implementer of the measure on the PQRS side, and so the --- typically, though, the way that any measure reporting works is you select your

denominator first. And it would generally behoove you to be specific in selecting your denominator, and having it --- so, in this case it's people who have extant cardiovascular or other ischemic vascular disease, so it would not include people who have non-IVD indications for aspirin.

CO-CHAIR GEORGE: Other comments? If not, we'll move on to a vote on the opportunity for improvement.

MS. IBRAGIMOVA: So, the importance to measure and report 1d performance gap. Data demonstrate a considerable variation or overall less than optimal performance across providers and/or population groups, disparities in care. One, high; two, moderate; three, low; four, insufficient.

MS. VICALE: Tom, if you could please cast your vote via the phone.

MEMBER JAMES: Yes, not email?

MS. VICALE: We're having trouble receiving the email, so if you could say your vote that would be helpful. Thank you.

MEMBER JAMES: Two.

(Voting)

MS. IBRAGIMOVA: The results are 50 percent high, 50 percent moderate, zero percent low, zero percent insufficient.

MEMBER SPANGLER: So, next is specifications. I thought the specifications were clearly defined. Dan kind of summarized that. He also went over kind of the changes that were made since the last maintenance regarding antithrombotic, antiplatelet, as well as the exclusions.

I thought the specifications were pretty consistent evidence, and because of that, thought it would be implemented pretty consistently.

One thing to note is the data sources. There's actually a lot of data sources for this, so it's claims, electronic data, paper medical records, et cetera, et cetera, et cetera, across all different types of sources. I don't know if you had anything else to add.

CO-CHAIR GEORGE: Any questions on reliability? Any other comments? Ellen.

MEMBER HILLEGAS: There were some concerns regarding the fact that there are these patients with atrial fib that would be on anticoagulants now. And with respect to that, I'd just like to talk with Linda. So, would they all be excluded from this measure now, and that would make a difference in your reliability, because you've got a different population that you were measuring before, and now a different population now, because of the anticoagulant exclusion. But the score, the reliability score was very high, .88, so --

CO-CHAIR GEORGE: Joe?

MEMBER CLEVELAND: I just have an overall general question that's in a lot of these, but I guess the data sources where we're using both ICD-9 and ICD-10 codes, we noted in a lot of these --- in a lot of our worksheets that there's not a conversion methodology. I'd like if the developers could maybe tell us a little bit about that?

CO-CHAIR GEORGE: Specifically, crosswalking?

MEMBER CLEVELAND: Yes, a crosswalking, so is that something that we're okay with for using as administrative claims? I am but I just wanted to know what if you all had thoughts about it.

MS. BARTON: Never asked to provide a crosswalk.

MEMBER CLEVELAND: Got you, so we're not.

MS. BARTON: We have them.

MS. WILBON: So, we don't require a crosswalk, but we do ask the developers to give like a summary, a description of how they did the crosswalk, so how they came up with the codes they failed to do the lines, so that should -- it's not required, but -- if it's something that you're interested in, we can follow-up with the developers.

CO-CHAIR GEORGE: Kristi.

MEMBER MITCHELL: So, point of clarification. In the current forms there's no

requirement to submit a crosswalk. Okay, thank you. Going forward, if we have an opportunity to kind of make a recommendation, we should probably have that in the forms.

MS. WILBON: We do, actually. I don't --- honestly, I don't know exactly the dates of when we started requiring that, but we have been -- for some time, been communicating that we do ask for a set of the IC-9 codes, a set of the IC-10 codes, and then a description from the developer on how they did that, if they're not able to provide the actual crosswalk so that we understand how that conversion was made.

CO-CHAIR GEORGE: Other comments on reliability?

MEMBER SPANGLER: I'm sorry. Actually, that's for all measurements, because I'm just wondering if because this is a --- the second time it's being maintained, some of the maintenance measures may not ---

MS. WILBON: Yes, it's for all measures that currently use ICD-10 codes, or billing codes,

or measure specification codes.

MEMBER SPANGLER: So, not just new, but even a maintained measure.

MS. WILBON: Yes.

MEMBER JAMES: Is not the standard the CDC CMS GEMs, particularly in cardiovascular where there should be a standard crosswalk?

MS. WILBON: We don't, necessarily, specify which crosswalk tool they use. GEMs I know is one that's publicly available, but I know there's other tools, electronic tools, that are out there that folks can use, so we don't necessarily say which tool they have to use, as long as they can explain and justify how they came up with their ICD-10 codes.

MEMBER JAMES: Okay.

MEMBER SPANGLER: Should I continue on with reliability?

CO-CHAIR GEORGE: Yes, anything you have on reliability.

MEMBER SPANGLER: So, testing was done through a beta binomial model measuring signal and

noise ratio. The results, as Ellen I think mentioned the number, showed high reliability, so I would give that a high rating.

CO-CHAIR GEORGE: If there are no comments, we'll move to voting on the reliability.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2a reliability, including 2a(1) precise specifications, and 2a(2) testing, appropriate method and scope with adequate results. One, high; two, moderate; three, low; four, insufficient.

MS. VICALE: Tom, if you could please state your vote over the phone?

MEMBER JAMES: One. One.

MS. VICALE: Thank you.

(Voting)

(Off the record comments)

MS. IBRAGIMOVA: And the results are 69 percent high, 31 percent moderate, zero percent low, zero percent insufficient.

CO-CHAIR GEORGE: Move on to validity.

MEMBER SPANGLER: So, face validity was

done. NCQA has a pretty thorough process for this, and they also have two expert panels. They didn't actually provide a system results but noted that the majority of panelists believe the measure can actually distinguish between doing them poor quality. So, overall, I think the validity is pretty good.

I mean, I would like to see the numbers instead of just saying majority, but it seems like based on the expert panel that it's good validity for the members.

MEMBER HILLEGAS: I would have to say if you go by the algorithm, since we have only expert opinion, I think the highest you can vote for this would be moderate.

MEMBER SPANGLER: Yes. And that's --- I would have given a moderate, but you're right. I agree with you.

CO-CHAIR GEORGE: Comments or questions? If not, we'll vote on validity.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2a validity

-- 2b validity including 2b(1) specifications consistent with evidence, 2b(2) testing appropriate method and scope with adequate results and threats addressed, 2b(3) exclusions, 2b(4) risk adjustment and stratification, 2b(5) meaningful differences, 2b(6) comparability, multiple specifications, 2b(7) missing data, eMeasures, composites and PRO-PMs. One, high; two moderate; three, low; four, insufficient.

MEMBER SPANGLER: I'm sorry. Can I make a note real quick, because we didn't --- I was only doing --- sorry, 2b(2). There was a question regarding threats to validity about --- I think I have this noted here. There were some differences noted in clinician performance. I don't know if there is any threats to the validity based on that, but I have it noted here that there are differences in clinician performance. So, I didn't know if they wanted to address it, because it says here there's a significant difference in the performance between the 20th and the 75th percentile. I don't know if the developer has any comments on that.

MS. BARTON: Actually, I'm glad you brought that up. Typically, a measure is most valuable when there is variation because, otherwise, it would be hard to use to improve quality. So, if a measure was at 98 percent, and everybody was reporting it at 98 percent, then you would say we don't really need a measure here. So, the fact that there is variation in clinician performance is part of what makes this a useful measure, I think.

MEMBER SPANGLER: So, I guess the question is --- I totally agree with that, Mary. The question is, is there a possibility that sometimes that if there's enough meaningful difference, that could be a threat to the validity of the measure?

Do you guys ever --- because I think there definitely needs to be some meaningful difference. If there's too much between the high end and the low end, do you guys ever consider that a threat to how valid the measure can be?

MS. BARTON: I could see the theoretical

possibility of that. The way that our Heart Stroke Recognition Program, which is where we use this data, works, it's a relatively small N, and clinicians can get full credit in the program by performing well on most of the measures.

So, we definitely see that some folks seem to not concentrate on one or two of the measures as they submit their data to us to achieve the recognition through our program. So, it's hard for me to make --- it's hard to entertain that exact hypothesis in this relatively small data set.

CO-CHAIR GEORGE: Any other comments or questions? All right, we'll continue voting.

MS. VICALE: Tom, if we can receive your vote.

MEMBER JAMES: Which way? Will you send me a text, a note for text?

MS. VICALE: Through text, please. Did you receive the email with the number to text?

MEMBER JAMES: No. Let's see. There it is. It just came. Okay.

(Voting)

MS. IBRAGIMOVA: The results are 7 percent high, 93 percent moderate, zero percent low, zero percent insufficient.

CO-CHAIR GEORGE: Move on to feasibility.

MEMBER SPANGLER: So, the data collection is done, as noted before, from the multiple sources. There doesn't seem to be any barriers, particularly with NCQA Heart Stroke Recognition Program, so it seems there's pretty good feasibility. I didn't have any issues.

The only question would be the conversion, but it seems like -- that they're prepared for that.

CO-CHAIR GEORGE: All right. We'll vote on feasibility.

MS. IBRAGIMOVA: Feasibility, 3a data generated during care, 3b electronic sources, and 3c data collected can be implemented, eMeasure, feasibility assessment of data elements and logic. One, high; two, moderate; three, low; four, insufficient.

MS. VICALE: Again, Tom, please text your vote.

MEMBER JAMES: Yes.

(Voting)

MS. IBRAGIMOVA: The results are 56 percent high, 44 percent moderate, zero percent low, zero percent insufficient.

CO-CHAIR GEORGE: We'll move on to usability.

MEMBER SPANGLER: So, as stated NCQ uses this in their own program. It's also used in three CMS programs, PQRS, their EHR Incentive Program, and the Medicare Shared Savings Program. And NQF has also recommended for further use in two other CMS programs, so I think it high usability.

CO-CHAIR GEORGE: Any comments or discussion? Sana?

MEMBER AL-KHATIB: Just a question regarding the data that you've collected so far, since this being used, this measure is being used through these different systems. Do you have any data on whether this has had an impact?

MS. BARTON: Unfortunately, we're not able to get that data from CMS.

CO-CHAIR GEORGE: Other comments or questions? If not, we'll move on to voting.

MS. IBRAGIMOVA: Usability and use, 4a accountability transparency, used in accountability within three year, public reporting within six year, or if new, credible plan. 4b improvement, progress demonstrated, if new, credible rationale. And 4c, benefits outweigh evidence of unintended negative consequences to patients and populations. One, high; two, moderate; three, low; four, insufficient information.

(Voting)

MS. IBRAGIMOVA: The results are 75 percent high, 25 percent moderate, zero percent low, zero percent insufficient information.

CO-CHAIR GEORGE: So, any last minute questions?

MEMBER SPANGLER: I just wanted to make a note. At the end, as you can see, I don't know

if people looked, but there's several relating or competing measures. And general comments by the public and Members were, you know, to somehow harmonize these or, you know, there is some preference for actually the composite measure, instead of these individual measures. So, I don't know if that's something we want to discuss, but I think that's probably something that, you know, should be raised.

CO-CHAIR GEORGE: Okay. So, we will save the related and competing comments for our after in-person meeting call. Any other questions or comments before we vote on the overall measure? Liz?

MEMBER DELONG: I guess I'm concerned about the proliferation of measures. As we go forward, there are more measures, plus there are the ones that have already been endorsed and come up for re-endorsement, and don't provide evidence of impact. And it seems that we're eventually going to have an enormous number of measures out there, and it's going to be very confusing.

And I'm just making a comment. I don't know what the solution is, but it does seem that the proliferation is concerning.

MEMBER SPANGLER: And some people would say we're already there.

DR. BURSTIN: Just to add, I mean, that is an expectation of having a Standing Committee, is you will help us right size the portfolio. I mean, in the instance we had earlier, in fact, there was so much work done, we had gotten down to one general hypertension measure.

So, again, I think there's room certainly to try to harmonize where we can. I know the developers have worked pretty closely to at least harmonize on the evidence, and the numbers, et cetera, but we'll do whatever we can to keep pushing on that.

CO-CHAIR GEORGE: Any other comments? If not, we'll vote on the overall suitability for endorsement.

MS. IBRAGIMOVA: Overall suitability for endorsement. Does this measure meet NQF

criteria for endorsement? Note, this may not yet be a recommendation for endorsement. Final recommendation for endorsement may depend on assessment of any related and competing measures. One, yes; two, no.

MS. VICALE: Tom, could you please text your vote?

MEMBER JAMES: I'm back online. I just sent it over in chat.

MS. VICALE: Thanks, Tom.

(Voting)

MS. IBRAGIMOVA: The results are 100 percent yes, zero percent no.

MEMBER HILLEGAS: Can I just make a comment? On our worksheet, we actually did have the composite measure 0076 as mentioned, that Jason said, that there was a competing measure.

In the future, is there a way to get a brief synopsis of what these competing measures have when we're discussing this, and looking at these measures? Could we learn about these competing measures, or is that not okay?

MS. WILBON: No, it's certainly something -- we could probably like provide a hyperlink or something too, so that you guys could get to it. We do try to just make sure, though, that the evaluation of the measure based on the measure in front of you and not kind of making comparisons to the other measures. You're evaluating the measure in front of you, but that's certainly something we can do. Thanks.

MEMBER MITCHELL: I have another question or comment around the harmonization process. And exactly how does that happen, because this is like the third round, and I don't feel like we are living up to our charge to harmonize.

And to Liz' point, we just keep adding more. And thankfully, you know, in many cases for good reason, things are coming off the table. But we're not --- I don't feel like we're harmonizing.

DR. BURSTIN: It's certainly been a struggle. I mean, it's not easy. I think, you know, what you see is when the developers come to the table they've got a fully baked measure, so at that

point it's difficult for them to say okay, so you're going to change this, this, and this. They're often in programs, et cetera, so we have been --- we need to go back and do the assessment.

We have been giving developers, you know, up to their necks annual update to try to make those changes to harmonize. And we've actually seen some progress, and it's probably worth quantifying that to share with committees. But, you know, we just fully recognize it can't happen on a dime.

The most important thing I think we can do is just prospectively, up front, try to get as many folks together as possible to work on measures collaboratively, so we don't wind up with a measure for health plan, a measure for provider, a measure for CAD, a measure for general. And that's, I think, a longstanding goal.

MEMBER SPANGLER: I mean, we're going to even talk about that tomorrow, right? Because your measure, 67, is very similar to this measure. It's just, you know, two different conditions. So, is there any way to combine them? I don't know.

MEMBER VIDOVICH: Yes. Thank you, everybody, for bringing this up because our measure tomorrow is almost verbatim, just minor modifications, so it is difficult to review both of these. Right?

CO-CHAIR GEORGE: All right. We'll move on to the next measure. I think Ellen and Jason, again. And we'll -- just a few brief comments.

MR. ROMAN: Sure. This is NCQA's persistence of beta blocker after heart attack measure. It was developed in 2005, and first endorsed in 2009. It's focused on patients 18 years and older who have had a heart attack, and assesses whether or not they received persistent beta blocker treatment for six months after discharge.

The measure is used in NCQA's Health Plan Ranking and Accreditation Programs, and included in our annual State of Health Care Quality Report. There are no changes for this maintenance round. That's all. Thanks.

MEMBER HILLEGAS: Okay. And I'm going to speak to this measure. And the first thing is, is

that this is an intermediate clinical outcome measure. So, when you're looking at your evidence, we go through the high, moderate, low rating. The level of analysis is at the health plan and an integrated delivery system. And this is a maintenance measure submission, as was said.

So, if you look at the evidence, what they supply is they supply guidelines, clinical practice guidelines, as well as a systematic review. And this measure is looking at beta blockers for 180 days post-discharge from acute MI. And the rationale is that the persistent beta blocker treatment after an MI will reduce the risk of mortality, reduce the risk and severity of reinfarction, and improve the preservation of left ventricular function.

The two guidelines, one is on management of a STEMI, and the other guideline is on management of a non-STEMI. The STEMI guideline is graded a Level B, that beta blockers should be continued during and after hospitalization for all patients, but it does not discuss reinfarction

rates, nor does it discuss outcomes.

The actual guideline says, the benefit of beta blockers for secondary prevention has been established on numerous trials conducted in the pre-reperfusion era and appears to be greatest for patients with MI complicated by heart failure, left ventricular dysfunction, or ventricular arrhythmias.

Long-term duration has not been prospectively addressed, and so there's a concern whether that guideline actually reinforces or supports the measure.

The non-STEMI guidelines are actually graded a Level C. And the systematic review was done actually through 1999, so there's concerns about the dating of the evidence. It's good evidence, it's strong evidence, but the date of the evidence is a concern. So, in looking at the evidence, there is concern about the evidence being strong.

So, the highest that it could be rated if you go by the STEMI guidelines or the systematic review to me would be, because of the STEMI

guidelines, would be moderate. I don't think you could go high evidence with this, but it also includes non-STEMI. And the non-STEMI guidelines are rated a C, which would be low.

So, Jason, I don't know if you wanted to chime in?

MEMBER SPANGLER: No, I don't have --- I agree. I mean, I think the evidence is moderate at best.

CO-CHAIR GEORGE: Discussion, comments, questions?

MEMBER AL-KHATIB: Quick question. When you're talking about persistence of beta blocker use, like what period of time are you talking here?

MR. ROMAN: So, the way the measure is constructed, we look at patients who had a heart attack six months prior to the measurement year and six months in, so we kind of create that calendar, that time period. And then we look at 180 days post discharge, and look to see that they had enough dispensed medication to hit 135 out of 180 days post-discharge. So, it becomes I think 75 percent

of the proportion of days covered in the treatment period.

MEMBER AL-KHATIB: I mean so the data are very strong for the use of beta blockers in patients post MI. I'm not sure that I agree with the comment that the evidence is moderate. The evidence is very strong for the use of beta blockers post MI.

I mean, where I think you might argue is if you're looking at it like five years after an MI. That's where really the evidence is lacking, because the clinical trials that looked at that didn't last that long. But if you're talking about within a year, within a couple of years, I actually would argue that the evidence is pretty strong.

CO-CHAIR GEORGE: Liz?

MEMBER DELONG: I'd like to ask, you're basing your evidence assessment on the fact that they supplied two guidelines.

MEMBER SPANGLER: Yes, that's not our opinion of what the evidence is. It's what they are saying the evidence is.

MEMBER DELONG: Right. So, is there

additional evidence that they haven't brought forward that Sana knows that the guideline developers ignored or whatever?

MEMBER AL-KHATIB: Certainly, in the setting of SD segment elevation MI there are multiple randomized clinical trials showing benefit. So, I have no concern especially about patients with SD segment elevation. I don't think the evidence is as strong for non-SD elevation MI, but there trials there, too. Maybe I'm missing something.

MEMBER VIDOVICH: I mean, the only evidence is that patients maybe in cardiogenic shock didn't do well, but overall there's very strong evidence that --- I would agree with you.

MEMBER PHILIPPIDES: It is significantly weaker in patients with non-STEMIs who have been revascularized, who have normal LV function. It actually is harder to show benefit out even out to about a year. I mean, it gets wishy-washy. No question with the STEMIs.

But in the modern era going to the Cath

within three seconds, opening up the vessel, LV function normal. It's not as strong as you would think.

CO-CHAIR GEORGE: Leslie?

MEMBER CHO: So, in the new ESE 2015 Guideline on Non-STEMI, actually beta blockers for long term management, especially with EF less than 40 is a 1a guideline, 1a indication, so there is very strong evidence for non-STEMI with EF less than 40. I agree with George that with EF greater than 40, or normal EF, the evidence is --- but for EF less than 40, for sure there's great evidence.

MEMBER SPANGLER: It seems like maybe some of the evidence is missing from this, because there's nothing that they supplied that is anything close to A, at all, it's only B or C.

MEMBER CHO: It's 1a.

MEMBER SPANGLER: Yes, so then it's missing.

MR. ROMAN: Just to be clear. We supplied two guidelines from the ACC and EHA, and we supplied their rating of the evidence. We did not rate the

evidence ourselves, so the recommendations we chose are from the ACC and EHA. They're the ones that fit with --- they're the ones that the measure is based on. And that is the rating of evidence that they supplied.

And as far as the dates of the systematic review, it is old. It is also kind of the seminal body of work that many of the recommendations when you tease out where they came from, and what other studies cite, it goes back to that one. So, since we had kind of this --- wide volumes, huge volume of evidence to summarize we chose one systematic review that supported the measure the best. So, I just wanted to clarify kind of where we got our ratings, and why we cited that one systematic review.

MEMBER CHO: Is there a time limit for this guideline? So, if you have a beta blocker after a heart attack, do you --- is it always, because --

PARTICIPANT: See, that's what I asked.

MEMBER CHO: Yes.

PARTICIPANT: That was the initial question.

MEMBER CHO: Because that's actually --- the data is very poor after two to three years.

MS. BARTON: Right. So, typically, it's very hard to make long term --- to make guidelines that are based on evidence for long term anything. But this measure is focusing on the six-month period after an event where it seemed as though there was not only a clinical agreement, but an evidence-base to support that.

MEMBER CHO: Only for six months. Okay.

MEMBER DELONG: I have another question. Do we vote on the evidence that the developer supplies, or do we vote on the consensus evidence of cardiologists in the room?

DR. BURSTIN: We generally have to look at what is on the form but, obviously, we convene a group of experts intentionally. Hearing this additional information is something you can consider, as well. But what's on the form is really

--- is primary.

CO-CHAIR GEORGE: Any other comments or questions?

MEMBER AL-KHATIB: I just want to say that, I mean, I think we definitely --- we cannot dismiss the evidence, even though maybe the developers did not include that very clearly, or vigorously as, you know, as we would have expected them to do.

I mean the evidence is there, and I don't think we can dismiss it. So, though I agree that we need to look at the worksheet, we also need to be open minded, and if we are aware of the evidence, I actually think we need to take that into account.

CO-CHAIR GEORGE: All right. I guess we'll go ahead and vote on the evidence.

MEMBER HILLEGAS: May I say one more thing?

CO-CHAIR GEORGE: Yes.

MEMBER HILLEGAS: The measure actually talks about acute MI. It does not differentiate

STEMI versus non-STEMI. And we are discussing the evidence, and there's a little different value for non-STEMI evidence versus STEMI evidence, and left ventricular function. So, if you're going to vote on the evidence, you need to consider all the evidence for a MI, and not differentiate between STEMI, non-STEMI, and STEMI or non-STEMI with heart failure.

MS. IBRAGIMOVA: Importance to measure and report, 1a evidence structure process intermediate outcome. One, high; only eligible if QQC submitted; two, moderate; three, low; four, insufficient.

(Voting)

MS. IBRAGIMOVA: The results are 25 percent high, 63 percent moderate, 13 percent low, zero percent insufficient.

MEMBER HILLEGAS: So, if you look at performance gap, was performance gap data on the measure provided? There's no statistical data to demonstrate gap and tear, but there is evidence that there is disparities issue. This appears to

be a disparity-sensitive measure.

CO-CHAIR GEORGE: Questions or comments? If not, we'll go to voting on the opportunity for improvement.

MS. IBRAGIMOVA: Importance to measure and report, 1b performance gap, data demonstrated considerable variation and/or overall less than optimal performance across providers and/or population groups, disparities in care. One, high; two, moderate; three, low; four, insufficient.

(Voting)

MEMBER SPANGLER: Dan or Mary, my understanding is that this measure is not a follow-up, but the measure --- the beta blocker measure that was retired, this is kind of an extension of that. You know, this is a persistence type measure. Can you remind me the number that it reached where you retired it? Does anybody know?

MS. BARTON: It was 99 percent.

MEMBER SPANGLER: Oh, it was 99. Okay. Got it. Thank you.

MS. IBRAGIMOVA: The results are 19

percent high, 81 percent moderate, zero percent low, zero percent insufficient.

CO-CHAIR GEORGE: We'll move on to reliability.

MEMBER HILLEGAS: So, if we look at the numerator, it's patients who had a 180-day course of treatment with beta blockers post discharge. The denominator is patients 18 or older as of December 31st of the measurement with a diagnosis of acute MI. And so, the concern is, is that is there a way to actually determine whether it was a non-STEMI or a STEMI? Does that make a difference in your denominator?

And then the other thing is, is what if some of your patients did have an MI, but they actually went on and had bypass, or stent, or whatever? So, is their diagnosis now not acute MI, but is it CABG, or is it post-stent? So, will the denominator capture the MIs that went on for a procedure? That was a question for the developer.

MR. ROMAN: So, I believe the way this works is we're looking at patients who are

discharged after having an acute MI. In our code set, we do have non-STEMI and STEMI codes, so it doesn't make a difference. It's just all included.

But as far as somebody who went on and had a CABG, I think that's a different --- you're discharged with having had a CABG procedure. That would not be included in this measure. We're looking at discharges with the discharge being from having an acute MI, so somebody who had a CABG procedure I don't believe would be included in the way that this is calculated.

MEMBER HILLEGAS: I don't know if that makes a difference to the expert panel, because you lose a lot of people who go on and have a procedure. And does that make a difference in what you find?

So reliability testing was done, and the reliability testing reported out at .81, which is high. So, you could vote this as a high reliability, or if there are concerns with the denominator you might say moderate reliability, in my opinion.

MEMBER SPANGLER: The only question I

had, too, for the developer was around the 75 percent threshold for persistence, and the number of allowed gap days, where that came from, why that number was picked? I mean I understand that's how you're defining persistence, but I'm just wondering the reasoning behind those specific numbers.

MS. BARTON: As this measure was developed, that was thought to be the best way to assign consistent use. Now, we are certainly aware that in the last 15 years, the Pharmacy Quality Alliance has moved forward with a number of measures that use an 80 percent of days covered as a threshold, and that's something that we're going to look at.

As we noted, this measure had not --- there are no changes in the specifications at this time that we bring it forward. This is one example when our timelines don't always overlap perfectly, so I think the next time we reevaluate this measure we would look at that question of whether aligning with what is now the sort of

accepted PQA threshold of 80 percent might make sense to align this measure better with other adherence measures.

MEMBER VIDOVICH: I'm a little bit confused by the answer you asked for, Ellen, about this treatment, because overwhelming majority of patients with acute MI will undergo some sort of diagnostic angiography, and then will get some sort of revascularization, unless they're treated medically, which also happens. You know, maybe in STEMI more than STEMI.

So, do we exclude patients who receive revascularization in this measure? Am I understanding this correctly?

MR. ROMAN: It's not that they're excluded. It's that they're not included in the denominator the way it's defined. We are looking at patients specifically who were discharged after having acute MI.

MEMBER VIDOVICH: So, then maybe the title of the measure may not be accurate then. Right? Because then this should be said patients

with acute MI who do not receive revascularization and are treated medically. Because it's --- at least in this country, almost everybody gets some sort of diagnostic angiography, and some sort of revascularization. I mean, there's some rare exceptions probably in the teens that don't.

MEMBER CHO: This is a DRG-based. This is an ICD-9-based code. Right? So, when --- just in general, you're not going to discharge a patient on a PCI code because you're going to get paid less. You're going to discharge a patient based on their MI code, because you're going to get paid more.

So, they're not going to --- right. So, their MI code, I don't think --- what we're saying is, is that what is the number one diagnosis is what you guys capture. Right? But any DRG.

MS. BARTON: We have to review the specifications to know the answer to that. I don't think it's only the first diagnosis.

MEMBER CHO: But, I mean, majority of patients that get admitted to America, they're going to be charged based on what is going to get

the most amount of billing, so you're not going to have a patient who gets admitted for a heart attack and just code them as having PCI. Nobody is going to do that, not even bypass. So, I don't think you're going to exclude all these patients.

MEMBER AL-KHATIB: Perhaps what you are trying to say is that those patients maybe are not excluded from the denominator, but maybe there is not a way to identify them. Is that what you're trying to say? Because I don't --- I mean, in reading all these specifications, I don't see how those patients are actually getting excluded.

MEMBER SPANGLER: I think that's what he said, that's what Dan said. Right? They're not excluded. He just said --- I think what --- he said basically what Leslie specified, that if they have that code, they're not going to be included, but there's going to be plenty who have had revascularization that have the MI code that would be included.

MS. BARTON: Could I specify? So, in reviewing the detailed specifications, the fact of

an acute MI being on the discharge --- being a discharge diagnosis is not limited in any way as to within the first, or first five, or first ten. So, I think actually the --- what Dan said before, we might need to amend to say that someone who had an MI and then had a procedure done in the hospital, in all likelihood, is still going to appear in this measure because they --- and, in fact, I think that the evidence still supports the use of beta blockers in those patients for the period of time after their discharge.

MEMBER PHILIPPIDES: Can I look at the other end of the spectrum? So, would this also include the 90-year old woman who's anemic, who goes into atrial fibrillation one night, has a component bump of .01, whatever it is that's positive in your hospital, and then somewhere on the list gets coded as an acute MI, a type 2, but an MI. Would that person also be included because administratively it was deemed an MI? I'm making the point between physiologically, you know, beta blockers helping, and administratively being coded

MI, we're not really sure there's benefit for that person. I don't think that they fit. But I think there are code ---

MS. BARTON: I can understand that, and the best that this measure comes --- the closest this measure comes to making sure that that's --- that there's not a harmful incentive is to exclude people with bradycardia, exclude people with asthma or COPD, and the other intolerance or allergy to beta blocker therapy. That's not really a big deal, but I think that the issues of people for whom beta blockers would clearly be harmful are excluded from the measure. But there's not an upper age limit.

MEMBER PHILIPPIDES: I think some of these people have really component bumps for lots of other reasons, some of which you've excluded, some you haven't, and it gets thrown into that.

MEMBER VIDOVICH: It's essentially the Type 2 MI. Right? By the World Health Organization --- yes, which does get captured. Right? I think we all know in our clinical practice that they do

contaminate your data, and you have to manually pluck them out to get your quality up, so I think this measure will contaminate them.

CO-CHAIR GEORGE: Any other questions or comments on the reliability? If not, we'll move to voting on this.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2a reliability, including 2a(1) precise specifications, and 2a(2) testing appropriate method and scope of adequate results. One, high; two, moderate; three, low; four, insufficient.

(Voting)

MS. IBRAGIMOVA: The results are 13 percent high; 81 percent moderate; 6 percent low; 0 percent insufficient.

CO-CHAIR GEORGE: Move on to validity.

MEMBER HILLEGAS: Okay, so validity was done with an expert panel. They determined it was valid empirical testing with moderate correlation with other measures. They compared it particularly to comprehensive diabetes care and cholesterol

management, and pharmacotherapy with COPD exacerbation, but actually did not do any kind of empirical testing.

If you go by your algorithm for validity, therefore, then you're talking about face validity, and the highest you could rate this validity would be moderate. They did have --- yes, so that's it. Sorry.

CO-CHAIR GEORGE: Any threats to validity?

MEMBER HILLEGAS: Yes. According to the Box Plus, there's a 7 to 11 percent gap in performance between the first quartile and the third quartile. The largest gap was found in the Medicaid plans. There are meaningful differences, but small.

CO-CHAIR GEORGE: Any comments or questions on validity?

MS. BARTON: Could I make a point? On our measure submission form under 2b(2), we do present data on empiric validity, so I'm a little confused. So, under NQF's instructions in measure

endorsement forms, construct validity is included as a form of empiric validity testing. And as such, that is where we put the data, where we were asked to put the data. And, in fact, there's not a requirement that there be other kind of validity testing.

MEMBER HILLEGAS: Basically, what they're talking about is their method of testing was they correlated the following measures with comprehensive diabetes care, the cholesterol management for patients with cardiovascular conditions, and the pharmacotherapy management of COPD. And in that, they came up with a value near .8, but the face validity was from two expert panels. So, they used construct validity for the empirical comparing it to these others, correlating it with the following measures, comprehensive diabetes, et cetera. And then they used expert participation.

Again, in my personal opinion, I feel that validity is no higher than a moderate, but you may all disagree.

MEMBER SPANGLER: Just a note; three expert panels, not two. But, yes. I mean, I think there --- to me, there's constructive face validity both done, but I think that they both show moderate validity. And even the construct validity, they said it's moderate.

CO-CHAIR GEORGE: Additional comments, discussion? If not, we'll vote on validity.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2b validity, including 2b(1) specifications consistent with evidence, 2b(2) testing appropriate method and scope with adequate results and threats addressed, 2b(3) exclusions, 2b(4) risk adjustment plus stratification, 2b(5) meaningful differences, 2b(6) comparability and multiple specifications, 2b(7) missing data, eMeasures, composites, PRO-PMs. One, high; two, moderate; three, low; four, insufficient.

(Voting)

MS. IBRAGIMOVA: The results are: 0 percent high; 94 percent moderate; 6 percent low,

and 0 percent insufficient.

CO-CHAIR GEORGE: Move on to feasibility.

MEMBER HILLEGAS: The data is from electronic clinical data, pharmacy, and administrative claims. The caution still is the denominator, identifying individuals with a MI and whether definition of a MI is reliable across the whole continuum. But it does seem feasible from --- based on the different sources that they have.

CO-CHAIR GEORGE: Questions or comments on feasibility? If not, we'll vote on feasibility.

MS. IBRAGIMOVA: Feasibility 3a, data generated during care, 3b electronic sources, and 3c data collected can be implemented, eMeasure feasibility assessment of data elements and logic. One, high; two, moderate; three, low; four, insufficient.

(Voting)

MS. IBRAGIMOVA: And the results are: 56 percent high; 44 percent moderate; 0 percent low; 0 percent insufficient.

CO-CHAIR GEORGE: We'll move on to usability.

MEMBER HILLEGAS: And, again, if you look at usability and use, the benefits outweigh the unintended consequences at this time. It's publicly reported on Quality Compass, and on annual State of Health Care Quality, so I would say it's usable.

CO-CHAIR GEORGE: Any comments or discussion on usability? If not, we'll move to a vote on usability.

MS. IBRAGIMOVA: Usability and use, 4a, accountability transparency, use and accountability within three-year, public reporting within six year, or if new, credible plan, 4b improvement progress demonstrated, if new, credible rationale, and 4c benefits outweigh evidence of unintended negative consequences to patients and populations. One high; two, moderate; three, low; four, insufficient information.

(Voting)

MS. IBRAGIMOVA: The results are: 63

percent high; 31 percent moderate; 6 percent low; 0 percent insufficient information.

CO-CHAIR GEORGE: So, any last minute comments or discussion? Ellen?

MEMBER HILLEGAS: Yes. There are competing measures we'll be discussing I think tomorrow, 0070, which actually measures these similar patients for 12 months. So, again, this is a concern about multiple measures.

The 0070 actually spells out a prior MI or current ejection fraction less than 40 percent, so it's a little different description than just AMI.

MEMBER MARTIN: So, I worry a little bit about the pediatric cardiologist speaking about an AMI measure, but one of the things that strikes me is, you know, this is about beta blocker use. Then, you know, this is really something that begs for a composite measure because you could send them all home on a beta blocker but not be on an aspirin. So, if your hospital only measures this, you know what, you really don't like an antiplatelet drug.

I mean, this should be --- this is a --- you know, several things that need to be done in these patients, and just having --- pulling out one element of it and measuring it seems a little bit silly to me. But that's a pediatric cardiologist.

CO-CHAIR GEORGE: I think we'll have a lot of things to think about when we get to the competing and related measures and recommendations. Any other comments? If not, we'll move on to voting on endorsement.

MS. IBRAGIMOVA: Overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? One, yes; two, no.

(Voting)

MS. IBRAGIMOVA: The results are: 94 percent yes; 6 percent no.

MS. VICALE: Okay. Thank you, everyone. At this time, we'll invite Karen Johnson -- our Senior Director -- to discuss composite measures and provide a little bit of guidance for you as we come up to reviewing our first composite measure of the day.

DR. JOHNSON: Thank you, Leslie. So, today I have the pleasure to just remind you all of some of our guidance and definitions, and things like that that we use for composite measures. And, hopefully, this will ring a bell, especially for Liz, who is our Composite Expert Panel. So, thanks again, Liz.

We have defined at NQF composite measures as a measure that's a combination of two or more component measures, each of which individually reflects quality of care into a single performance measure with a single scope. And you can read all about it in the link that was noted there.

But there are actually --- NQF actually also has defined for the purposes of evaluation and endorsement some other kinds of measures as composite measures. And this is a little bit different sometimes than what you might see out in the world. But at NQF, we define what are called all or none measures, or any or none measures or none measures as composite measures. So all or none

measures, it's the kind of measure that all the things that are listed in a measure have to be done in order to get credit for meeting that measure. Any or none is kind of the flip of that. If you have any of several things listed in the measure, then you meet that measure. Often meeting an any or none measure is not a good thing because it's often complications or things like that. Next slide.

So, basically, when it comes to evaluating composite measures, particularly the any or none, or all or none composites, such as the ones that you have in your list of measures this time around, is that there are a couple of extra criteria that we ask you to consider. So, one comes under importance to measuring report, and it's basically the idea that measure developers need to be able to state their thinking behind how they constructed their measure. And that's pretty much all this is. But we would like to see a description of the quality construct, so what's the overall thinking behind it? What components are included, and how those components work together? And really

how the measure is constructed -- glued together is my scientific way of saying it -- to reflect that construct that they are trying to come up with. So, that's what 1d is about. Next slide, please. Thank you.

And then the second additional criterion for composite measures comes under the scientific acceptability criterion, and it has to do with the empirical analysis that support the construct of the composite. So, basically, this is the idea that the component measure should fit the construct, and that the aggregation and weighting roles are consistent with the construct. And, also, we ask about missing data, as well.

These are --- can you go back just a little --- so, this --- for all or none and any or none measures, this is a pretty easy thing to do, because for any or none, or all or none measures, a lot of --- most of the empirical kinds of analyses that we would expect would be things like frequency distributions of the different components. It's pretty simple to do.

Okay. These next slides really just break down all the various sub-criteria under our major criteria to show you that in some cases we have to think about things for the composite as a whole, sometimes we have to think about things for the components, and sometimes both. Okay? So, specifically on importance to measure a report, there needs to be evidence at the component level. Right? So, everything that goes into the composite needs to be evidence-based. That's what that is.

Performance gap, we definitely want to think about it at the composite level. We also want to think about it for components for those more traditional kinds of composite measures. But for the any or none, all or none, not so much because they don't work independently within the measure. They are an any or none, all or none, so I don't know if I'm explaining that well, but really for today's work, you need to think about performance gap at the composite level. And then, clearly, 1d is for the composite itself. Let's go to the next one.

Okay. This is pretty busy, but basically what this is saying is, as you would expect, specifications --- you have to have good specs so that you know how to compute the components of the composite. So, components have to be --- the specs have to be considered for components. And testing really needs to be done at the composite level, not the component level. And that's true whether or not we're talking about more traditional composites, or the any or none, all or none types of composites.

For validity, we really want --- we prefer to have validity testing at the composite level. We can also look at it at the component level. It doesn't, again, make that much sense in your none, all or none, so you don't have to worry about it so much. But the other threats to validity really come into play for the components particularly, so you have to think about any exclusions that happen for the components. And let's go to the next slide. I don't have to go into detail here.

These are really just reflecting the notes that go with the other slide. And I think we'll get there. For feasibility, we're thinking about feasibility of the composite overall, but clearly, you have to also think about the components. But that kind of gets --- you think about that together when you're thinking about feasibility of the composite.

And then usability and use, for the most part, we're thinking about things at the composite level, not so much the individual components, but we're just saying that for accountability and transparency, we have to at minimum have the various components listed so that everybody understands what goes into the composite. But we also are interested in benefits outweighing the harms and improvement at the composite level.

Okay. And related competing, when that comes up, and I know you guys are just chomping at the bit to talk about related competing measures, unlike a lot of other people who don't want to talk about it. We have --- we need to think about related

and competing measures at both the composite level and the component level. And I would imagine that makes sense to you. You have to think about it for both.

And I think that's all about composites. So, let me start there and see if anybody has any questions about composites. It can be tricky, but it's less tricky with the all or none, any or none types of composites that you guys are working with this time around.

Okay. And now shall we just slide on into the SDS trial? Okay. So, you have probably heard about hopefully several times that NQF is in the middle of a two-year trial for SDS risk adjustment. So, the background is that about a year and a half ago, or so -- late 2013 -- NQF convened an expert panel to consider if, when, and how outcome performance measures should be adjusted for socioeconomic status or other demographic factors. So, up until then, NQF actually had a policy that those kinds of factors were not to be included in risk adjustment approaches.

And there's actually two diverging perspectives, and this is why we convened the panel. One is that adjusting for --- I'm going to SDS instead of saying sociodemographic factors every time. Adjusting for SDS factors will mask disparities. So, a lot of people are very concerned that if you do this, you'll mask disparities, and you don't want to do that. So, that's on one side.

The other side said you really need to adjust for these factors because, otherwise, you might make incorrect inferences about quality of care that's being provided.

So, after a lot of deliberations, the panel recommended and NQF's Board approved a two-year trial period. And, basically, during this two-year trial period, we have lifted that prohibition, so now developers are allowed to include SDS factors in their risk adjustment approach, if that is appropriate to do.

So, in terms of how this plays out in our day to day operations, it's --- we're hoping that it doesn't feel a lot different to you than

it did before. So, just like any other measure that comes through, each measure has to be assessed individually on all the different criteria, so SDS is just an extra piece that we're adding into your consideration.

What this means is that we're not saying that all outcome measures must be adjusted for SDS factors. That's not what this trial is about. What we are saying is that developers should, at least, consider it and think about the conceptual basis -- if there is one -- about whether or not it should be done. And I'm sure that you already know that when you're doing risk adjustment, you're already doing that for your clinical kinds of variables. Right? You have a reason to put those into your models, so it's the same thing with SDS factors.

And we also realize that efforts to implement this adjustment can be constrained by data limitations and data collection burden. So, in other words, there might be a very valid conceptual basis to do this kind of adjustment, but the data may not be there to make it possible. So,

we'll work our way through those discussions as we go through the day.

So, just so you know what our process is, April 15 was not only Tax Day, it was also a big day at NQF because that was the day that we said all measures submitted to NQF after April 15th is considered part of the trial period. So, it's not that just a few are a part; everything that comes under or comes through is part of the trial period, and committees will consider whether these measures have been appropriately adjusted, or not, including potentially for SDS factors.

For previously endorsed measures, when they come up for maintenance review, those are also part of the trial. And then there are also some other ways that measures can be part of trial, either as ad hoc requests, or if there was a conditional endorsement. And that happened with Re-admissions and Cost and Resource Use, and basically that happened because those two projects were underway at the same time that the SDS panel was making its deliberations and its final

recommendations. So, at the end of the day, the Board put some conditions on the endorsement of those measures pending what happened from the SDS panel. And those are going back around now. You can stay tuned for those.

So, how will this impact your evaluation of measures? Well, as I've already said, you'll still continue to evaluate measures just like you always have. So, you will think about the risk adjustment approach and whether it's appropriate or not, so nothing new there.

The --- you will continue to use the validity criterion to evaluate the appropriateness associated with demographic factors, as well as clinical factors used in risk adjustment. So, you may recall that when you talk about validity, you talk about how the specs conform or not to evidence. That's part of it. And then you talk about testing -- what was done and what were the results. And then you talk about the various threats to validity; that's when you talk about exclusions, missing data, things like that, but that's also where your

risk adjustment conversations come into play. So, again, that is where we will talk about SDS, where we always have talked about it.

This is --- sorry, you're going too fast for me. NQF has completed a preliminary analysis, as you notice, and we did try to identify in this preliminary analysis some of the questions that you'd have to think about when you consider the SDS factors. And we tried to make sure it was included, but it didn't overwhelm your analysis. Because, again, this is not about SDS; it's about evaluation of the measure and all of the criteria. Okay, thank you.

So, you'll be asked to consider the following questions. First, and most basic, is there a conceptual relationship between the SDS factor or factors -- there may be several -- and the measure focus? Measure focus is NQF's speak for what's being measured. Okay? What are the SDS factors that are available and analyzed? Does empirical analysis show that the SDS factor has a significant and unique effect on the outcome in

question? And does the reliability and validity testing match the final specs? Okay.

So, talking about it a little bit more in depth, we actually do ask specifically if there's a conceptual relationship between the SDS factors and the measure focus. So, that's one thing that we expect all the developers after April 15th, you know, once we're in this trial, we do expect the developers to at least discuss any conceptual relationships between SDS factors and the measure focus.

Now, it could be that they will say that there are no conceptual relationships between, and that's fine. If they say that, that's fine, that's their discussion. They may not say that, in which case they will probably want to be more verbose about that.

We don't have any rules really about what they have to show, so it doesn't have to be a systematic review of all these things. It doesn't even have to be, necessarily, published literature; although, I think most of what we've

seen so far is that sort of thing. Just like any kind of factor that would be included in a case mix adjustment or a risk adjustment, is the SDS factor or factors present at the start of care, and are they caused by the care that's being evaluated? So, in other words, you don't want to put in your adjustment something that happens because of the way quality, or the way care was delivered. So, you don't adjust out things like something went wrong and there was a complication, so you wouldn't put a complication in there. Most people wouldn't consider a complication an SDS factor, but that's the example that we would put in there.

So, data and variables. We would ask you to review the variables that are available and analyzed, and consider whether or not those that are available and analyzed -- so together -- do they align with the conceptual description? So, if there's a conceptual rationale or relationship between say income and the measure focus, maybe income, as such, is not available, but maybe poverty status -- yes or no -- is. So, does that

really reflect that conceptual idea of income? So you could think about that and have discussions about that. And are these variables available and generally accessible for the measured patient population? Okay?

Then we basically are asking for two sets of empirical analysis, and let me say that we expect empirical analysis if it actually is the case that there is, or the developer believes that there is a conceptual relationship. So, if they say right off the bat we do not think there's a conceptual relationship between an SDS factor and the measure focus, then it doesn't make sense to ask them to do empirical analysis. Right? But if they say there is, then --- and really if there's data available to look at it -- then we would ask them to talk about the importance of the SDS variables in the risk adjustment model. Okay, so that's where you come into your model diagnostics.

And, secondly, at least for the trial period we're actually asking developers to go a little bit further for us, and we're saying please

tell us about the scores that you get if you do include SDS factors in the risk adjustment approach compared to if you don't include them in your risk adjustment approach. And let's actually see what happens to those results. And the question there is, are the differences --- because you will see differences. Right? Any tiny, little change that you would have in a case mix adjustment, or risk adjustment approach, will make a difference. The question is: is it a substantial difference? Okay?

If so, so if they go that far, they did find a conceptual rationale. They had some data, and they analyzed it, and it turns out that they feel like that it's important in the model, and makes a difference in results, then at that point we would assume that they would say this is --- we actually want to continue, or to actually include those SDS factors in our model. So, what we would say is they should provide updated reliability and validity testing on the measure as specified.

So, really what that's getting at is it actually doesn't apply if it's a new measure

because there is no updated reliability and validity testing if it's a new measure. But if it's a maintenance measure that came in maybe before without SDS adjustment, now it's coming in with it, we're just saying update your reliability and validity testing, because we always say that testing should be for the measure as specified.

And then, finally, we say that we're asking the developer to provide information required to stratify a clinically adjusted only version of the measure results by the relevant SDS variables. So, all that means is if they do include SDS factors in their risk model, we're asking them to provide the specs for a model without those factors so that, if you want to -- you being you -- if you want to, you could actually stratify according to those SDS variables. So, if they included poverty status, yes or no, you would know that, and then you know that you could go in and stratify the results, above poverty versus below, and be able to look to see if you are seeing disparities in your results. Okay?

I think that was the last slide, so let me stop there and see if anybody has questions about our SDS trial. All right, that's it. Thanks.

MS. VICALE: Thanks, Karen. We'll continue on with the consideration of the next measure, 0694, and I'll turn it over to Mary and Tom.

CO-CHAIR GEORGE: Measure developers, can you give us just a brief overview?

DR. MASOUDI: Absolutely. Thank you. I'm Fred Masoudi from the University of Colorado, and I'm here representing the American College of Cardiology with Jensen Chiu from ACC Staff. There are also some of our methodologists on the phone line, if I understand correctly.

This measure is number 0694, hospital risk-standardized complications rate following implantation of an implantable cardioverter-defibrillator. This is a previously endorsed outcome measure of complications at the hospital level after ICD implantation. It employs a risk-adjusted modeling using validated

hierarchical logistic regression that is based upon established modeling principles to create hospital-level risk-standardized complication rates. The data to develop this measure were from both the NCDR ICD registry and linked to Medicare claims data. It is a composite defined as an any or none measure by the standards that were previously described of short-term device complications deemed important by a technical expert panel convened by the Centers for Medicare and Medicaid Services, with either a 30 or 90-day measurement time frame depending upon the complication.

There are nine clinical risk factors in the model which is slightly more parsimonious than the previous version. There are now nine clinical variables that are used in risk adjustment. That's an overall summary.

CO-CHAIR GEORGE: Joel and Tom?

MEMBER MARRS: So, to start off with the evidence piece, overall they provided a sample of looking at risk scores of these composites of 30

and 90 days, and so if you look at hospitals across the U.S., it ranges anywhere from as low as 4 percent, all the way up to 30 percent with --- in that range. With using this registry, they're able to capture about 90 percent ICDs that are actually reported. And the key piece of the evidence is they use CMS and Medicare patients above the age of 65, because that's the only available data to actually support and actually validate this measure.

CO-CHAIR KOTTKE: I don't have anything more.

CO-CHAIR GEORGE: Any comments or questions about the evidence?

CO-CHAIR KOTTKE: I guess the only question I had was in your ranked hospitals, does the complication rate range from zero to 17.8 percent?

DR. MASOUDI: That's the range, yes, in terms of the distribution of the complications as presented in the document.

CO-CHAIR GEORGE: Any other --- Liz?

MEMBER DELONG: So, for clarification

the deciles, are you saying the lowest decile had a zero complication rate?

DR. MASOUDI: No, I'd have to look back at the data. I don't know what the range of --- I mean, I don't know what the range --- I will have to look back at the data to understand range of complications within a specific decile. I think that was the lowest number for ---

MEMBER DELONG: That was the lowest number overall, not the number for the lowest decile. Is that right?

DR. MASOUDI: Yes. I would have to review that --- do you have the ---

MEMBER MARRS: What was reported and was submitted was that it was zero to 17.8 percent was the range.

MEMBER DELONG: The range over all hospitals or providers.

MEMBER MARRS: Yes. Right.

MEMBER DELONG: Not --- okay. I guess my point would be, if the lowest decile is composed of hospitals, all of which had a zero percent

complication rate, I would be a little worried. But you're talking about not the lowest decile, but the lowest hospital.

DR. MASOUDI: Yes. I mean, again, I'd have to look back at the source data. I'm not so sure that I would share the conviction that the lowest decile having a zero percent complication rate is necessarily concerning because, again, the --- you know, with overall complication rates that run about 4 percent, you can imagine that particularly in smaller sample size hospitals -- although we do truncate the sample at I think 25 hospitals -- that in a sampling period that number could be zero.

CO-CHAIR KOTTKE: To quote the worksheet, it said, "In these preliminary analyses, complications were seen in 5.7 percent of ICD admissions. There were 3,818 complications. The median complication rate following ICD implementation ranged from zero percent to 17.8 percent across deciles of hospitals' complication rate. So, it's possible that the lowest 10 percent,

no complication perceived, or in this limited data set.

Now, this is an outcome measure and so all we have to do is demonstrate that there's some relationship between the measure and the outcome. We don't have to ---

MEMBER DELONG: Yes, I guess my point would be if --- whether you're getting all the data, if 10 percent of your hospital, and I have no sense of whether you could expect 10 percent of hospitals to not have a complication ---

DR. MASOUDI: Yes. So, again, just to be clear how the outcomes are ascertained. This is not self-reported ascertainment; this is ascertainment through Medicare claims data, and the complications are all those that are either include death, or complications that are significant enough to require a hospitalization or a procedure. And so the ascertainment is actually quite --- you know, as complete as claims data can be for this population. It's not a self-reported complication, which I think is important.

CO-CHAIR KOTTKE: And these complications are significant; pneumothorax or hemothorax requiring a chest tube, a hematoma plus a blood transfusion or evacuation, cardiac tamponade or a pericardiocentesis. Yes, those are the 30 days, and then 90 days include system revision, open system revision, device-related infection, or additional ICD implantation, so it's not trivial or minor stuff; it's nine major things -- seven major things.

CO-CHAIR GEORGE: Any other discussion or questions on this, the evidence? All right, we'll go to a vote.

MS. IBRAGIMOVA: Importance to measure and report 1a evidence, health outcome or PRO, rationale supports the relationship of the health outcome or PRO to at least one health care structure or process, intervention or service. One, yes; two, no.

(Voting)

MS. IBRAGIMOVA: And the results are: 100 percent yes; 0 percent no.

MEMBER MARRS: So, the comment on opportunity for improvement or performance gap issues, some of the percents were already talked about, so we talked about there's a huge range of complications from hospital to hospital. I think a key thing to talk about was they did look at controlling for socioeconomic status and race, and there really wasn't a difference hospital to hospital from that standpoint, so they did look into some performance gaps relative to that.

CO-CHAIR KOTTKE: I have nothing to add.

CO-CHAIR GEORGE: Any comments or discussion on the opportunity for improvement? If not, we'll move to a vote.

MS. IBRAGIMOVA: Importance to measure and report 1b performance gap, data demonstrated considerable variation or overall less than optimal performance across providers and/or population groups disparities in care. One, high; two, moderate; three, low; four, insufficient.

(Voting)

MS. VICALE: I just want to ensure, Tom

James, that you've cast your vote.

MEMBER JAMES: I sent that on the chat function.

(Off the record comments)

MS. IBRAGIMOVA: The results are 76 percent high; 24 percent moderate; 0 percent low; 0 percent insufficient.

CO-CHAIR GEORGE: Okay, we'll move on to the quality construct.

MEMBER MARRS: So, relative to the quality construct, this is one of those any or none, whether you're looking at 30 days as well as the 90 days from the Medicare claims data. So, the quality construct felt it was appropriate based on what data was available, and looking at the individual patient level of did you have an event at 30 or 90 days from a complication rate standpoint?

CO-CHAIR KOTTKE: And the composites, or the components are not weighted, which is fine with me. I don't know how you weight a pneumothorax with chest tube versus a hemo pericardium or something.

It's just one point for each.

CO-CHAIR GEORGE: Linda?

MEMBER BRIGGS: I did have one question about the denominator in this composite that we're looking at, because when I looked at the individual ICD-9s that were involved, it included recent indicator pacers, not necessarily defibrillators, and the indicator is supposed to be defibrillators. I was wondering why those codes got included?

DR. MASOUDI: Yes. So, the re-synchronization devices are all --- any first-time with a management device, be it single or dual lead ICD, and CRT are included. The vast preponderance of CRT devices or CRT-D devices, like 95 plus percent, so almost all the CRT devices also include defibrillator functionality.

MEMBER AL-KHATIB: And also, in the context of the NCDR ICD registry, just to build on that, you know, that just captures ICDs. We do not capture CRT-Ds within the ICD registry.

I have a couple of questions about the measure. To go back to Tom's point about the

different weighting, you know, while I completely agree with pneumothorax, tamponade, what have you, I think death certainly carries a different weight. And what --- did you --- I mean, I know it's not easy to kind of decide on what weight to give it compared with the others, but I certainly can see the argument for weighting it differently.

And then in relation to that, how do you handle patients who have more than one complication? Let's say they have, you know, cardiac tamponade and then they end up dying? How do those --- how are those handled?

DR. MASOUDI: Yes. So, to answer the first --- the second question first. It's any or nothing, so if a patient had two complications, they count as having had a complication. With respect to death, there was some discussion about this issue of potential weighting at the time of development. It turns out that, you know, first of all, it would be relatively arbitrary, however it was weighted. Second of all, the rates of death themselves are actually extremely small. And

1 although one could say well, the rates of death are
2 almost, you know, relatively small compared with
3 the other complications, it seemed also awkward not
4 to include at least very short-term mortality in
5 the measure. So, it was decided as a means of sort
6 of including that, and not being arbitrary to
7 include --- it was a relatively low frequency
8 event, but it felt awkward not to include it at all,
9 even though it's quite low frequency.
10 CO-CHAIR KOTTKE: The other side of the
11 coin might be, basically, aversive selection where
12 like the physiologist knows that death is five
13 times as bad as pericardiocentesis and, therefore,
14 doesn't operate on the sickest patients. And I
15 think they have to consider that, too.
16 CO-CHAIR GEORGE: Girard.
17 MEMBER MARTIN: Could you talk a little
18 bit more about the decision on the any or none?
19 Obviously, you chose the any. You probably have
20 data on the number per site, so you probably have
21 both of those pieces of information within the
22 registry. I would imagine if you're at 4 percent,

you could say that's great. I'm right at the national average, but it's 4 percent of the time I have four complications. And one might say you're a little bit worse performing.

DR. MASOUDI: Yes, I don't recall the data specifically. You know, again, the overall complication rate is about 5 percent of any patient. I suppose there could be a very small proportion of patients who experience more than one complication.

I think from a --- sort of a usability and sort of patient-centered point of view, the idea of do you have a complication or not, I think is reasonably important. I don't know that, you know, counting additional complications within a patient would make that much difference, but I don't know for sure.

MEMBER MARTIN: In surgery, or at least in children, is what we've seen as the number of complications go up, the likelihood of death being an outcome goes. So, I ---

CO-CHAIR GEORGE: Any other

considerations, discussion? Okay, we will vote on the construct.

MS. IBRAGIMOVA: Importance to measure and report, 1c composite explicitly articulated and logical, 1c(1) quality construct including components, 1c(2) rationale for distinctive/additive value, 1c(3) aggregation and weighting. One, high; two, moderate; three, low; four, insufficient.

(Voting)

MS. IBRAGIMOVA: The results are: 59 percent high; 41 percent moderate; 0 percent low; 0 percent insufficient.

CO-CHAIR GEORGE: Move on to reliability.

MEMBER MARRS: So, from a reliability standpoint, they're basically matching Medicare claims to the NCDR registry, and so basically describe face validity, or face reliability relative to utilizing claims data from a standpoint --- that standpoint.

And then the one issue is currently

using ICD-9 codes and so subsequently have to switch to ICD-10 codes, and that issue has come up with a number of measures.

CO-CHAIR GEORGE: Any comments on the reliability testing?

MEMBER MARRS: So, from a reliability testing standpoint, they did some analysis looking at the agreement between two RSCRs and each hospital was considered to have slight agreement with this --- with their agreement score that they used when looking at reliability from an across-hospital standpoint.

CO-CHAIR KOTTKE: So, the NCDR has a very rigorous reliability program. It was only tested at the measure score level, and following the algorithm, I get to moderate.

CO-CHAIR GEORGE: Any discussion on reliability?

MR. CHIU: This is Jensen Chiu here. I think somebody asked about the crosswalk. We actually did note --- we did do a crosswalk, ICD-9, ICD-10 using the GEMs crosswalk. We also, I

believe, had some manual --- a clinician looked at it manually just to make sure, kind of sniff test, everything kind of fit, so we did include that, as well as an appendix. It's somewhere buried in here, but we definitely have the ICD-9 to 10 realizing the changeover.

CO-CHAIR GEORGE: Sana.

MEMBER AL-KHATIB: So, I heard you, Fred, saying that here you're not really relying on self-reporting; you're relying on claims data to capture those outcomes. So, then how are you handling the self-reported data of in-hospital complications which are captured within the context of the registry?

DR. MASOUDI: Those aren't used in this measure.

MEMBER AL-KHATIB: Got it; thank you.

CO-CHAIR GEORGE: Any other comments? Liz.

MEMBER DELONG: I am a little confused by the agreement within hospital of 14 percent. Wouldn't you expect that --- the way you tested

this was to randomly divide the data so that you had the data set with all hospitals in it, and you had a second data set with all the --- those same hospitals in it. And if you look at their rankings in the first data set, they're not correlated very well with the rankings in the second data set. Do you have any sense of why that would be?

DR. MASOUDI: Yes, I think some of that is a function of the sample size within the --- the sample size within our testing samples, unfortunately. I think, ultimately, you know, as the sample sizes get more robust, we'll be able to investigate that in bigger detail, but we're limited to some extent by the sample size and then the frequency of the events.

MEMBER DELONG: It isn't as though they may have changed the performance.

CO-CHAIR GEORGE: Any other concerns, discussion? If not, we'll vote on reliability.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2a reliability, including 2a(1) precise

specifications, and 2a(2) testing appropriate method and scope of adequate results. One, high; two, moderate; three, low; four, insufficient.

(Voting)

MS. IBRAGIMOVA: The results are 24 percent high; 71 percent moderate; 6 percent low; 0 percent insufficient.

MEMBER MARRS: All right. So, to comment on the validity specifications, as described before, they're basically looking at claims data with the 30 and 90-day composite there, so the dichotomous variable of did you have any event or not? So, no threats to validity were seen.

And then to move on to validity testing. They did do validity testing and showed when they evaluated chart versus claims, that there was 91.5 agreement with the kappa coefficient of .83, which the developers noted as "almost perfect range."

CO-CHAIR KOTTKE: You --- applying the algorithm, again, the measure would be rated as moderate against the testing that was done just at the data element level.

CO-CHAIR GEORGE: Any comments or questions on validity? Sana.

MEMBER AL-KHATIB: I just have a question about the SDS. Is this a good time to bring it up?

CO-CHAIR KOTTKE: You could. They did look at SDS and race and found really no ---

MEMBER AL-KHATIB: What about gender, because we --- I mean, we have had several studies showing that women are more likely to have complications from ICD implants.

MR. CHIU: So, while Dr. Masoudi is pulling that up, I just have one other really quick point to bring up to circle back to Dr. Martin's point about complications. Actually, these materials, all the specific complications, the rates are listed in the testing document, the second to last page. And just to summarize it, basically, all the complications are about --- you know, there's like 5 percent is for all the complications, and death is only 1.3 percent, so they actually did break out all the elements, as

well, so it isn't just --- I know it kind of came late in the application, but it was buried in there. We did put that in there.

DR. MASOUDI: To answer your question, Sana, gender is one of the factors in the risk model itself, so it's factored into the --- it's factored in specifically the risk model, because as you point out, it's a very --- it's been shown to be a very strong risk factor for complications after ICDs.

CO-CHAIR GEORGE: Other comments, questions? If not, we'll move to voting on the validity.

MS. IBRAGIMOVA: Scientific ---

CO-CHAIR KOTTKE: You better request comment on SDS, and it doesn't seem that this is something that SDS ought to factor in, so I think that it's not appropriate to adjust for a sociodemographic. I mean, in the old days this was students and societies. Some of us never knew it.

CO-CHAIR GEORGE: We'll move on to voting on validity.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2b validity, including 2b(1) specifications consistent with evidence, 2b(2) testing appropriate method and scope of adequate results and threats addressed, 2b(3) exclusions, 2b(4) risk adjustments and stratification, 2b(5) meaningful differences, 2b(6) comparability ultimate specifications, 2b(7) missing data, eMeasures, composites PRO-PMs. One, high; two, moderate; three, low; four, insufficient. We're just missing one vote.

(Voting)

MS. IBRAGIMOVA: The results are 12 percent high; 88 percent moderate; 0 percent low; 0 percent insufficient.

CO-CHAIR GEORGE: Now, I think we move on to the empiric analyses to support the composite.

MEMBER MARRS: So, the data provided was that the empirical analysis demonstrated the individual component measures, that the overall quality construct was currently underway when this

submission happened, and didn't know if there was an update relative to that from the developers.

MR. CHIU: At this time, there's still no update on the empirical analysis piece.

CO-CHAIR KOTTKE: But I didn't think there were anything that raised concerns, so I think it's okay.

CO-CHAIR GEORGE: Any comments on the empiric analysis to support the composite?

MEMBER MITCHELL: Is there or is there not analysis to support this composite? I couldn't hear.

DR. MASOUDI: Well, the questions around that is whether it adds value, does it fit the quality construct, and is it parsimonious? And I think those are yes; there was no mathematical analysis as far as I could tell.

MR. CHIU: Yes, there is no mathematical analysis. It's really just based on the literature, you know, the complications and ---

DR. MASOUDI: Yes, and the extent to which these are important factors for patients who

undergo these procedures.

CO-CHAIR GEORGE: If not, we'll vote on this.

MS. IBRAGIMOVA: Scientific acceptability of measure properties, 2d composite empirical analysis support composite construction and demonstrate 2d(1) component measures fit quality construct, add value, parsimony to extent possible, 2d(2) aggregation and weighting fit quality construct, simplicity to extent possible. One, high; two, moderate; three, low; four, insufficient.

MS. VICALE: Tom, are you able to cast your vote via the chat window, or through text message?

MEMBER JAMES: Just did text; the system went down again.

MS. VICALE: Thanks, Tom. We received it.

MEMBER JAMES: Okay; good.

(Voting)

MS. IBRAGIMOVA: The results are 24

percent high; 71 percent moderate; 0 percent low; 6 percent insufficient.

MS. WILBON: So, I just wanted to clarify because there was no empirical analysis that was provided by the developer. And based on our criteria, it's pass, so it would be really helpful for us, I think, to understand and maybe have a good understanding of what the committee's rationale was for the vote based on the criteria. I know that Tom mentioned a little bit about just general comfort, but can you guys give us a little bit more on how that vote aligns with the criteria in terms of them having not provided any empirical analysis for that element, but still the high moderate rating, I think is the clarification that we'd need.

MEMBER CHO: I mean, for me it was common sense. It's a composite end point where you look at --- I don't need them to show me empiric evidence point by point on all the risk. It's a composite end point detailing the risk of a known procedure. It makes perfect sense to me.

MEMBER AL-KHATIB: And from a clinical perspective, also, I mean it's not just commonsense, but from a clinical perspective they've captured all the important and major complications from this procedure.

MEMBER VIDOVICH: And I would say it's an expensive device in a very high-risk population. I think this is a very important clinical measure for quality and for patients, you know, who are undergoing the procedure, third-party payers, and petitioners, so I think it's a high-quality important measure.

CO-CHAIR KOTTKE: So, the quote in the instructions is, "Empirical analysis should demonstrate that the component measures add value to the composite, and that the aggregation and weighting rules are consistent with quality construct." So, I --- you know, not being a mathematician, I just --- I bought the argument that these are relatively rare events. And if you just look at the single events, it looks like everything is pretty cool. But then you get this

huge deciles from zero to 17 percent, some centers don't do very well at all. And it's sort of non-mathematical analysis that says yes, it makes sense to lump these together to pick up --- increase your signal.

CO-CHAIR GEORGE: Karen.

DR. JOHNSON: So, to point out this is a pretty simple composite. It is any or none composite, so really the empirical analysis that we would have expected to see is really just the frequency distributions of the different components. And my guess is you have those.

CO-CHAIR KOTTKE: That's on page 61.

DR. JOHNSON: If you have those, then I would say you do have empirical analysis. And it goes back to the question --- it really is the question, okay, if you had something in there that was topped out, then the question would be why is it in there if it's topped out? Or the flip, if it's zero -- kind of like death or really close -- why is it in there? And you've already had that discussion, so ---

MR. CHIU: Just to add to Karen's point, we just --- we actually --- so, I know we broke out the complications. We also --- any of the complications, and then death separately, because realizing death is --- we don't want to say it's 5 points more but it's, you know --- we wanted to break it out so you guys could see --- I believe that's on the second to last page of the testing document.

MEMBER DELONG: I'd just like to make the comment that I think we would be concerned about a composite that contained something that was a relatively common complication that kind of lumped in and counted the same way as some of these severe complications, but it seems that that is not the case in this venture.

MEMBER DELONG: Thank you; that's very helpful. Thanks, Karen.

CO-CHAIR GEORGE: So, we'll move on, feasibility.

MEMBER MARRS: So, the measure met the criteria for feasibility from the standpoint of

it's directly being pulled and matched from Medicare claims data, and then linked to the NCPR registry.

CO-CHAIR KOTTKE: Yes; this is theoretically feasible. The problem was that the ACC assumed that they could use the ResDAC data set for this measure, and they were informed that they could not. Fred, do you want to update us on where you are with your negotiations with CMS?

DR. MASOUDI: I'll defer to Lara Slattery from ACC, if I may, who's coming to the microphone there. She give us the very latest information, if that's okay, Dr. Kottke.

MS. SLATTERY: Thanks. Hi, Lara Slattery speaking. It actually wasn't an assumption. Our part with the ResDAC application is we were going through the process of being approved as an organization qualified to receive the CMS data through the ResDAC process. We did identify this is one of the intended uses for receiving the data. When we went back specifically on a project-by-project basis to update ResDAC on the

use, that's when they flagged that the intended use for both quality reporting back to hospitals, and then the intended public reporting component, which we would then pull through because that's part of why the measure was developed, did not fit within their authorization for use.

They have since proposed that we go through the qualified entity mechanism to become qualified to receive the CMS data for these purposes, and we this week have submitted that application. It's rigorous; it mirrors a lot of the ResDAC requirements, but it is different. And we'll have to wait to see how we progress through that mechanism.

And then we are, as we stated, have been tracking language that was initially introduced in the House bill for 21st Century Care, as it did seem to open up a pathway. When that bill was finally approved and the House version of that language disappeared, we're tracking to see if it might get introduced on the Senate side. Our advocacy colleagues are not optimistic that will happen, and

at which point we will have to either see if get approval through qualified entity, or we will have to start talking with CMS about what other avenues might be available. And just to further clarify, it's because the College intends to do this reporting without any type of contract in place with CMS.

CO-CHAIR KOTTKE: Okay. So, my interpretation is that it's --- as a construct it's feasible, even though CMS ---

CO-CHAIR GEORGE: Any discussion on this? If not, we'll vote on feasibility.

MS. IBRAGIMOVA: Feasibility, 3a, data generated during care, 3b, electronic sources, and 3c, data collected can be implemented, eMeasure, feasibility, assessment of data elements and logic. One, high; two, moderate; three, low; four, insufficient.

(Voting)

MS. IBRAGIMOVA: The results are 35 percent high; 65 percent moderate; 0 percent low; 0 percent insufficient.

CO-CHAIR GEORGE: Usability.

MEMBER MARRS: So, the main issue with usability is the same issue that we talked with feasibility, is lacking access to the data currently is kind of the main limitation. So, currently not being used as a measure since there's no access to the data.

CO-CHAIR KOTTKE: Right. But that being said, it would be very useful under certain conditions.

CO-CHAIR GEORGE: Any discussion on usability? If not, we'll move to a vote.

MS. IBRAGIMOVA: Usability and use, 4a, accountability/transparency, used in accountability within three-year, public reporting within six year, or if new, credible plan, 4b improvement progress demonstrated, if new, credible rationale, and 4c the benefits outweigh evidence of unintended negative consequences to patients and populations. One, high; two, moderate; three, low; four, insufficient information.

(Voting)

MS. IBRAGIMOVA: The results are: 35 percent high; 59 percent moderate; 6 percent low; 0 percent insufficient information.

CO-CHAIR GEORGE: Any final comments, or discussion, or questions before we vote on the measure? All right, we'll vote.

MS. IBRAGIMOVA: Overall suitability for endorsement, does the measure meet NQF criteria for endorsement? One, yes; two, no.

(Voting)

MS. IBRAGIMOVA: The results are: 100 percent yes; 0 percent no.

CO-CHAIR GEORGE: I think we're going to break for lunch now and take care of everyone's hypoglycemia. But it will be a very abbreviated lunch; we'll try to convene back at about 20 after. That's only 15 minutes, but you can continue eating while we talk.

MS. VICALE: Thank you, everyone. And the eMeasure review will continue after lunch, with the continuation of 2764 following that. Thank you.

(Whereupon, the above-entitled matter went off the record at 1:04 p.m., and resumed at 1:24 p.m.)

DR. BURSTIN: So, I'm not Jason but just a quick review for the committee and we have talked about this before. We will talk about it again but, very briefly, we have got the definition here of what an eMeasure is and it is in the accepted format of what is called an HQMF, using a specific data model actually developed by NQF called the QDM and then a set of approved value sets. So, that is what we would expect to see as part of an eMeasure.

We do have a team of folks. Jason Goldwater, who was here earlier but had to leave for another meeting oversees our work on eMeasure review. They will, in fact, check to make sure these very technical elements are here.

Just to orient you to what you will see, next slide please. So, we are now considering these separate measures if there is a related measure that comes off another data source and then

one is an eMeasure, we would like people to understand that those are, at times, quite different. And in fact, we know very little about the comparability of an eMeasure compared to, for example, a claims-based measure. We want to make sure there is an opportunity for people to see that.

We do the technical review, as I mentioned, with our internal staff, and there is a set of criteria that are required that are listed here, including testing for reliability and validity in more than one EHR vendor system and that this feasibility logic that assesses whether you can actually find the data elements and the logic is sound. Next.

We also have measures, at time, that come in that are retooled. An eMeasure versus an existing measure is not applicable to this measure. But, as I mentioned, we would consider those separate, and we have provided some guidance that allows for some testing of them using a simulated testing tool developed by HHS called BONNIE. Next.

Okay and lastly, the eMeasure trial

approval. So, the one directly applicable to today is that we very quickly recognized that as much as everybody wants eMeasures, it is really, really hard to find EHRs ready to test new eMeasures. And so we didn't want to hold up innovative measures from getting to market. We didn't want to fully label them as endorsed either, not necessarily assuming that as they hadn't had their full reliability and validity testing, they would be ready, potentially, for all uses, but we wanted to at least allow a pathway in for those innovative eMeasures to come in, where they have otherwise met all the other criteria, with the exception of the fact that they have not yet been able to be tested in an EHR in a practice environment.

So, the idea here is we approve them for trial use, and they then come back to this committee with just the additional testing that is provided to finalize endorsement.

I don't think we need to go through this. Next. Too much. I think we are good.

And Jason just came back in the room. So, if you have any questions, as he's chewing, you can ask him. But I just did the overview. So, thanks.

CO-CHAIR KOTTKE: I have a question. Epic only has about 60 percent of the U.S. market and each Epic system is different. We have, even within -- Health Partners and Park Nicollet Medical Group were one organization, but they are two very different systems.

So, does an eMeasure have to work in every single different vendor's record, or how do you figure that out? And what are the --

MR. GOLDWATER: Good afternoon, everybody. I apologize for running in and out. It has been kind of a nutty day. The qualifications, the criteria for testing an eMeasure is it has to be in more than one EHR. So, theoretically, Epic Inpatient and Epic Outpatient are two separate EHRs, even though they are in the same vendor.

While it would be beneficial to know

that it would work on every EHR, that is not the current requirement. So, one of the things that we are watching because the criteria used to be it had to be in three EHRs, and then it was switched most recently to more than one.

So, what we are watching is when they start testing eMeasures, are they going to test them in just Epic systems because, as you said, they are a large, dominant player in the industry. And where that might pose a problem could very well be in measures such as those for cardiovascular disease, particularly those that are affecting populations that are going to seek care in community health centers or rural health clinics or areas that do not or could not carry an Epic system and carry systems such as NextGen or Netsmart or some of the more second tier vendors.

Right now, the data, it's too premature to see if it would not, but that is something that we are watching. Could it actually work on every EHR? It is too difficult to extrapolate that out because some of these measures would, if they rely

on purely structured data that is contained in every system then, yes, it would work regardless of platform. If it has some unstructured data, that poses some --

MS. VICALE: Before we get started, can I just remind everyone to speak clearly into your microphones, specifically for the records and transcripts. Thank you.

CO-CHAIR KOTTKE: So, developers, please introduce yourself, and give us three or four minutes of introduction.

DR. PUCKREIN: Good afternoon. I'm Gary Puckrein. I'm president of the National Minority Quality Forum. I am joined by Dr. Elizabeth Ofili, who is Director and Senior Associate Dean, Clinical Research Center in Clinical and Translational Research at Morehouse College of Medicine; and Dr. David N. Smith, who is an interventional cardiologist in South Carolina; and Ms. Heidi Bossley, who is President of Bossley Consulting.

The National Minority Quality Forum

appreciates this opportunity to present for your consideration a proposed trial eMeasure to advance the quality of care to patients with chronic heart failure. Our measure, NQF 2764, is entitled Fixed Dose Combination of Hydralazine and Isosorbide Dinitrate Therapy for Self-identified Black or African American Patients with Heart Failure with Left Ventricle Ejection Fractions under 40 on ACEs/ARBs or Beta-Blocker Therapy -- and Beta-Blocker Therapy. Sorry.

Measure 2764 aligns perfectly with the objectives and values espoused by the National Quality Forum and its members and our constituents. The science is sound and the need is great. Our application clearly documents the importance of measuring and reporting the prescribing of the fixed dose combination to the eligible patient population.

The results of the African American Heart Failure Trial published in 2004 demonstrated a significant benefit for African American Heart Failure patients who received the fixed dose.

AHAF was stopped early due to high mortality rates in the placebo population. The fixed dose combination demonstrated a 43 percent reduction in mortality, a 33 percent increase in initial hospitalizations and a 50 percent improvement in patient-reported quality of life.

The importance of this measure is supported by the 2013 ACC/AHA Guidelines for the management of heart failure, which recommends the combination for the treatment of heart failure in blacks with Class 3 and 4 heart failure. The recommendation was made with a Class 1A rating, the highest possible rating.

The need for the measure is amplified by data from peer-reviewed literature, including the application which documents that only a small percentage of eligible patients are receiving the medication contributing to over 7,000 premature deaths a year. It is important to note that the Federal Drug Administration recognizes no substitute or generic for the fixed dose combination. Neither of the component compounds

are indicated for the treatment of heart failure and prescribing them separately as an alternative to the fixed dose is labeled off-label use.

The technical foundation for Measure 2764 is sound through testing of two different EHR vendor products. We have demonstrated that it is feasible to collect the data elements needed to capture the performance measure in current electronic health record systems.

Measure 2764 harmonizes with and does not duplicate or compete with other performance measures. In fact, the subject therapy is adjunctive to therapies in other performance measures that have been approved by NQF.

Measure 2764 is a sound candidate for the trial measure program, testing of validity and reliability is being designed and will be completed within the next 12 to 18 months to enable NQF to endorse this measure.

Measure 2764 represents the best of American medicine. It harmonizes the untapped capacity of current science and knowledge to

provide care that recognizes the heterogeneity of the American patient population. It mitigates the potential harm of one-size-fits-all medicine, and it is a step along the path to precision medicine for all populations in all disease states.

The National Minority Quality Forum is pleased to have this opportunity to present before the standing committee, Dr. Ofili, Dr. Smith. Ms. Bossley and I look forward to the opportunity to respond to your questions.

CO-CHAIR KOTTKE: Thank you very much. Liz, I think you are primary discussant. Oh, Sana, I'm sorry.

MEMBER AL-KHATIB: So, I just have a really burning question. First of all, I really like the focus of this measure but can you help me understand how self-identified black race will be captured through electronic medical record? Because from my experience, that is usually not self-identified. And I am sure you have done some testing. You know I am not the primary reviewer for this measure, but through your testing, how did

you make sure that race was, indeed, self-identified?

DR. PUCKREIN: So, the Joint Commission's criteria for accreditation in the inpatient setting requires the collection of race and ethnic data. The new self-meaningful use guidelines require the collection of race and ethnic data.

So, where we are now is health systems have the capacity to collect race and ethnic data. There isn't any challenge there. And certainly, when we looked at the EHR vendors that we work with, it was possible to identify race and ethnicity.

MEMBER AL-KHATIB: The evidence that the data that they are entering into EHR is data coming from the patient, or is somebody checking the patient in making assumptions about their race?

DR. PUCKREIN: I think the answer to the question is, if I am reading the guidelines correctly, self-identification, I'm assuming that they are asking the patient whether they are African American or not.

DR. OFILI: Okay, in terms of the intake process, but I think you are right, going back, you don't know if someone helped them along. But the intake process does allow you to identify those demographic elements.

CO-CHAIR KOTTKE: Okay, Liz.

MEMBER DELONG: In terms of the evidence, number one, I think we do need to be aware of disparities in care and differential care. And I'm very concerned about unanticipated consequences.

And there was something when I looked up the V-HeFT trial on the web, I found something that -- and I am not an authority on this but some questioned the underpinnings of the trial, including the fact that V-HeFT studied patients in an era predating ACE inhibitors and beta blockers.

Further the self-identified black cohort in V-HeFT had a significantly higher BP, the control of which with ISDN, this combination, may have accounted in large part for the therapeutic benefits seen in AHAF.

And given the lack of -- there doesn't seem to be, and I assume this will follow, a genetic underpinning for this finding, I think it needs more substance than these trials.

DR. OFILI: So, just to respond to the data from V-HeFT. So, the AHAF comparison actually took care of that because everybody who is in AHAF was already, about 97 percent of them were either on ACE inhibitors or ARBs. So, that was the standard of care and that took care of the difference between V-HeFT and AHAF.

In terms of data, we do have data in a limited number of patients that there is better effectiveness of this drug in some genetic polymorphisms. It is just not, you know, everybody didn't get measured within the trial. So, we use that and say, basically, mortality is an important outcome. And based on the standard of care at the time, we know that everybody else who is getting therapy today and not this particular compound that was tested is actually getting sub-optimal therapy.

CO-CHAIR KOTTKE: Mary.

CO-CHAIR GEORGE: I would just add that just for starters that this is a ACC/AHF Class 1 Level A recommendation from their 2013 heart failure guidelines. It is also a Level A recommendation from Heart Failure Society for Classes 3 and 4 heart failure.

Based on four randomized trials that the developers presented with a 33 percent relative reduction in hospitalizations and a 43 percent reduction in mortality in this specific population. The review of the guidelines was covered through April of 2013.

And they also submitted an estimated benefit study that was done by Fonarow that over 6,000 lives could be saved per year. There is a 50 percent improvement in quality of life in this treatment.

That's all I wanted to add for the evidence.

CO-CHAIR KOTTKE: Leslie.

MEMBER CHO: So, I have a question.

The AHAF trial is for Class 3/4 heart failure. This is for what? This is for the -- it just says heart failure. Is it for Class 3 to 4, New York Heart Association? Because that is not --

DR. OFILI: Essentially symptomatic, symptomatic heart failure. And that is the guideline recommendation.

CO-CHAIR KOTTKE: Joel.

MEMBER MARRS: I guess my question is for the issue of fixed dose. So neither HFSA or ACC/AHA guidelines actually specified fixed dose, but this is advocating for fixed dose. And can you comment on the issue that cost relative to generic products versus brand new products?

DR. OFILI: Yes, I think that from the standpoint of what we know about off-label use of medications versus what has been tested, we know that physicians do not have adequate compliance, based on heart failure patients' list of medications that currently exist and the dosing that has been tested actually does not currently exist. So, you, essentially, would have to run a

trial. You are breaking up pills. You have no idea what you are actually offering to the patient. So, even though the guidelines did not specify, physicians who prescribe understand that in order for the patient to get the total dose that was tested, they would have to prescribe fixed dose combination.

CO-CHAIR KOTTKE: Mladen.

MEMBER VIDOVICH: I have two questions because I practice in two different environments. One is an intercity university hospital and the other is an intercity VA.

At the intercity university hospital, I cannot get BiDil because the patient just cannot afford the combination. So, we end up splitting it.

So, my fear is that there would be an unintended consequence of this that while A-HeFT trial is clearly positive, I mean there is no question about this, we may create an unintended consequence. And I just logged into my VA actually remotely to try to order BiDil and I can't order

it because the VA just doesn't have it. Because it would have to be a non-formulary drug request.

So, again, I think there may be some unintended -- I mean by trying to do the right thing, we actually get unintended consequences.

DR. PUCKREIN: Yes, so there are lots of issues here. The first, what you are proposing is that we have off-label use as a performance measure. I mean that is bizarre. On the face of it, it is bizarre.

The VA did a study, did a national coverage study that is fundamentally flawed. If you look at that study, it has contributed to the confusion because the VA, in that study -- you have to go look at it to really understand what was done in that study. I mean it didn't even necessarily deal with African Americans in that study.

So, the problem that you are having with the VA needs to be corrected because they are not following the science. And that is one of the reasons why we have this performance measure. If we are not going to follow the science, and you are

going to sit here and tell me that we can do off-label use for African Americans, that is bizarre, and it is unacceptable, period.

DR. SMITH: Just to add a quick thing to that, the FDA actually did send out a flier saying that there is no bio-equivalent to the fixed dose of isosorbide dinitrate and hydralazine.

There actually were pharmacokinetic studies that show the combination, at the doses used in A-HeFT trial peaked earlier and much higher than any other combination.

And actually, if you were to use the number of pills of generics, one, you will never get the total peak concentration; and, as Dr. Puckrein is saying, you will never see the exact same outcome or efficacy.

CO-CHAIR KOTTKE: So, we are on evidence. And I think this is a very interesting discussion that ought to come a little later. Does anybody disagree with me?

Do you have something that you want to say, Leslie?

MEMBER CHO: My point is is that in soft trial, lisinopril was 40 milligrams and we don't recommend ACE inhibitor at 40 milligram dose for our heart failure measures. We haven't, as an NQF committee, recommended any certain dose, in my understanding. Right? Is there any measure you can think of where we recommend a dose of a medicine?

So, we have given a class of medicine but never a fixed dose -- or never a dose as a mandate.

DR. OFILI: So, I think the issue of class effect with ACE inhibitors actually came later. Because when the trials were done, it was enalapril and that is what we were prescribing for heart failure.

In this particular instance, what you have is, because of the juxtaposition, actually, if you look history of Medicare Part D and when the trial was released, that is how we got into this hole. But I don't think we should perpetuate the problem by not recognizing what has been tested,

what has been proven.

And we know, actually, that because hospitals are not stocking the appropriate therapy, people are not getting treated. If you look at Fonarow's data, it doesn't matter what combination you are talking about, people are not getting treated. So, we have come to NQF almost as a last resort to enable hospitals to do the right thing, to enable medical doctors to do the right thing by their patients. And to me, this is really a pretty -- an issue of life and death for my patients because I have to fight Medicaid in Georgia. Every week I prescribe this drug. They eventually give in, but it is a lot of time, and most doctors just don't have the bandwidth for that. And that is why I think this measure is so critical.

CO-CHAIR KOTTKE: So basically the question on the floor is there evidence that giving BiDil changes outcome. Right? That is the evidence question.

So, are we ready to vote?

MS. MARINELARENA: And just to be clear, you haven't made recommendations on the other measures on dosage because the guidelines didn't have dosages. It was just the class of medication. So, it was based on the evidence.

None of the measures do. So, it is not NQF criteria.

CO-CHAIR KOTTKE: Okay. Ready to vote?

MEMBER CROUCH: It costs $200 a month right now through all the pharmacies. I am just looking online. It costs $200 a month, unless you get some kind of free deal from the company.

DR. SMITH: Actually, all my patients pay $35 a month. There is coverage for it.

CO-CHAIR KOTTKE: But I don't think cost right here -- that is not relevant to the evidence.

MS. IBRAGIMOVA: Importance to measure and report, 1a. Evidence: structure, process, intermediate outcome. One, high, only eligible if QQC submitted; two, moderate; three, low; four,

insufficient.

MS. VICALE: Can Tom James please place his vote as well, through text message or via the chat window. Thank you.

MEMBER JAMES: I sent it through chat window. I can do it again.

MS. VICALE: I think we are having an issue with the chat window. If you could text that, that would be great. Thank you.

MEMBER JAMES: Oh, it's your side now. Okay.

(Laughter.)

MS. VICALE: While we are waiting for that final vote, I did want to mention to any of the public and members that are joining us via phone or through the web, any public member comments, we have a designated time at the end of the day. I believe it is at 4:45 p.m. If we are running early, we will make note of that and allow sufficient time for public member comment. And if any comments are communicated via the chat window, we will announce those comments at that time. Just so everyone is

aware of when that public member comment is available. And that is for all the measures being reviewed for Phase 3 today. We don't have an individual public member comment after each measure review. Thank you.

MS. IBRAGIMOVA: And the results are 35 percent high; 59 percent moderate; 6 percent low; zero percent insufficient.

CO-CHAIR KOTTKE: Okay, opportunity for improvement.

MEMBER DELONG: Well, there was evidence of gaps in care but they didn't seem to be specific to this measure. I quote, much of the disparity can be assigned to modifiable risk factors, such as uncontrolled hypertension and suboptimal healthcare. Therefore, when African Americans are treated according to guidelines, discrepant outcomes can be minimized.

But it is not demonstrated specifically that it is a gap in this particular measure, although, you have spoken to a gap in this particular measure.

DR. OFILI: Right. So if I may, the data by Fonarow was pretty -- he specifically said we need to figure out a way to get heart failure patients in the hospital on this appropriate therapy because by the time they leave the hospital, the gap widens.

There was a gap that was identified in the Fonarow papers specific to what we are trying to address.

CO-CHAIR GEORGE: The data that I think Dr. Ofili is referring to was a review from Get With the Guidelines that found 7.3 percent of the African American heart failure population were on that, compared to an estimated 27 percent that should have been. And I was a little unclear about why only 27 percent should have been. Maybe you can explain that.

DR. PUCKREIN: Yes, I think they were looking at the eligible population. So, there is roughly about 500,000 African Americans with heart failure. And I think by their calculation, this is Fonarow's calculation, about 27 percent of that

500,000 should be eligible for therapy.

If you actually look in the Medicare data, and we have actually done a lot of work and are about to publish an article, in the Medicare, it is about two percent of African Americans with heart failure who are eligible are on the therapy. I mean this is a really serious problem. This is not borderline. This is a large number of patients not getting access to appropriate therapy.

CO-CHAIR KOTTKE: Michael -- Sana.

MEMBER AL-KHATIB: Just to get further clarification from you, so those other people weren't eligible because they had diastolic heart failure, because their Neural Cardiac Association class was not in the group that they needed to be? I mean those are the reasons that -- okay.

DR. PUCKREIN: Yes, Class 1 and 2.

MEMBER AL-KHATIB: Thank you.

CO-CHAIR KOTTKE: Any other discussion about opportunity for improvement? Seeing none, let's vote.

MS. IBRAGIMOVA: Importance to measure

and report, 1b. Performance Gap: data demonstrated considerable variation, or overall less-than-optimal performance across providers and/or population groups; disparities in care. One, high; two, moderate; three, low; four, insufficient.

And the results are 35 percent high; 59 percent moderate; 6 percent low; zero percent insufficient.

CO-CHAIR KOTTKE: Reliability.

MEMBER DELONG: I'm sorry. I didn't think we were doing reliability and validity on this one.

CO-CHAIR KOTTKE: Oh, did it --

DR. BURSTIN: So, essentially since it is an eMeasure for trial use, you can't speak to reliability and validity. But scientific acceptability includes things like precision of specifications. So, we would ask that you just vote on that. Have I got that right, Jason, Karen?

MR. GOLDWATER: Yes, that's correct.

CO-CHAIR KOTTKE: Mary, do you want to

do specification?

CO-CHAIR GEORGE: So, all the data elements were defined with the VSAC registered value sets specified in the HQMF format using the QDM. Sorry for using all the acronyms. But anyway, it satisfied all of those requirements that NQF has for the specifications.

Do I do the testing plan now or --

CO-CHAIR KOTTKE: Sure, why not?

CO-CHAIR GEORGE: So, they did submit a testing plan that appears to comply with all the required testing that they will do during the trial period. This is intended for an outpatient population and hospital acute care population. They did do BONNIE testing with 100 percent pass rate and it covered 85 percent of the data elements, if I interpreted that correctly.

Threats to validity?

CO-CHAIR KOTTKE: Should we let Sana chime in?

MEMBER AL-KHATIB: Yes, just a couple of questions about the specifications, especially

in terms of the exclusions that you used. So, what if a patient can't tolerate them because of other side effects? I don't see those listed. And we may also have patients who may tolerate a lower dose than the fixed dose that you proposed. So, how do we count those people?

And then you had here severe lupus. How are you defining the severity of lupus?

DR. OFILI: So from the standpoint of the actual dosing, there is a guideline for physicians. So, even though it is called fixed dose, I mean there is one tablet, half a tablet, a quarter tablet that they start patients with and depending on the blood pressure.

So, there is a clinical sort of education that we provide doctors around how to start it. And most physicians get comfortable with that, once they follow those guidelines. And we have tested that in some of our Get With the Guideline Initiatives.

The other question you had about severe lupus, the reason that is there is really because

in the -- what is that book called -- the PDF, the book, yes, that has there lupus recorded in the past when hydralazine was used in megadoses. And in fact, we never saw this in any of the trials, but that kind of follows, and so severity allows individuals to know when not to use it. But in fact, it is not drug-induced at all in any of our heart failure trials.

CO-CHAIR KOTTKE: So, a question. If somebody has a code for lupus, do you use BiDil?

DR. OFILI: Right now, if they have heart failure and -- just carrying a diagnosis of lupus does not mean we don't use BiDil and that is because it is not drug-induced lupus. So, this is in the setting of drug-induced that it would cause -- not the traditional autoimmune -- I'm sorry. Go ahead.

MS. BOSSLEY: No, and to get to your question about whether it is in the EHR as severe lupus or not, we looked at the codes. ICD-10 does not include severity in most of their codes.

So, in there, you will see there in the

specifications there is a SNOMED code that looks at severity. Testing will tell us whether or not we can get that piece of information or not. Is it, indeed, a true severe lupus code that we are pulling? That is still to be determined but that is part of the testing.

CO-CHAIR GEORGE: All of their exclusions were either contraindications or warnings in the FDA Guidance.

MEMBER CHO: What about if a patient couldn't afford it? So, like you prescribed it.

DR. OFILI: Now, this is actually similar to other drugs that are out there. Doctors have access to, based on the patient's level of affordability and there is a very strong, at least in my practice, and I know that is happening around the country where individuals can get access to the drug, based on who cannot afford it. But the practice has to request that.

MEMBER CHO: No, no. What I am asking is is the measure. You prescribe it; they don't fill it because they can't afford it. Is it still

in your denominator?

MS. BOSSLEY: Yes, the group specifically discussed whether they wanted to include that as, in essence, an exception where a patient wasn't on the drug, and we noticed a documentation that they couldn't afford it. But given the importance of the issue, they really didn't want to include that as something that would be removed from the denominator. So, it is -- that would be included as a ding or not qualifying.

MEMBER CHO: But why is that a ding for the physicians? They did prescribe it.

MS. BOSSLEY: I'm sorry. So, it will depend -- I'm sorry. I misspoke. You are right. If a patient didn't receive it, then it wouldn't. I'm sorry I've got myself confused.

MEMBER CHO: If I prescribe it, they won't fill it.

MS. BOSSLEY: Then you would get credit for it.

MEMBER CHO: Then I will get credit for it.

MS. BOSSLEY: If it is on the medication list as a prescription. I'm sorry.

MEMBER CHO: So, if I don't prescribe it because a patient says you know what, don't even prescribe it because I can't afford it, that is a ding against me.

MS. BOSSLEY: Depending on how they document that. Correct.

MEMBER CHO: Well, let's say that is how we document it.

MS. BOSSLEY: That is correct.

MEMBER CHO: So, then it will be a ding against the physician or the hospital.

CO-CHAIR KOTTKE: But that is like all the other measures. It is all the non-SDS adjusted measures, the rest of them there, the cost of therapy or lack of insurance enters into the picture.

It is time to vote on reliability. They just told me. Specifications. Sorry.

MS. IBRAGIMOVA: So, eMeasure approval for trial use. Measure Specifications: 2b.1

specifications consistent with evidence must pass. One, high; two, moderate; three, low; four, insufficient.

The results are 18 percent high; 71 percent moderate; 12 percent low; zero percent insufficient.

CO-CHAIR KOTTKE: Feasibility.

MR. GOLDWATER: Right. So, reliability and validity we can't because we haven't actually tested the measure yet, only in a synthetic way. So, the next thing would be feasibility.

CO-CHAIR KOTTKE: Okay, feasibility.

MEMBER DELONG: I can't really find it addressed except that they don't have the data yet to test it.

CO-CHAIR KOTTKE: Mary?

CO-CHAIR GEORGE: They did provide some feasibility scorecard from one inpatient and one outpatient EHR from the same vendor, which we heard is acceptable. The scorecard measured data availability, data accuracy, data standards,

workflow, and total data element feasibility score. And each data element scored at the highest level, with the exception of ejection fraction, which scored the intermediate category, and their testing will continue. I think they made some note that there may be some issues with ejection fraction not being recorded in the EHR.

MEMBER DELONG: And I actually wondered, I mean I would think that would be pretty important.

CO-CHAIR GEORGE: And so it basically met all the requirements for a trial measure on feasibility.

DR. OFILI: That issue is an issue that we are overcoming across all of heart failure. So, I think this is not unique. Right?

CO-CHAIR KOTTKE: I would agree. I'm really shocked at the number of, I mean going around the country, I mean, you call people up and they ask them what, you know, what proportion of your patients code with heart failure have an EF less than 40 percent, and they say we don't because we

don't record it. It is shocking. But I don't think that should stand in the way of this measurement. It is one of those name and shame things.

So, I think we are -- any further discussion on feasibility? Time to vote?

MEMBER BRIGGS: I guess we need to know when we can talk about this issue of people being able to get the drug. I mean just because we prescribe it, doesn't mean they are actually going to fill it or that they get the medication. You know if you can't get it at the VA, that is a problem. If you can't get it through the hospital where you are working, that is a problem, too. And you know some of you have said yes, my patients can get it for $35, but that, again, that requires a lot of work on the part of both the patient and the provider to get to the drug company and fill out the proper paperwork to get that to happen.

So, while I think this is a very worthy thing to do, a particular drug at a particular dose from a particular drug company is kind of pushing

the edge of what we should be looking in terms of quality measures.

CO-CHAIR KOTTKE: So, you are suggesting that because it is hard to do, even though there is strong science, we shouldn't endorse it?

MEMBER BRIGGS: No, I'm saying that if a patient can't afford it, and it can't happen, you know, the issue is more with the insurance companies and so forth actually covering this drug. It is not with people wanting to do the right thing. Providers want to give this drug. Now you don't want us to give it off label, so that means that we are basically being asked to prescribe a particular drug from one particular drug company. And that is not what we are used to doing and it is not something that -- you know, there should be a generic for this drug, or we need to cover it universally, if you are going to measure people on an outcome related to this.

CO-CHAIR KOTTKE: That doesn't stop oncologists from prescribing neoplastics.

Mladen.

MEMBER VIDOVICH: I was wondering, could you get some objective information about formulary status around the country among major hospital systems? Could you get information about formulary status? How is this drug available to, let's say, large hospital systems, integrated hospital systems? Is it available for providers?

Because as you mentioned, I will just take it as an example, as an interventionalist, I frequently make a decision between a bare metal or drug-eluting stent whether a patient has insurance or not. So, whether I want to put a drug-eluting stent or not, I may not do it just because the patient may not have means to afford it. And then if I were to get dinged by my percentage of how I place stents, we may not be okay because they make a clinical decision prior to even implanting the device or writing the prescription.

So maybe if we were able to get objective data on formulary availability of the drug, that would be helpful.

DR. OFILI: So with all due respect, I think that when drugs are approved to treat a certain condition based on the scientific data behind it, we go along with that. Generics don't come online until several years, and patients continue to get the therapy. I just think we should apply the same standard here. When I talk to my colleagues who are physicians, and I ask them, I said are you making an assumption that this patient cannot afford the drug, or have you prescribed it and they didn't fill it, just like you do other prescriptions, many physicians stop and think about why they were making that assumption that the patient could not afford the drug.

So, all we are asking is should it be the chicken or the egg? In this case, if hospital formularies don't stock it, and it is a performance measure, what do you think is going to happen?

CO-CHAIR KOTTKE: You are not quite up, Liz.

MEMBER DELONG: It won't stand up.

CO-CHAIR KOTTKE: Oh, I'm sorry, that's you. I'm sorry. I couldn't read from all the screens. Okay, go ahead, Liz.

MEMBER DELONG: So, I think you have made a very good point that if NQF were to endorse this measure, it might force the issue. But I'm concerned about two things, one of which is that maybe this is not a quality measure for the physicians who are prescribing it. Maybe it should be a quality measure for the insurance or some other entity.

My second concern is this is an expensive drug and we are talking about giving it to, in a lot of cases, people who can't afford a lot of drugs, and this is on top of a lot of other drugs. And are they going to pick -- are people who can't afford it going to pick and choose and maybe suffer unintended consequences because they won't take maybe some of their other drugs that are more important?

MR. GOLDWATER: Can I interject here for just a second? I'm sorry to do this but the

conversation is trending away. I think we need to sort of recalibrate this a bit.

You are not reviewing a measure for endorsement. That is not the point of this exercise. The point of the exercise is you are reviewing a measure to be put into a trial use program. And a measure can be put into the program for a period not to exceed three years, in which you will collect data and be able to tell, at that point in time, whether what you are thinking actually comes to pass.

If this were a measure that you were looking at for endorsement, you are absolutely correct in making these assumptions but that is not what you are doing. What you are doing is do we put this measure in the field and let it be tested to see if what our assumptions are are true or accurate and may come to pass?

So, please keep that in mind. This is not to be an endorsed measure. It is a measure to be accepted into a very specific program.

CO-CHAIR KOTTKE: Also, this

conversation is reminding me a bit in tobacco cessation counseling where the doc said well, the guy had a blue shirt so I interpreted that as meaning he wasn't interested in quitting smoking.

I don't think we belong in a position of second-guessing the patient. And I don't disagree that maybe when this comes back it also ought to be a health plan measure of is this on formula? And what is it the PCSK9s, $14,600 a year. I mean talk about expensive. Or for an antineoplastic, a course of $189,000. I mean this is -- Sana.

MEMBER AL-KHATIB: Just one quick comment because I heard several people making a reference to the fact that, you know, the physicians are actually prescribing it, but we do not have data that the physicians are prescribing it on a regular basis. And if you reflect on our experiences with other evidence-based medications, if I were to guess, I would say that actually a good number of physicians are not prescribing it, not because they don't want to do

their best. Because they are very, very busy, and they are having to address so many different things when they see the patients. So, having a performance measure like this, I think will push them to think about this medication like it did for ACE inhibitors, and ARBs and beta blockers.

CO-CHAIR KOTTKE: Ellen.

MEMBER HILLEGAS: Just one quick comment. I think it is great. Jason, you really clarified a lot. So, thank you very much. It put it better in my mind. But the key thing I also want to say is we have argued in this room before about measures where they come back in three years, and they haven't collected data. And one of the frustrating things is sitting on this panel and hearing that you haven't collected data, but you want it renewed.

So, I think that we should have a policy that if this is going to be going forward, and we are using it as a test, that we don't pass it unless we have the data next time. I mean we pass it now if we want, but in three years, if it comes back

without data, we don't pass it at all because we have seen a bunch of different measures come without any data, even though they have been out there for three years. That is just from my --

MS. MARINELARENA: So just to clarify. When it comes back in three years, it has to be submitted for endorsement like every other measure. So, it has to have the full testing data and everything. So, it will go through the full cycle in three years.

MR. GOLDWATER: Right. Right. They have to have enough data to be able to justify reliability, validity, feasibility, all of it. If they don't get that data, and that could very well happen, I mean don't discount the fact that the measure could go into the field, and the testing fails. That is the point of a trial use program. So, that is a possibility. But they cannot submit the measure for endorsement until enough data is collected in order for this committee to make an informed judgment about the potential endorsement of a measure.

CO-CHAIR KOTTKE: What I would like to do is take Leslie's question. Gerard has the next measure and he has to leave at 3:00. So, I would like to sort of move this along. I think we have visited this --

MEMBER CHO: So, is the intention of the measure to -- of this measure, 2764, to see how many African Americans or self-identified blacks are prescribed the BiDil? In three years, we will have that number, whatever that number is, 20 percent, let's say, and then we will then endorse measuring the fixed dose combination. I mean is this a fishing expedition for us to find out the low rates of fixed combination of hydral and isordil that is currently being prescribed? Because if that is the expedition, we already know the answer.

CO-CHAIR KOTTKE: The question is can they get the data. Does the measure work in an EHR?

MS. WILBON: Right, so what they are going to be testing is the use of the actual measure. So, through using the measure, they may

have better information about how often it is being prescribed. But the purpose of the testing and the three-year trial period is for them to implement the measure in various EHRs, see how the data is collected. Are they able to find the data elements in the EHR? Does it collect all of the information, the data, that they intended to collect in order -- and when they get that data, they can run the reliability and validity testing. And based on that, they will come back to the committee.

So, having analyzed that data, they may have some of the information that you are describing but that is not the intent of the three-year trial period.

CO-CHAIR KOTTKE: Okay, Nick, and then I would like to call the question on feasibility.

MEMBER RUGGIERO: The one question that I have is even in my own practice, where you can't get it for your patient, you prescribe two separate agents. So, I wonder what percentage of people we're actually going to miss who are being

prescribed separate agents because of the fact that we have been conditioned to not be able to prescribe the fixed dose pill. So, they may be on the appropriate therapy, but because of the fact that you can't get a fixed dose pill, you are actually getting two separate medications. Is there any way to correlate for that and see what -- because people are actually getting therapy. That may not be with the fixed dose pill.

DR. OFILI: Well, so if you look at the data, the Get With the Guidelines data, the good news for all of us is the American Heart Association is consistently collecting data on the Get With the Guidelines, and they will collect any form of the combination that you prescribe. So, you are not going to lose that prescription. In fact, we are hoping a rising tide would lift all boats, if you know what I mean.

CO-CHAIR KOTTKE: Okay, Leslie.

MS. VICALE: Just before we go on, I know Tom James had a question come in over the chat box. Tom, do you still want to ask that question?

MEMBER JAMES: I think that has been already addressed, as far as coverage.

MS. VICALE: Okay, great. Thank you so much.

CO-CHAIR KOTTKE: Okay, let's vote on feasibility.

MS. IBRAGIMOVA: Feasibility: 3a data generated during care, 3b electronic sources, and 3c data collection can be implemented; eMeasure feasibility, assessment of data elements, and logic.

One, high; two, moderate; three, low; four, insufficient.

And the results are 6 percent high; 82 percent moderate; 12 percent low; zero percent insufficient.

CO-CHAIR KOTTKE: So usability and use -- usability.

MEMBER DELONG: I think we have discussed usability.

CO-CHAIR KOTTKE: Okay. Mary, do you have anything additional?

CO-CHAIR GEORGE: Yes, I would just add it is used, as we mentioned, in Get With the Guidelines for QI and benchmarking.

CO-CHAIR KOTTKE: Okay, can we vote on usability?

MEMBER PHILIPPIDES: First of all, this has been a great conversation, by the way. I really have liked this conversation. I think it is an important one. But just for clarification, if I cobble together hydralazine nitrate, it is not in the fixed combination but in the old fashioned ones, and treat it that way, that would not meet the specification. Is that correct? That would be a ding.

DR. PUCKREIN: That's correct.

MEMBER PHILIPPIDES: Okay.

DR. PUCKREIN: I would point out and advise there is a conversation about the fixed dose -- I'm sorry -- about using the two drugs separately that occurred in 1993 at a FDA hearing, when the fixed dose was then being offered to treat patients who could not tolerate an ACE or an ARB. And the

FDA conclusion there I think is appropriate here, which is, there is no science that allows us to predict what will happen if you provide a patient with the two drugs separately. You cannot say that you can define what the prognosis of that patient is going to be.

And that is why we use AHAF. Because AHAF is a randomized clinical trial study against standard of care, ACEs and ARBs plus the injunctive therapy. That is why we don't use the two drugs separately, because it is not science. You can't predict. You cannot tell your patient or any patient when I give you this medication, I am going to lower your mortality rates; I am going to improve the quality of your life; I am going to slow up your hospital admissions. There is no science to say that.

MEMBER PHILIPPIDES: Okay, thank you.

CO-CHAIR KOTTKE: Okay, time to vote.

MS. IBRAGIMOVA: Usability and Use: 4a. Accountability/Transparency used in accountability within three year, public reporting

within six years or if in a credible plan. 4b. Improvement. Progress demonstrated if a credible rationale and 4c., benefits outweigh evidence of unintended consequences to patients and populations.

One, high; two, moderate; three, low; four, insufficient information.

The results are 12 high; 53 percent moderate; 35 percent low; zero percent insufficient information.

CO-CHAIR KOTTKE: So, overall, we'll vote.

MS. IBRAGIMOVA: Overall suitability for eMeasure approval for trial use. Does the measure meet NQF criteria for eMeasure approval for trial use? One, yes; two, no.

The results are 82 percent yes; 18 percent no.

CO-CHAIR KOTTKE: It says it is time for lunch.

(Laughter.)

CO-CHAIR GEORGE: We will be moving on

to Measure 0730, AMI Mortality Rate. Developers, and then our discussants are George and Gerard.

Go ahead for just a few minutes.

DR. ROMANO: Hello. My name is Patrick Romano. I am a general internist on faculty at UC Davis School of Medicine in Sacramento, California. And I am here representing AHRQ, the Agency for Healthcare Research and Quality on behalf of the contract team that is responsible for enhancing the measures and performing the analytic work that is represented in our NQF submission.

I think joining me on the phone are a couple of other members of our team. From Stanford University, from Truven Health Analytics, and from AHRQ.

I will just say, to open the discussion, that this is a measure that has been endorsed previously, so it is up for maintenance of endorsement. It is a measure of risk-adjusted in-hospital mortality for patients with acute myocardial infarction. It is a measure that is

predominantly used by hospitals and health systems, as well as state health data organizations that don't have access to linked data on post-discharge mortality. So, it is an alternative to the CMS-Yale measure of 30-day risk-adjusted or risk-standardized mortality for AMI patients.

So, there is extensive experience with its use and feasibility, and I will just leave it for questions later.

MEMBER PHILIPPIDES: So, just trying to move forward quickly, everything that you mentioned in regard to this is an outcome measure, well-established process of care that showed that this does impact on performance and outcomes.

As far as opportunity for improvement -- actually, should I stop there?

CO-CHAIR GEORGE: Review the evidence.

MEMBER PHILIPPIDES: The evidence is strong as far as processes and past data. It is strong.

MEMBER MARTIN: Agreed. And the

processes that they included all support the --

CO-CHAIR GEORGE: Any discussion on the evidence for this measure? We will vote on the evidence.

MS. IBRAGIMOVA: Importance to measure and report, 1a. Evidence: health outcome or PRO, rationale supports the relationship of the health outcome or PRO to at least one healthcare structure, process, intervention or service. One, yes; two, no.

The results are 100 percent yes, zero percent no.

CO-CHAIR GEORGE: Move on to the opportunity for improvement.

MEMBER PHILIPPIDES: The developers cite several large data bases of healthcare cost minimalization and HCUP databases on inpatient performance, both of which show significant performance gaps and, I think, an opportunity for improvement.

In regards to disparities -- do I bring it up now -- they do look at certain patient and

system characteristics in the HCUP database that they cite, including age, they expect it to go up; gender, which seems to be associated with an increased rate in mortality; zip codes in low income areas, large central metropolitan hospitals, and Medicare payers.

So, they do, in fact, cite some areas where there are disparities of care but after we read these, I never know what to do with that. I never know how to move on, based on what we have heard before but there is some evidence of disparities outcome. Let me leave it at that.

MEMBER MARTIN: Nothing to add.

CO-CHAIR GEORGE: Any comments on the performance gap? If not, we will move to a vote.

MS. IBRAGIMOVA: Importance to measure and report, 1b. Performance Gap: data demonstrated considerable variation or overall less than optimal performance across providers and/or population groups, disparities in care. One, high; two, moderate; three, low; four, insufficient.

MS. VICALE: I think we might be missing one more vote. Did everyone vote?

MEMBER JAMES: Yes, I did.

(Laughter.)

MS. IBRAGIMOVA: And the results are 82 percent high; 18 percent moderate; zero percent low; zero percent insufficient.

CO-CHAIR GEORGE: Specifications and reliability testing.

MEMBER PHILIPPIDES: We are all reasonably in order with this metric. As far as reliability, it was tested at the performance measure level. They looked at over 2,000 hospitals in a hospital network and measured signal to noise ratio. They weighted it such that hospitals with fewer discharges had a lower weight than those with larger hospitals, except for the two or three sort of lowest discharging hospitals. The S to N ratio was quite good and overall, was 0.75, which we are told is good.

As I meandered through the algorithm, as I tripped through it, it seemed like the

reliability rating was still high, best I could tell. So, I have no problem with the reliability.

MEMBER MARTIN: I thought reliability and validity were both strong.

CO-CHAIR GEORGE: Any discussion on reliability? It looks like everybody is ready to vote.

MS. IBRAGIMOVA: Scientific Acceptability of Measure Properties: 2a reliability, including 2a1, precise specifications, and 2a2 testing appropriate method and scope of the adequate results. One, high; two, moderate; three, low; four, insufficient.

The results are 82 percent high; 18 percent moderate; zero percent low; zero percent insufficient.

MEMBER PHILIPPIDES: In regards to validity, the specifications do align with the evidence. This measure was tested at the data element level and the performance measure score level. They noted test samples, where they looked to make sure that the diagnosis of AMI was accurate

and a large Canadian database viewing ICD-9 codes and they had good results there. They also looked at -- they said the fact that the measure has been in use for ten years without any issues in regards to the risk factors that were put in the risk model, there were, I think, 23 that were evaluated formally. We came up with a C statistic, based on over 44,000 discharges and there was a good C statistic of 0.8867. So, it seems like, empirically, the measure is valid.

They looked to see if there were any exclusions that might have muddied the waters and led to a threat to the validity. And the exclusion rate when you take out missing data, I believe, was what really dinged it. It didn't ding it much at all, 0.01 percent. Exclusion is quite low. So, it seems as though the data is whole. It is consistent and valid the best I could tell.

Gerard, any thoughts?

MEMBER MARTIN: Agreed.

CO-CHAIR GEORGE: I just had a question on the risk adjustment, given that what you noted

with the SDS, was there any testing in risk adjustment?

MEMBER PHILIPPIDES: There was no testing, specifically, that I could tell but I will push back to the developer.

DR. ROMANO: Yes, we included a literature review, basically indicating that a substantial portion, if not most of the observed disparities on the other sociodemographic characteristics, such as race, ethnicity, and income, appear to be driven by differences in access to care and by utilization of certain services in the hospital, including early intervention with PCI for patients with STEMI.

So, for that reason, AHRQ opted not to include those additional sociodemographic factors.

So, age and gender are included in the model and transfer status of patients transferring in from other hospitals, those are included in the model. But race, income, ethnicity are not.

CO-CHAIR KOTTKE: So, is this one of

the measures that should be calculated both ways, adjusted and unadjusted?

DR. ROMANO: Well, AHRQ staff may be able to address this but we were given the option of making an argument based on prior evidence and we chose to take advantage of that option, because we think the evidence is pretty clear.

MS. WILBON: Yes, so just to recap, if you recall when Karen presented earlier, so what Patrick is describing is their conceptual analysis basically, that based on their review of the literature that they didn't -- while there is a link, it is not direct and there are other factors mitigating it. So, they didn't feel that there was a need to do additional empirical analysis and their conceptual analysis of the variables, at this point, don't warrant inclusion in the risk model.

So, we don't -- if the developer finds that there is no conceptual link and the committee is onboard with that, as long as they are able to provide a rationale for that and the committee is okay with that, then we won't require them to do

any additional analysis.

DR. ROMANO: Right. The idea is not to mask disparities in outcomes that are actually driven by disparities in processes that contribute to outcomes because those processes are at least partially under the control of the healthcare system.

MS. WILBON: It could be a consideration for the committee if they felt that there would be a need to stratify the results of the measure, based on some of these factors, which would allow you to still see the differences among, and the results among the population but it wouldn't be actually baked into the risk adjustment model. So, if that is something that the committee feels would be helpful or necessary to be able to see what differences there were among the different populations, that could be something you could consider.

CO-CHAIR GEORGE: Any comments about validity or stratification? George.

MEMBER PHILIPPIDES: If you were to

consider that kind of stratification, when would that be considered?

CO-CHAIR GEORGE: Now. Yes, it would --

MEMBER PHILIPPIDES: I would like to consider --

CO-CHAIR GEORGE: Okay.

MEMBER PHILIPPIDES: I think sometimes we make assumptions that these issues don't impact on care, that it is all the hospital systems. And I'm not sure that is the case. I'm not sure that it's not the case. But if we are about to have a robust database presented by good developers, why not take a look at that?

I mean, for example, educational level. I can find, I think, a valid way of saying that that might in fact impact care. I'm not sure if that is related to the health system in place. So, is there a compelling argument not to ask for it to be stratified as well?

CO-CHAIR KOTTKE: If I can just make a comment that having looked at this in Finland

along, I mean the Finns have no barriers to care on cost or anything else but they seem persistent and in face they are widening gaps in outcomes by SES, even though they have a very narrow band compared to us. It just appears that lower SES, less educated people, are less able to manipulate the health system to their own ends. And so it may be an argument for stratification.

CO-CHAIR GEORGE: I mean there is also the consideration but by stratifying, we have the information and we know where we need to work harder with performance improvement as well.

DR. ROMANO: I would just say two things. One is that I think the data that you are referring to are data looking at longer term outcomes after cardiac events. And those findings are less applicable to studies where the outcome is in-hospital mortality.

Now, I will say now actually Agarwal and colleagues had a very nice paper in the *Journal of American Heart Association* 2014, in which they explored this issue, using median household income

of the residential zip code as a proxy for socioeconomic status. And they did find a 1.11 times a higher mortality in the lowest SES quartile but there was also 0.80 reduced timely reperfusion. It was exactly the process factor that appeared to be explaining much of that difference.

So but stratification is always an option with the AHRQ quality indicators and we are certainly open to further discussion.

CO-CHAIR GEORGE: Do we just note that we discussed this but we don't vote on that?

MS. WILBON: Yes, so, obviously, you are going to have to vote on the measure as-is but we can work with the developer after the vote is done and what have you to come back with the measure with the specifications, with the stratification included in this as part of the specification. So but we would like you to kind of vote on the measure as you see it in front of you.

CO-CHAIR GEORGE: Any other discussions on the validity? If not, we will vote.

MS. IBRAGIMOVA: Scientific

Acceptability of Measure Properties: 2b. Validity, including 2b1, specifications consistent with evidence; 2b2, testing appropriate method and scope with adequate results and threats addressed; 2b3, exclusions; 2b4, risk adjustments/stratification; 2b5, meaningful defenses; 2b6, comparability, multiple specifications; 2b7 missing data, eMeasures, composites, PRO-PMs.

One, high; two, moderate; three, low; four, insufficient.

DR. ROMANO: While you are voting, Pam Owens from AHRQ staff reminds me that AHRQ does look at disparities by race/ethnicity for this measure in the National Healthcare Disparities Report, which is available online.

MS. IBRAGIMOVA: We are missing one vote, if everyone can just try again. Oh, okay.

The results are 63 percent high; 38 percent moderate; zero percent low; zero percent insufficient.

CO-CHAIR GEORGE: We will move on to

feasibility.

MEMBER PHILIPPIDES: All the measure elements are available in electronic records via administrative claims data. It has been used for over ten years, wide experience with the AHRQ software. I couldn't find any major concerns as far as the feasibility of it.

CO-CHAIR GEORGE: Comments or discussion on feasibility? If not, we will vote.

MS. IBRAGIMOVA: Feasibility: 3a, data generated during care; 3b, electronic sources; and 3c, data collection can be implemented, eMeasure feasibility, assessment of data elements and logic. One, high; two, moderate; three, low; four, insufficient.

And the results are 94 percent high; six percent moderate; and zero percent low; zero percent insufficient.

CO-CHAIR GEORGE: Move on to usability.

MEMBER PHILIPPIDES: This was first released, I think, in 2003. It has been broadly

used, if I am not mistaken, in many quality programs. I believe it is publicly recorded and I couldn't think of any major, nor could the developers cite any major unintended consequences. So, I think it is useable. Gerard?

CO-CHAIR GEORGE: Any discussion on usability? We will vote.

MS. IBRAGIMOVA: Usability and Use: 4a, Accountability/Transparency - used in accountability within three year public reporting, within six year or if in a credible plan; 4b, Improvement - progress demonstrated in a credible rationale; and 4c, benefits outweigh evidence of unintended negative consequences to patients and populations.

One, high; two, moderate; three, low; four, insufficient.

And the results are 100 percent high, zero percent moderate, zero percent low, zero percent insufficient.

CO-CHAIR GEORGE: Any points of discussion before we vote up or down on the measure?

All right, we will vote on the measure.

MS. IBRAGIMOVA: Overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? One, yes; two, no.

Gerard, did you vote?

(Laughter.)

MS. VICALE: We just want to make sure everyone has voted. We are missing one vote.

MS. IBRAGIMOVA: The results are 100 percent yes; zero percent no.

CO-CHAIR KOTTKE: Okay, 2396. American College of Cardiology. Discussants are Michael Crouch and Nick Ruggiero.

Dr. Masoudi, would you give us a brief intro?

DR. MASOUDI: Thank you. I'm back here with Traci Connolly, staff from the ACC, and Kristen McCoy, also staff from the ACC. This is measure 2396, Evaluation of Vital Status and NIH Stroke Scale and Follow-up after Carotid Artery Stenting.

This is a process measure characterizes

the systematic assessment of outcomes after percutaneous carotid revascularization. It involves the documentation of vital status and, among patients who are alive, measurement of stroke symptoms or severity, based on the NIH Stroke Scale.

And I would note the NIH Stroke Scale is implemented in several measures that are used within the joint commission stroke certification program and is widely acknowledged as a means of standardizing the assessment of patients with symptoms of stroke.

It is a hospital-level assessment that includes, again, the ascertaining of vital status and stroke scale amongst living patients within 21 to 60 days following carotid revascularization via stenting. These data that are the data for this measure are derived from the NCDR care registry and have been used in care quality reports to facilities. And as you can see, there is fairly market variation in performance in this measure. And although it is a process measure, it helps

address one of the substantial barriers to understanding meaningful neurologic outcomes of patients after carotid stenting, mainly, the ascertainment of outcomes using a standardized approach after the procedure is performed.

CO-CHAIR KOTTKE: Thank you. Michael, are you first?

MEMBER CROUCH: Yes. We may have an opportunity to catch up on time here, I am hope. Nick and I agree that there are serious problems with the measure, from our perspective.

The evidence addresses the two-year restenosis rate after carotid artery stenting, two years after the stenting procedure and the follow-up period of this process measure is 21 to 60 days. That doesn't seem to be directly applicable to the -- the evidence doesn't seem to be directly applicable to the interval of the follow-up measure, unless I am misreading it.

I'm not aware of any reason for exempting this from the usual evidence criteria. It just doesn't seem acceptable to me to hold

providers accountable for measure performance without empiric evidence of benefit, including performance.

The specific things that were to be measured are is the person dead or alive at 21 to 60 days and, if they are alive, have they had a stroke scale performed by a qualified examiner certified by the American Stroke Association. It is not clear how that person's American Stroke Association approvals -- certification status was to be ascertained. That wasn't available in electronic medical record.

Overall, it just seems not like a very solid measure and not very well tied to evidence.

CO-CHAIR KOTTKE: Nick.

MEMBER RUGGIERO: So, I think the idea is good. The problem is is that the data they use to support it is pretty much consensus evidence and some expert opinion also. Some of the studies which were actually presented like EVA-3S, SPACE trial, a lot of it is data that were actually sort of disproved a lot of times because of the fact that

the way the trials were carried out.

So, I think the idea is good but I do not think that the data that was presented support the measure that they are trying to support here.

CO-CHAIR KOTTKE: Fred.

DR. MASOUDI: Can I just respond to a few things? It is a little bit of a tricky measure because it really is involving the ascertainment of an outcome. I think that we could all agree that understanding an outcome is valuable to patients.

And in this case, it is difficult to really understand what those outcomes are beyond mortality. Mortality is relatively easy to ascertain but in terms of neurologic impairment following revascularization, it is hard to understand that if it is not being performed in a standardized manner, it is a little bit -- you know the whole issue about evidence supporting this is a little bit difficult to -- is a little difficult to provide in terms of is this, are these outcomes that are important to patients or are they not? I would argue they are.

The 30-day time frame, the evidence was provided yes, a two-year time frame. I don't know that a patient really cares that much in terms of what time point their outcomes are ascertained. Certainly, this helps tie the ascertainment of the outcomes to the proceduralist who did it, in that it is a relatively short time frame and a reasonable time frame for follow-up.

And the NIH Stroke Scale is something that can be -- certification in NIH Stroke Scale is something that can be obtained free and at relatively little burden.

MEMBER RUGGIERO: The question I have is was it ever thought about the measure being proposed for any carotid intervention, rather than just carotid stenting, such that you have an even playing field with endarterectomy versus stenting?

DR. MASOUDI: Yes. It is funny you should ask because, indeed, the first time this measure was proposed, it was proposed both after carotid stenting and carotid endarterectomy and we were asked to re-propose as following carotid

stenting only.

CO-CHAIR KOTTKE: Mary, did you have words to say?

CO-CHAIR GEORGE: I would just echo the comment that the NIH Stroke Scale score is a major of the outcome after the procedure. And to say a person didn't have a stroke, did have a stroke, doesn't quantify so much as do they have something with an NIH Stroke Scale score of 2 or was it 20. So, it does provide some further information on that outcome.

DR. MASOUDI: One other comment I would make on the 30-day time frame is that it does also permit some perspective on outcomes related to those that have been obtained in the clinical trials, at least in the short-term. And that is important because it has been widely demonstrated in the literature the outcomes in clinical trials in experienced centers, particularly with respect to carotid revascularization, are markedly different than those in the typical center. That has been shown very clearly with the carotid

revascularization data in a variety of different context.

MEMBER CROUCH: Is that true for 30 days, the 30-day follow-up period?

DR. MASOUDI: Yes, the follow-up periods in those studies, I believe vary. But again, I think the bottom line is that from a patient's perspective, I think they care about being ascertained and ascertained in a way that is meaningful. And this is a meaningful way to assess neurologic outcomes after carotid revascularization.

And many of those outcomes, particularly those that were related to the procedure themselves, will occur in the shorter term period.

MEMBER CROUCH: Was there data on 30 days at all?

DR. MASOUDI: What's that?

MEMBER CROUCH: I didn't see any data on the 30-day interval outcomes at all. I have no way of knowing whether we could expect a

significant difference at 30 days.

DR. MASOUDI: Well, we have data on the extent to which these outcomes are ascertained at 30 days, presumably by the center that is performing the procedure. And there is obviously this tension. If you want to look at two-year outcomes as an example, ascertaining two-year outcomes by the proceduralist is going to be more difficult.

This allows the ascertainment of outcomes in a standardized fashion in the site that the procedure was performed after the procedure.

And again, it is hard for us to know what the outcomes are at 30 days if people aren't actually systematically collecting them.

MEMBER DELONG: Could we clarify --

CO-CHAIR KOTTKE: Fred, aren't we just asking did somebody actually check up on the patient?

DR. MASOUDI: That is exactly what we are saying. Are they document if they are alive or dead; and, if they are alive, what is their

neurologic status using a granular, widely-accepted, freely-available form of ascertainment.

MEMBER DELONG: So, we are only assessing here the performance measure of whether they evaluated the patient, not the outcome of the Stroke Scale.

DR. MASOUDI: That is correct. This is, as I said, a process measure of whether or not they ascertained the patient, not what the outcomes were.

And it is an important distinction but you actually can't get to that second step if we don't have the first step.

CO-CHAIR KOTTKE: And is -- I missed this. I had a neural event.

The amount of the gap, I mean the number, the proportion of patients who do not have any assessment?

DR. MASOUDI: You can see the data that are provided here show probably one of the most marked discrepancies in performance that I have

personally seen in measures ranging from zero to 100 percent with quartiles that are all across the range.

CO-CHAIR KOTTKE: Okay. Any other -- Leslie, did you want to?

MEMBER CHO: So, why? I guess the confusion of this measure is that you are measuring whether they are alive and a Stroke Scale. But at the same time, there is all this stuff about revascularization and the carotid stenting. I think that is what is sort of confusing.

DR. MASOUDI: Right, so point well taken. But our denominator are the patients within a registry who are getting an invasive neurologic procedure and the process measure is whether or not they get follow-up in the short-term in the 21 to 60 days following that carotid revascularization procedure.

MEMBER CHO: I mean I agree with you that somebody should be measuring them, checking them. I totally agree.

But I think it is the lingo in here about

the revascularization rate and the CEA versus CA. Do you know what I am saying? I think that is what is what's confusing and I think that is what, I think, initially I also had no idea what this was talking about.

DR. MASOUDI: Okay. Well, I certainly apologize for any confusion around that and hope the discussion will clarify it without taking too much of the group's time.

But I think it is now clear now what we are talking about in terms of this being a process measurement that is the documentation of the ascertainment of an outcome, not the outcome of itself, which is certainly the springboard for understanding outcomes at 30 days.

CO-CHAIR KOTTKE: And the reason you chose not to make it an outcome measure?

DR. MASOUDI: Well because of the marked variability in ascertainment of the outcomes. As you can see, again, a lot of sites are reporting it not at all. And so it is very difficult to be reporting outcomes on a level

playing field if at first you are not setting the stage for measuring them in a systematic way.

CO-CHAIR KOTTKE: Thanks.

MEMBER DELONG: Given the vagueness of the evidence, is there evidence that -- and this is a clinical question -- if you were to do the follow-up within 30 days, it would make a difference in patient outcomes?

DR. MASOUDI: Well again, it is really impossible -- that is a tautology, almost, in the sense that if we don't know what the outcomes are, we can't say whether or not doing something improves outcomes.

And that is why this measure, I admit, is really kind of tricky because it is a process measure that reflects the extent to which you document outcomes, which makes it conceptually a little more challenging, I agree. But it also makes it kind of, again, tautological to say well, is this related to better outcomes? Well, we can't really know, particularly in those sites that really aren't measuring outcomes at all. It is

impossible to know.

CO-CHAIR KOTTKE: So, I don't know if NQF has a stroke rehabilitation measure but it is theoretically possible that if somebody is identified as having had a completed neurological event after a carotid procedure that referral to rehabilitation may improve the outcome. That may be one reason to check on them. It is just like had a myocardial infarction, refer them to cardiac rehab so they can mend.

I think it is, unless I -- yes.

MEMBER BRIGGS: This question is for the developers. And that is did you have any data related to the reportability by tertiary centers? There may be a number of places like say Cleveland Clinic or even here in D.C. where people are coming from a distance to a certain provider for carotid stenting, and then leave the area again, not to necessarily see that provider or that facility again. And that 20-day window or whatever would not catch those people where they would be lost to follow-up.

DR. MASOUDI: We were looking into the data in terms of facility type. I don't know. We have a breakdown in the data tables regarding specific facility types.

I think that those sorts of things are challenges with any measure that requires longitudinal follow-up and you can see that that is the case here.

MEMBER BRIGGS: The big thing with the stroke is that you can't do it on the phone. You have to see the person, in order to make that evaluation.

DR. MASOUDI: Yes.

MEMBER BRIGGS: So, does that create a hardship for certain facilities that are doing a lot of tertiary kind of referral types of things?

DR. MASOUDI: Yes, I can't -- again, we can look at the data tables to see how things broke down by center type. I know that -- I mean I can only speak for my site and knowing that we perform procedures like this that we tend to follow-up our patients directly and we serve a fairly large

catchment area but that is anecdotal evidence.

CO-CHAIR KOTTKE: So, anything else about evidence? Seeing nobody moving -- oh, yes.

MEMBER HILLEGAS: I have one question. So, we are looking at between 21 and 60 days. So, there is strong evidence that says that it happens within that time frame? If there is going to be a stroke, it happens between 21 and 60 days.

DR. MASOUDI: No. So, the 30-day follow-up, again, provides some comparability with say the clinical trials.

MEMBER HILLEGAS: Okay.

DR. MASOUDI: But we think it would be too stringent to say to a provider you must document this on day 30. It is just not feasible. And so there is a window of time that is permitted to allow for a post-procedural assessment that is both feasible and one that conforms to that that would provide some comparability with outcomes from trials.

MEMBER HILLEGAS: Okay and with this stenting, do they all have medications the same or

similar medications post-op? Do they all have like calcium channel blockers to prevent spasm or they don't do medications? I'm not -- so this is just the procedure. It doesn't matter if they have other things.

DR. MASOUDI: It doesn't. This is just the index, the denominator are people who undergo this procedure, independent of whatever medical therapy their physician or their clinician has decided to use in their case.

MEMBER HILLEGAS: Okay, thank you.

CO-CHAIR KOTTKE: So, I think the question for evidence is if you have done a procedure on somebody's carotid, are you obligated to check to see if they are still alive or functional within 60 days? Okay.

Yes, Fred.

DR. MASOUDI: One brief comment. I will say there was a -- there is, in the data table here, a comparison of rural versus suburban, versus urban hospitals, just for your reference. And there is not statistically significant difference

in the proportion of people meeting the quartiles of performance.

CO-CHAIR KOTTKE: Okay, can we vote? Let's vote on evidence.

MS. IBRAGIMOVA: Importance to Measure and Report: 1a, evidence structure process, intermediate outcome. One, high, only eligible if QQC submitted; two, moderate; three, low; four, insufficient.

Just missing one vote. We got one more. We've got one more committee member. Someone walked out?

Okay, so the results are zero percent high; 56 percent moderate; 38 percent low; 6 percent insufficient.

MEMBER DELONG: In the interest of time, and we are losing it rapidly, do all of these definitions have to be read at each vote?

CO-CHAIR KOTTKE: For the record so that the transcriptionist knows what we are voting on. And I don't think that is what is eating up the time.

Okay in the interest of time, opportunity for improvement. Stop? Gray zone.

Okay, opportunity for improvement. Mike.

MEMBER CROUCH: Performance gap now or what are we doing?

CO-CHAIR KOTTKE: Opportunities for improvement. Sure, if that is performance gap.

MEMBER CROUCH: Oh, right. They already cited a performance range of zero to 100 percent that would appear to warrant a national performance measure. That is the only criteria they met.

I did not see evidence of significant disparities in care. I don't think it should be indicated disparities sensitive.

CO-CHAIR KOTTKE: Thank you, Nick?

MEMBER RUGGIERO: I think there is a great performance gap that can be measured based upon the wide variety of reporting that you actually have. So, I agree with what Mike said.

CO-CHAIR KOTTKE: Any burning --

Linda? Okay, Ellen, did you want to say something? Naming and shaming. Ready to vote?

MS. IBRAGIMOVA: Okay, the Importance to Measure and Report: 1b. Performance Gap. One, high; two, moderate; three, low; four, insufficient.

Tom James, if you could cast your vote. Thank you. The results are 50 percent high; 38 percent moderate; 6 percent low; 6 percent insufficient.

CO-CHAIR KOTTKE: Reliability.

MEMBER CROUCH: I thought there were some issues with the clarity of the numerator and the denominator. Age wasn't specified, although somewhere else in the application, it was. It wasn't in that.

The method of ascertaining the examiner certification has been mentioned as being feasible. So, I will let that go.

It was mentioned that the literature supports the person who does the neurological exam shouldn't be the person who did the procedure. I

didn't see that included in the specifications. And then the other thing was that if there is someone who dies prior to 21 days and that is documented prior to the 21 days, do they get credit for having done something prior to the time frame starting?

CO-CHAIR KOTTKE: Fred.

DR. MASOUDI: Yes, so if it is not in the specifications, it should be. Again, as Lara Slattery mentioned, it is an independent ascertainment.

If a patient dies before 21 days and it is documented that they have died at whatever time frame during this period, they are counted as having satisfied the measure.

CO-CHAIR KOTTKE: Nick.

MEMBER RUGGIERO: The only question that I have and it is sort of counterintuitive is the fact that the data collection tool appears to allow patient reasons. For example, if you don't have follow-up on a patient and it is not included, which can steer your data.

So, that would be the only thing concerning me is that if you have a -- if we are talking about if the patients refused, or are unavailable, or what have you, then those patients are not included, which may or not have had a stroke which would change your data. So, that is the only thing that you can't control for.

DR. MASOUDI: Though it is a pretty standard exclusionary approach and is one that is always sort of a two-edge sword in that respect.

CO-CHAIR KOTTKE: Okay, any other questions? Are you ready to vote?

MS. IBRAGIMOVA: Scientific Acceptability of Measure Properties: 2a. Reliability. One, high; two, moderate; three, low; four, insufficient.

The results are 6 percent high, 88 percent moderate, 6 percent low; and zero percent insufficient.

CO-CHAIR KOTTKE: Validity.

MEMBER CROUCH: Reliability -- we're on reliability now, right?

CO-CHAIR KOTTKE: We just did reliability.

MEMBER CROUCH: Oh, validity already. Okay.

The non-CS, carotid artery stenting operator being a certified examiner was an inconsistency there. I think the time frame is reasonable around the 30 day, the 21 to 60. They are based on -- face and content validity arguments are based on expert opinion. It is not clear to me that they support the validity as a quality indicator, although it keeps coming back to we need to get data to see if there is a quality problem.

So, I guess I was reading it a little bit too critically at first.

There is no analysis of the effects of the exclusions that I saw. So, under threats to validity, the lack of analysis and the effects of exclusions left me a little unclear on the effects of that.

And overall, it is not real clear to me that it measures meaningful differences in quality

of care but if it is just survival and whether that exam was done or not, then maybe my criticisms aren't as valid.

CO-CHAIR KOTTKE: Nick?

MEMBER RUGGIERO: I agree with Mike. I think we were probably looking at it in the beginning and it was clarified, that if it is just a matter of it being a test if performed; a test is not performed, I mean I think it is fine in that respect.

DR. MASOUDI: Just one quick point. There is some assessment of exclusions analysis on page 12 of the document that I believe was submitted to the staff.

CO-CHAIR KOTTKE: Any discussion? Seeing no discussion, let's vote on validity.

MS. IBRAGIMOVA: Scientific Acceptability of Measure Properties: 2b. Validity. One, high; two, moderate; three, low; four, insufficient.

The results are 6 percent high; 65 percent moderate; 24 percent low; 6 percent

insufficient.

CO-CHAIR KOTTKE: Feasibility.

MEMBER CROUCH: The elements are available in electronic form. I wasn't clear that the examiners' or the patients' status was but someone else commented that that was really attainable. So, they appear feasible.

CO-CHAIR KOTTKE: Nick?

MEMBER RUGGIERO: I think it is just based upon -- it is site reporting. So, it is going to be based upon the data that is entered by the given sites.

And the other question I have is can you perform -- I don't know the answer to this, can you do carotid stenting at your institution without submitting your data to the NCDR registry?

DR. MASOUDI: Yes.

MEMBER RUGGIERO: So, that is the only -- you are going to miss that but if there is no --

DR. MASOUDI: Although it doesn't, per se, require registry participation to do this and

that one could independently assess patients who had carotid stenting and determine whether or not they have this follow-up independent of registry participation.

CO-CHAIR KOTTKE: Do you know what proportion was participating, Fred? Do you know what proportion participated in NCDR?

DR. MASOUDI: I'm not sure right off the top of my head for care what the participation rate is. It is probably sub-50, greater than 20 but that is just off the top of my head.

CO-CHAIR KOTTKE: Further discussion around feasibility? Okay, let's vote.

MS. IBRAGIMOVA: Feasibility. One, high; two, moderate; three, low; four, insufficient.

The results are 18 percent high, 65 percent moderate, 18 percent low, zero percent insufficient.

CO-CHAIR KOTTKE: Usability and use.

MEMBER CROUCH: The measure is not currently publicly reported. It is not clear

whether it will be used by the NCDR. Benefits are unclear but that has already addressed.

They identified no unintended consequences. I couldn't think of any.

That is pretty much that.

CO-CHAIR KOTTKE: Fred.

DR. MASOUDI: I'm sorry, just for clarification, it is used in the quality reports for the care registries. So, it is not publicly reported.

MS. SLATTERY: Just to clarify, it is currently reported back to the sites.

And also to clarify, if a site is participating in the registry and chooses to submit any carotid stenting patients, they must submit all carotid stenting patients. So, they don't selectively within the site get to choose which ones are submitting to us or not.

The ACC Board has decided that in order for us to be able to consider measures currently for our public reporting effort, they first must be NQF endorsed.

So by virtue of the fact that we are submitting this for NQF endorsement, it means that we are viewing this toward public reporting. But at this point in time, we can't state that because it doesn't have the NQF endorsement.

CO-CHAIR KOTTKE: Sure. How do you know that somebody is reporting all their cases?

MS. SLATTERY: So, we do have an audit program that is in place and it is a challenge but we try to do similar validation strategies like we do with like half PCI registry and others to get at that.

CO-CHAIR KOTTKE: Okay, thank you.

MS. SLATTERY: That is one of the challenges.

And also just to add in that we came up regarding the carotid endarterectomy, so there is a -- this measure also exists for the carotid endarterectomy population. It is not appropriate for this group to consider because that is a surgical procedure. So, we do plan to take that forward also under an NQF surgical group for

endorsement consideration.

CO-CHAIR KOTTKE: So, any further discussion on usability and use? Yes, Leslie.

MEMBER CHO: There are other carotid registries out there, the Society for Vascular Surgery. So, how is that going to compete with this? Do you have to buy the NCDR?

DR. MASOUDI: No. Again, I think to qualify for this measure, one would have to identify the patients that are getting carotid stenting and document whether or not they have follow-up. So, our testing data is all based on the NCDR data but there is no reason why -- I mean to me it is not really a question of competition. It is something that can be measured in a variety of different ways. But all the testing that has been performed, provided here, is performed within the NCDR Care Registry.

But there is no reason why a site couldn't, in theory, identify their carotid stenting patients and document whether or not they satisfied this measure, independent of their

participation with the NCDR Care Registry, the SES Registry, or even potentially no registry.

CO-CHAIR KOTTKE: Okay, vote on usability and use.

MS. IBRAGIMOVA: Usability and use. One, high; two, moderate; three, low; four, insufficient information.

The results are 12 percent high, 76 percent moderate, 12 percent low, zero percent insufficient information.

CO-CHAIR KOTTKE: Okay, time to vote up or down.

MS. IBRAGIMOVA: Overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? One, yes; two, no.

The results are 76 percent yes, 24 percent no.

CO-CHAIR KOTTKE: Measure 0965.

MS. VICALE: Thank you, everyone. At this time, we are going to go ahead and take our break originally scheduled for 3:00 p.m. At this time it is 3:15 and we will take a 15-minute break

and come back to the room at 3:30.

(Whereupon, the above-entitled matter went off the record at 3:15 p.m. and resumed at 3:30 p.m.)

CO-CHAIR GEORGE: So, we want to welcome Judd Hollander. If you could, just give us your conflict of interest and disclosures.

MEMBER HOLLANDER: Yes, I have no conflicts with any of the measures.

CO-CHAIR GEORGE: Thank you. So, we will move along to the measures. A few minutes with the developers.

DR. MASOUDI: This is measure 0965, discharge medications, including ACE inhibitors or ARBs and beta blockers in eligible patients receiving implantable cardioverter defibrillators at discharge.

This is a previously endorsed measure. It is an all or nothing composite process measure that has not been changed since prior endorsement.

It is a hospital-level measure that assesses the use of guideline-based medications

for patients who have received an implantable cardioverter defibrillator. As I said, it is an all or nothing measure of opportunities for each patient, measuring those who have left ventricular systolic dysfunction and receive an ACE inhibitor and those who have either left ventricular systolic dysfunction or myocardial infarction and receive a beta blocker.

The use of both of these medications in these specified populations is strongly supported by guideline recommendations; both the heart failure guidelines, as well as the secondary prevention ACC/AHA guidelines and the data that were used to derive this measure come from the NCDR ICD Registry.

CO-CHAIR KOTTKE: Joe or John.

MEMBER CLEVELAND: You bet. I will go ahead and take the lead on this. So, as you have heard, this is -- I think the only other to add is a facility-level analysis for this measure's facility. It is a resubmission.

As we have heard, the evidence is, I

think, quite strong. I'll cut to the chase since we are running a little low on time but these are all guideline support measures. The guidelines direct that you have to.

Specific recommendations with Class 1a recommendations for these medications. Again, two different populations. Again, two different populations.

And then just to get everybody thinking again, because I had to rethink the composite measure talk that Gary helped qualify a lot of this this morning, it has two component measures in this composite.

So, that is it. Judd?

MEMBER HOLLANDER: I will start off with a big bang for the first measure is I have trouble understanding this. I only have one issue with this measure and it is up-front, that none of the evidence that is cited in here has anything to do with ICDs. It has to do with heart failure and MIs. And all the evidence is about heart failure and MIs and these medications. And there are

measures that directly apply to heart failure and MIs.

So, I will take an example, and I could be wrong because I am not a cardiologist, but take some weird structural cardiovascular disease where someone has no MI and normal LV function and has dysrhythmias, would they need an ACE inhibitor? Would that fit? But they may have an AICD placed anyway.

And so I can't find the incremental benefit of this measure over other measures and I know that is not really what I am supposed to do here.

But then I could say well, looking at the evidence, there is no evidence that ICDs, as a whole, in the absence of heart failure or AMI benefit from these agents.

And so I would say that there is no evidence to demonstrate it should be used.

DR. MASOUDI: Well, so these medications are restricted specifically to those patients who would qualify them by the heart

failure prevention guidelines.

So, only the patients with systolic dysfunction would qualify for an ACE inhibitor, absent contraindications. Only those patients with systolic dysfunction or prior MI would qualify for a beta blocker absent contraindications.

So, the evidence from the secondary prevention and heart failure guidelines are directly applicable to the denominators -- for what would be the denominator of the population.

The reason we think this is important is that this is not an unusual procedure. There are hundreds of thousands of defibrillators that are implanted every year in the United States, most of which occur in patients with structural heart disease, either left ventricular systolic dysfunction or ischemic heart disease.

And as you can see from the data on performance, by virtue of this measure, this is situation where it seems like in a lot of cases, patients are coming in; they are having a very expensive procedure and no attention is being paid

to the medical therapy, the guideline-based medical therapy that they should be getting.

And gain, this is occurring in hundreds of thousands of patients every year.

MEMBER HOLLANDER: So, let me just push back a little and say why is that not picked up by the other measures that directly say everybody with heart failure or a systolic ejection fraction below 40 in an MI should be on these medications?

So, what we are doing is we are using the ICD as a proxy to identify patients that are already identified in other measures, in reality.

DR. MASOUDI: Yes, I mean again, all I can say is we can illustrate a gap in treatment in these patient who are coming in for a $30,000 device and there is no attention being paid to the medications that they are being given at the time of discharge.

And you know that is -- it is a slice of the heart failure pie but it is one where, again, we are taking patients, we are doing something very expensive to them and not necessarily paying any

attention to the sort of bread and butter medicines that these patients should be getting.

And again, we are talking about a denominator of hundreds of thousands of patients a year in the United States.

MEMBER HOLLANDER: So, let me just ask one follow-up question because I can understand that logic. Is there data to support that once you have an ICD-9 you benefit -- an ICD-9 -- you saw where my mind is -- that you benefit from these agents in that population?

DR. MASOUDI: So, the medical therapy studies were done to a varying degree in patients who had defibrillators and some of the very earliest were not done in patients with defibrillators. That is true.

However, the guideline recommendations are quite clear that patients with LV systolic dysfunction, independent of whether or not they have a defibrillator should be getting ACE inhibitors and beta blockers in the absence of a contraindication. And then similarly, with the

beta blockers after MI.

So, although there is not a prospective study in an ICD patient population only, the guideline recommendations speak very strongly to this as a meaningful process of care.

CO-CHAIR GEORGE: Discussion? Liz and then Sana.

MEMBER DELONG: Well follow-up to one of my previous comments. We have a proliferation of measures. And if this measure is measuring the same thing that we are measuring in the heart failure population in general, do we need another measure to measure the same thing?

DR. MASOUDI: Well, I obviously defer to the panel about that. That is not mine to answer. Again, I would say we have identified substantial gaps in a large population of patients. Sort of what is a teachable moment, I think, for practitioners in a large population of patients who are, again, getting an expensive technology.

So, it is certainly different than saying ambulatory heart failure performance

measure and one can decide whether or not that seems to be meritorious or not.

MEMBER AL-KHATIB: So, I want to echo what Fred said regarding the guideline documents. They clearly say that patients have to be on optimal medical therapy. What we actually face in clinical practice is we get asked to implant ICDs in patients who were just diagnosed with heart failure, who are not on any optimal medical therapy. This is definitely a gap in care that I think this performance measure will definitely address.

Remember, this performance measure has been around. This is just up for renewal.

The last thing I would say, to go back to your comment, Judd. So, yes, there hasn't been a randomized clinical trial of ICD recipients being randomized to medications versus not because the pivotal randomized clinical trials of ICDs require that patients be on optimal medical therapy.

So, all of the evidence that we have on ICDs is actually in this study of patients being

on optimal medical therapy.

CO-CHAIR GEORGE: I would just like to remind people that the staff are putting up their cards. I think we have got --

MEMBER CLEVELAND: I think I would add one other thing, too. In my practice as, obviously, a cardiac surgeon, seeing advanced heart failure, you know the advanced, advanced heart failure -- transplants, VADs, we have a very cohesive heart failure system but people still work their way into, if you will heart failure ICDs from different perspectives.

So, I think this offers a chance to do a cross-cut where, again, you may miss the isolated heart failure measure population because they have got a distinct clinic from where they are going into the EP route.

So, I think it is just another chance to catch these and it will get to, we'll talk about in a little bit, if there are substantial gaps.

So, it really does cut across facilities and potentially health systems.

CO-CHAIR GEORGE: Michael.

MEMBER CROUCH: This is up for renewal. Is that correct? How long has it been implemented and what evidence is there that it is has had any effect on improving care so far?

DR. MASOUDI: Yes, so it has been, I think it was endorsed three years ago, if I am not mistaken. And there are data from two sequential periods that would suggest two things, one of which is that there have been increases in rates of the composite, on the one hand. But on the other hand, there are residual gaps in care. So, I can't say there is a causal relationship between the two but we have identified increases.

MEMBER CROUCH: Can you be more specific about how much it has improved and how big the gap is, still?

DR. MASOUDI: Yes, sure, I would be happy to. With deference to the yellow flag holder, I will --

MS. MARINELARENA: Can we move on and continue the conversation about evidence and then

talk about that in gap?

DR. MASOUDI: Yes.

MEMBER CLEVELAND: I've got that in here. I'll get to that in our next segment.

MS. VICALE: I do have a question from Tom. Tom, are you on the line? Would you like to ask?

MEMBER JAMES: Yes. I guess I am somewhat confused. This sounds like it is a two-part measure. One is is there consistent criteria for implantation of the ICD. And from the discussion, it sounds like people have to be under optimal medical management before an ICD is implanted; although, I have never seen that in my own practice.

But then the second part is if there is consistent population of ICD-implanted patients, then are they getting these drugs?

So, I am just confused by this being what seems to be requiring two elements before it could be approved.

DR. MASOUDI: No, it doesn't -- well,

I may not be understanding you correctly. But the denominator is patients who had an ICD implanted. So, that is the denominator. So, it is not really a part of it, per se, as is the composite of whether or not they got either a beta blocker or an ACE inhibitor, should those be indicated by guideline indications.

MEMBER JAMES: The title "eligible" just suggested to me that there was a set of criteria for implantation but that is not in fact the case.

DR. MASOUDI: Yes, that is not the case. When it says eligible, it pertains to the eligibility for the medications, not for the ICD.

MEMBER JAMES: Okay, thank you.

DR. MASOUDI: The fact that they got an ICD is consistent.

MEMBER JAMES: Thanks.

CO-CHAIR KOTTKE: I just want to make a comment about should we, for a re-endorsement, require that there has been improvement? George Isham and I and Nico Pronk wrote a paper in

preventing chronic disease about five years ago asking the question: what does it take? And we would say that this is just one of five factors of mutually agreed upon goals. And you need public reporting, you need resources, you need alignment of incentives, imperatives and sanctions, and you need leadership.

So, I don't think we ought to put on the table that we are not going to endorse if people haven't shown improvement. It would be really nice if they did, but just having a measure alone, you can't say that would result in improvement.

CO-CHAIR GEORGE: Mladen?

MEMBER VIDOVICH: I just have a quick -- maybe not splitting hairs. But there are many other reasons why ICDs are placed and other requirements that an ICD is placed such as assessment of ejection fraction, and -- you know I am not an electrophysiologist but other than beta blockers and meds, maybe should that be included in the measure? Should the measure have a little bit more elements than just beta blockers?

Because I see this also in clinical practice. An ICD gets put in, somebody hasn't been on a trial of meds. Maybe the EF wasn't assessed appropriately.

Should we maybe refine this measure a little bit more than just meds? That would be just my question. There is more to putting an ICD in than just being on meds.

DR. MASOUDI: Yes, I would agree with that but this is looking at, again, a couple of medications across a few indications. So, it is really trying to get to this issue of optimal medical therapy.

MEMBER VIDOVICH: I disagree. I mean we see this clinically all the time; this is a good step to measure.

CO-CHAIR GEORGE: Any other comments that we haven't discussed on the evidence? If not, we will vote.

MS. IBRAGIMOVA: Importance to Measure and Report: 1a, evidence in structure, process, and intermediate outcome.

One, high, only eligible if QQC submitted; two, moderate; three, low; four, insufficient.

The results are: 35 percent high; 53 percent moderate; 12 percent low.

CO-CHAIR GEORGE: Opportunity for improvement.

MEMBER CLEVELAND: So, I think I have got these numbers and, Fred, make sure I am quoting these correctly. But the developer provided analysis from the NCDR ICD registry during the period of 2011-2012 and subsequently 2013-14.

So, in the 2011-12 cohort, there are 243,000 patients in 1,552 hospitals. So the mean compliance rate, if you will, or meeting this composite measure is 74 percent. The 50th percentile was 76 percent. The subsequent '13-'14 analysis showed a small improvement from 74 percent, if you will, to 78 percent. Now, there were fewer patients but more hospitals in that analysis. And the 50th percentile was still, again, at 76 percent. A wide standard deviation

of 16 and 17 percent are present.

So, I think based upon the very wide standard deviation gaps there, definitely a performance gap exists.

There were no racial or gender disparities, I should say, present in this measure.

CO-CHAIR GEORGE: Any comments on opportunity for improvement? If not, we will vote.

MS. IBRAGIMOVA: Importance to Measure and Report: 1b. Performance Gap. One, high; two, moderate; three, low; four, insufficient.

The results are: 71 percent high; 29 percent moderate.

CO-CHAIR GEORGE: Quality construct.

MEMBER CLEVELAND: Again, this composite measure has two components. The proportion of patients that are going on ICD who received prescriptions for either beta blockers or ACE/ARBs for which they are eligible. It is an all or none composite measure.

I thought, personally, that the

components both had support for them together. It is logical and the rationale does have additive value to consider this as a composite measure, they are getting both, and say that you have to have both or you don't satisfy the measure.

CO-CHAIR GEORGE: Any comments on the quality construct? If not, we will vote.

MS. IBRAGIMOVA: Importance to Measure and Report: 1c, composite. One, high; two, moderate; three, low; four, insufficient.

The results are: 65 percent high; 35 percent moderate.

CO-CHAIR GEORGE: We will move on to specifications and reliability testing.

MEMBER CLEVELAND: The specifications, I think, are solid. The reliability testing occurred at the facility level with NCDR ICD Registry data, again from 1,606 hospitals. The reliability testing of the measure score was performed using a correlation of random split halves, which is an NQF accepted method. The see a correlation coefficient of 0.87, quite high.

The developer also included a description of the NCDR IDC data registry or data quality program, which its description involves Inter-rater Reliability Assessments in on-site audits. Kappa scores for that have a 95 percent confidence interval, which is quite high. So, I think based on these things, the reliability is high.

CO-CHAIR GEORGE: Any discussion on reliability? We'll vote.

MS. IBRAGIMOVA: Scientific Acceptably of Measure Properties: 2a, reliability. One, high; two, moderate; three, low; four, insufficient.

We're just missing one vote.

MS. VICALE: Has everyone placed their vote?

MS. IBRAGIMOVA: The results are: 71 percent high; 29 percent moderate.

CO-CHAIR GEORGE: Validity.

MEMBER CLEVELAND: So, validity testing was conducted with empiric testing at the

data element and also measure score. It was actually using a sample of 93,000 Medicare Fee-For-Service patients greater than 65 years of age who underwent ICD implantation.

There was an analysis, so this analysis did reveal an association of both patient and hospital performance in the composite measure with adverse outcomes, specifically with mortality readmission.

So, there was a significantly smaller proportion of patients discharged on appropriate medical therapy who died or were readmitted within six months of hospital discharge; 28 percent without meds versus 36.3 percent with meds. So, again, the composite measure seemed to reduce death and readmission.

And then at the facility level, patients treated at hospitals that performed better on this measure had been unadjusted outcomes than those treated at hospitals that performed worse.

There was also face validity that was

completed with a variety of content experts that were submitted by the measure developer.

So, based on those two things, I think validity is met.

In terms of threats, the exclusions are relatively straightforward. If you, obviously, died, you weren't included. And those patients not eligible for ACE or ARB or beta blockers, I assume those are people that had defined contraindications in the NCDR database. Exclusions are rare.

The measure is not risk-adjusted. And lastly, the missing data are treated as performance not met.

CO-CHAIR GEORGE: Any discussion on the validity? Judd.

MEMBER HOLLANDER: I don't have a big issue with this, and you have probably had this discussion 13 times this morning already before I got here, but it doesn't take into account the sociodemographic status of the patients for this particular measure. And when you are talking

about getting two medications after an expensive therapy, there are people who may not be able to afford it in certain sociodemographic groups.

So, I just want to throw that on the table. You know I don't know if you have been around this all day already or not.

CO-CHAIR GEORGE: Would you be suggesting that they stratify this by SDS?

MEMBER HOLLANDER: Yes, I mean I think it makes sense. If you are catering to a higher population of poor people who can't get medications and you are measuring whether they got it, then that might be relevant. And it is unfair to hold the hospital against it. The patients probably still benefit more from the therapy than not, in terms of the things that get you into this pool.

So, you wouldn't say okay, don't put anything in as a result of it but yet, I guess this is all new and I know you have had the discussion already. So, I don't want to recreate the discussion if you guys have reached peace at this.

CO-CHAIR KOTTKE: It's described I

feel that there is generics available for both of these classes of drugs. It is not BiDil.

MEMBER HOLLANDER: You know, I am not saying I would vote against it. I am just saying I am just throwing that out. And having generics and spending $10 or $20, for some people, it is still $10 or $20 that they can't pay enough.

DR. MASOUDI: I mean short of aspirin, perhaps, this is an argument that could be made for any medication process measure, I think. And more so for some than others, as you pointed out there.

CO-CHAIR GEORGE: Any other comments on the validity? If not, we will vote.

MS. IBRAGIMOVA: Scientific Acceptability of Measure Properties: 2b, validity. One, high; two, moderate; three, low; four, insufficient.

The results are: 53 percent high; 47 percent moderate.

CO-CHAIR GEORGE: The analysis to support the composite.

MEMBER CLEVELAND: The developer

states that an empirical analysis demonstrating the individual component measures fit into this quality construct design going in to be published in medical literature. There has been nothing else done in regard to that.

CO-CHAIR GEORGE: Any discussion on this? If not, we will vote on it.

MS. IBRAGIMOVA: Scientific Acceptability of Measure Properties: 2d, composite. One, high; two, moderate; three, low; four, insufficient.

The results are: 35 percent high; 59 percent moderate; 0 percent low; 6 percent insufficient.

CO-CHAIR GEORGE: Feasibility.

MEMBER CLEVELAND: These data sources are readily available in electronic sources transferred to the NCDR or web entry are both available. Also, the developer states that centers actually, I guess, this is mandated participation in this registry for reimbursement of ICDs. So, you have those data there already.

So, I think it is feasible.

CO-CHAIR GEORGE: Any comments on feasibility? We'll vote.

MS. IBRAGIMOVA: Feasibility: One, high; two, moderate; three, low; four, insufficient.

The results are: 88 percent high; 12 percent moderate.

CO-CHAIR GEORGE: Usability.

MEMBER CLEVELAND: So, while the measure has been used, it is not currently publicly reported. The planned use, however, is for public reporting with external benchmarking and also internal benchmarking specific to the organization.

I could not identify any unintended consequences, other than the systemic, obviously, there are some inaccuracies, data entry perhaps, potential for medication exclusions, but I think those are very small. Therefore, I think usability use is fine.

CO-CHAIR GEORGE: Any comments on

usability? If not, we will vote.

MS. IBRAGIMOVA: Usability and use. One, high; two, moderate; three, low; four, insufficient information.

The results are: 94 percent high; 6 percent moderate.

CO-CHAIR GEORGE: Any comments before we vote on the measure itself? We will vote on the measure for endorsement.

MS. IBRAGIMOVA: Overall Suitability for Endorsement: Does the measure meet NQF's criteria for endorsement? One, yes; two, no.

The results are: 100 percent yes; 0 percent no.

CO-CHAIR KOTTKE: Okay; thanks. The next measure 2712, Statin Use in Persons with Diabetes, PQA. Dr. Eisenberg. Liz and Tom James.

MS. VICALE: As the panel settle in, we do, again, want to appreciate the succinctness of the presentations, as well as any responses to questions and the conversation in not really repeating what others have already stated. Again,

we are trying to keep the measures to about 15 minutes.

DR. EISENBERG: Thank you and hello. I am Woody Eisenberg. I am the Senior Vice President for Performance Measurements at PQA and I am joined by my colleagues Kristen Butterfield, who is Director of Research and Analytics, and by Julie Khule, who is Vice President for Performance Measure Operations. And we are here to talk to you about Measure NQF 2712, Statin Use in Persons with Diabetes.

The American College of Cardiology and the American Heart Association guidelines recommend moderate to high-intensity statin therapy for primary prevention for persons aged 40 to 75 years with diabetes. And that is a Class I recommendation.

The proposed measure, Statin Use in Persons with Diabetes, is intended to be used at the healthcare level for plans that have access only to pharmacy claims data, such as Medicare Part D plans, in which this measure is currently being

used as a patient safety reporting-only metric.

Just for your information, as background, such programs that have access only to pharmacy claims data cover, today, 29.3 million Medicare lives, and that represents about 61 percent of the Medicare Part D population.

The measure, however, is also suitable for state Medicaid programs, some of which have prescription-only data. Of course, they are using delivery models.

Because this measure -- Statin Use in Persons with Diabetes -- uses prescription claims as a data source, it uses this prescription data as a proxy for diabetes diagnosis. However, we have tested the measure using medical claims data and what we found is that the denominator criteria of two prescription claims for a hypoglycemic agent identified a population where the great majority -- 90 percent -- had a diagnosis of diabetes confirmed using medical claims data.

Additionally, this denominator included very few persons, less than one percent,

with select conditions to which the guidelines would not apply, such as patients with polycystic ovary syndrome, gestational diabetes, or diabetes secondary to another condition.

Administrative pharmacy claims demonstrate a high degree of reliability and are generated as a standard, essential part of care. So, no additional burden is placed on the health plans to recommend this data.

CO-CHAIR KOTTKE: Thank you. Liz, evidence.

MEMBER DELONG: In terms of evidence, it is not clear that using prescription claims data could adhere to whatever evidence there is for this measure. You can't identify individuals with contraindications to statin therapy or recommend exceptions. So, even though the statin therapy may have significant evidence, I don't see that this measure would measure what the evidence claims.

DR. EISENBERG: May I respond to that?

MEMBER DELONG: Tom might have

another, a better feeling for the evidence.

DR. EISENBERG: I think the objection is that there might be people included who have adverse reactions to the statins. Is that it?

MEMBER DELONG: That is one of the problems.

DR. EISENBERG: Yes, and that is so. Keep in mind, though, that this is not a provider-based measure. This is a population-based measure at health plans with hundreds of thousands of millions of numbers.

Also, I point out that the recommendations for people that are intolerant in one way or another of using these medications is to try again. You know, cut down on the dose, switch to a different medication, so that there is a great effort, really, to keep people on it -- to place them on statins and keep them on statins.

CO-CHAIR KOTTKE: Tom? Tom James, do you have anything?

MEMBER JAMES: Yes. Yes, the comment that I had, and Woody, I appreciate what you have

put together, I think that the evidence is very solid for what this measure purports, but we have to recognize that this is not a measurement of all patients with diabetes. This is a subset, and there is a bias that is built into this particular measure.

Those people how are on medications, therefore, are the individuals with diabetes who are probably at higher risk than those who could be managed by diet alone or by exercise.

But invariably, those people who are taking medications also represent a biased population in that they are willing to accept medication management more readily, I suspect, than those people who refuse to take medications and are not included in this set.

So, there are two directions that come in impacting whether we are really treating people with diabetes appropriately with the use of statins, but recognizing those biases, it is still a good measure. It has got good evidence behind it.

CO-CHAIR KOTTKE: Okay, thanks, Tom.

Anybody else have a burning urge to say anything about evidence?

Is that Linda down there or is that Mladen? Joel.

MEMBER MARRS: I guess I have a question of citing the recent ACC/AHA Guidelines as your evidence to support this. Why not actually evaluate intensity of statins, since you actually have this from a claims -- a prescription claims standpoint?

DR. EISENBERG: We thought about that and talked about it a lot. And we were advised by our testers that getting the details of the intensity of statin therapy would pose perhaps insurmountable problems and might lead to incorrect information.

So, we decided, given the fact that this is a primary prevention-focused measure to go along with the use of any statins. I know other measure developers have done the same thing.

CO-CHAIR KOTTKE: Seeing nobody

moving, let's vote on evidence.

MS. IBRAGIMOVA: Importance to Measure and Report: 1a, evidence, structure, process, intermediate outcome. One, high, only eligible if QQC submitted; two, moderate; three, low; four, insufficient.

The results are: 24 percent high; 65 percent moderate; 12 percent low, and 0 percent insufficient.

CO-CHAIR KOTTKE: Thank you. Opportunity for improvement?

MEMBER DELONG: Well, the figures quoted have to consider the fact that 100 percent is not the target because you don't really know how many people really qualify for this measure. It is a biased measure and you can't identify diabetics who should be on care. It is only the diabetics who you pick up who have been prescribed but you don't even know if they are taking it. And you cannot identify contraindications.

I think there is probably a gap, but there was no way of proving that.

CO-CHAIR KOTTKE: Tom.

DR. EISENBERG: I would just point out once again that there are tens of millions of people in plans that don't have the kind of information that we would need in order to address that issue. You are correct in what you are saying.

MEMBER DELONG: But we are endorsing it as a measure.

CO-CHAIR KOTTKE: Yes, Tom, go ahead.

MEMBER JAMES: Yes, this is Tom James. The utility in this particular measure is how it moves over time. That is a health plan that is the unit of measurement here should be able to benchmark itself from one year to the next and see improvement in this level.

But you are absolutely right that it is not going to be 100 percent, but we should see movement in the positive direction.

CO-CHAIR KOTTKE: Seeing no movement, let's vote on opportunity for improvement.

MS. IBRAGIMOVA: Importance to Measure and Report: 1b, performance gap. One, high; two,

moderate; three, low; four, insufficient.

The results are: 24 percent high; 71 percent moderate; 6 percent low.

MEMBER DELONG: I don't know, maybe you can do this, but you specified that these are patients who are continuously enrolled during the measurement period. It is not clear what you mean by enrolled because patients change their enrollment plans. They could drop out of the pharmacy benefit plan that they were on six months ago. Once again, there are the biases. You have got gestational diabetes, steroid-induced diabetes, and polycystic ovarian disease, who could be prescribed these agents and they are not excluded.

I don't think the reliability is high.

DR. EISENBERG: The issue of changing health plans exists for, I guess, any measure that is a health plan measure. So, yes.

We did separately analyze patients that had other reasons for having diabetes, gestational diabetes, secondary diabetes, polycystic ovary

syndrome. Those numbers were less than 1 percent for every population that we examined, which included commercial Medicaid and Medicare.

MEMBER JAMES: In addition, you did signal to noise types of testing and a variety of other testing to ensure the reliability. So, I felt comfortable with this. Again, coming from a health plan background, I understand that whole issue of the enrollment. This is not something would be measuring physicians. This is measuring health plans.

MEMBER DELONG: I couldn't find the results of the signal to noise analysis. Tom, do you have them?

MEMBER JAMES: You have to go through the clicks and then getting it --

MEMBER DELONG: I went through the clicks and I couldn't get them.

CO-CHAIR KOTTKE: While you are clicking, Leslie, you had a --

MEMBER CHO: So, the use of a diabetic medication was actually used for diagnosis of

diabetes. Correct? So, I think that eliminates a lot of Liz's concern about gestational diabetes and whatnot.

I mean I personally think this is a very good measure.

CO-CHAIR KOTTKE: Okay; thank you. Tom, did you find what you were looking for?

MEMBER JAMES: I left my password at home. There is a way I found so you can get through there.

MEMBER DELONG: Well, they are here. They could tell me.

MS. BUTTERFIELD: So, we used a mixed effects regression model to model individual --

MEMBER DELONG: Right and I saw no -- you didn't provide any details on what that mixed regression model was.

MS. BUTTERFIELD: So we looked at -- we modeled the individual's likelihood of being on a statin to the varying health plan mean. So, looking at a random effect of the health plan.

And what we found is we looked at the

standard deviation of the intercept for the random effects, so looking at the contribution of the health plan is making in the model. We found that there was a significant difference, based on a confidence interval around the standard deviation of the intercept, which can be interpreted as saying that there are significant variation between performance scores at the health plan level.

MEMBER DELONG: You didn't provide any data to that effect. I mean you looked at the standard deviation and assumed that that meant there was a gap. But it would have been helpful to provide some of those numbers so that we could see what that variation is.

CO-CHAIR KOTTKE: Okay, we'll call that a deficiency, but let's vote on reliability.

MS. IBRAGIMOVA: Scientific Acceptability of Measure Properties: 2a, reliability. One, high; two, moderate; three, low; four, insufficient.

The results are: 12 percent high; 82

percent moderate; 6 percent low.

CO-CHAIR KOTTKE: Validity, Liz.

MEMBER DELONG: They present results of a consensus panel. I guess that works.

CO-CHAIR KOTTKE: Okay, Tom, any thoughts? Validity.

MEMBER JAMES: This has a lot of good face validity, and that is as much statistical analysis as I can understand.

CO-CHAIR KOTTKE: Seeing no movement, let's vote on validity.

MS. IBRAGIMOVA: Scientific Acceptability of Measure Properties: 2b, validity. One, high; two, moderate; three, low; four, insufficient.

The results are: 12 percent high; 88 percent moderate.

CO-CHAIR KOTTKE: I just had a thought. Remember the old days when we used to hold up our hand? You guys didn't remember that. We used to have hold up our hands; that was terrible.

(Laughter.)

CO-CHAIR KOTTKE: Now, I have forgotten where we are. Feasibility.

MEMBER DELONG: Given what they are doing, it is definitely feasible.

CO-CHAIR KOTTKE: Good. Tom, feasible?

MEMBER JAMES: Agreed.

CO-CHAIR KOTTKE: Okay, let's vote.

MS. IBRAGIMOVA: Feasibility. One, high; two, moderate; three, low; four, insufficient.

The results are: 94 percent high; 6 percent moderate.

CO-CHAIR KOTTKE: Usability.

MEMBER DELONG: Well, it is being used by several health plans and whatever. It is still based on pharmacy data that are notoriously unreliable.

DR. EISENBERG: I really must take exception to that.

MEMBER DELONG: I know you will.

CO-CHAIR KOTTKE: We will let you take

exception.

DR. EISENBERG: Yes, pharmacy data are notoriously reliable. They all use NCPDP standards; all of the transmission is the same; the data elements are all the same.

MEMBER DELONG: For diagnosis they are reliable?

CO-CHAIR KOTTKE: Well, we are missing diabetics, but the denominator here is people taking diabetic medications. Tom?

MEMBER JAMES: Yes, recognizing that the inherent population bias that I mentioned earlier, I think this is still reliable. Pharmacy data is some of our better data. In fact, it is probably better than EHR data at this juncture in time.

CO-CHAIR KOTTKE: Okay, let's vote on usability.

MS. IBRAGIMOVA: Usability and use. One, high; two, moderate; three, low; four, insufficient information.

The results are: 59 percent high; 35

percent moderate; 6 percent low.

CO-CHAIR KOTTKE: So, now it is time to tell us what you really think. Vote yes to recommend the measure or no to decline it.

MS. IBRAGIMOVA: So overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? One, yes; two, no.

The results are: 94 percent yes; 6 percent no.

MEMBER JAMES: We won the speed record for the day.

DR. EISENBERG: Many thanks to the committee.

CO-CHAIR GEORGE: So, we will be moving on to Measure 0669. May we have a few comments from the developers, briefly?

DR. BRUETMAN: Thank you. Good afternoon. I'm Dr. Bruetman. I am from the Lewin Group, and with me are my colleagues, Kelly Anderson, beside me, and Colleen McKiernan, who have been working on the development and

maintenance of the measure. We are presenting today the Cardiac Imaging for Non-Cardiac, Low-Risk Surgery. And we want to thank the NQF and the committee for the opportunity to present and clarify any concerns or questions you might have. Also, thank CMS for their support.

And we have been developing and maintaining this measure together with CORE, which is the Center for Outcomes Research Evaluation at Yale.

So, today the measure -- just to give to you an overview of the measure -- the measure calculates the percent of the stress echocardiography single photon emission computed tomography myocardial perfusion imaging, or SPECT MPI, or stress magnetic resonance imaging studies performed at a hospital outpatient facility in the 30 days prior to an ambulatory low-risk non-cardiac surgery performed anywhere.

And the denominator includes the number of stress echo, SPECT MPI, and stress MRI studies performed at the hospital outpatient department

and the numerator includes, of those patients, the denominator, those that have a stress echocardio, SPECT MPI, and stress MRI performed at the hospital outpatient department within 30 days of an ambulatory low-risk non-cardiac surgery performed at any location.

So, just to give you an idea, this measure is part of the Hospital Patient Quality Reporting Program at CMS, and the goals are to promote high quality and efficient care, reduce unnecessary studies, contrast and radiation exposure. It is based on adherence to evidence-based medicine and guidelines and provides consumers with information on facility imaging use.

Finally, the guidelines suggest that cardiac imaging is not recommended for pre-op assessment for non-cardiac low-risk surgeries, since these tests do not change a patient's clinical management or outcomes.

The measure has been initially endorsed by NQF in 2011; it has been publicly reported by

CMS in 2012.

Thank you.

CO-CHAIR GEORGE: Thank you. Joe? Nick?

MEMBER RUGGIERO: This is a maintenance, and I think it is a good measure because of the fact -- as an interventional cardiologist -- we still see a fair number of people come to the cath lab for an abnormal stress test prior to cataract surgery.

If we look at the evidence here, the evidence is based upon guidelines, non-guideline statements, with varying degrees of level of evidence of support, along with an additional 14 articles to support the measure's intent.

So, I think, overall, this has a high level of evidence.

CO-CHAIR GEORGE: Any comments? We will vote on the scientific evidence.

MS. IBRAGIMOVA: Importance to Measure and Report: 1a, evidence, structure, process, intermediate outcome. One, high, only eligible if

QQC submitted; two, moderate; three, low; four, insufficient.

The results are: 71 percent high; 24 percent moderate; 6 percent low.

CO-CHAIR GEORGE: Opportunity for improvement.

MEMBER RUGGIERO: So, if you look at the data that they present, based upon the maximum performance rates, which range from about 15 percent to 18 percent and the mean performance rate, which is about five percent, it shows that there is still a significant disparity between facilities performing these studies. It also shows that there is also a race and ethnicity, as well as the location of the facility, as far as how its testing is performed. So, I think you have the opportunity for improvement gap in care with the disparity and also in SDS here.

CO-CHAIR GEORGE: Any comments on the opportunity for improvement? We'll vote.

MS. IBRAGIMOVA: Importance to Measure and Report: 1b performance gap. One, high; two,

moderate; three, low; four, insufficient.

The results are: 82 percent high; 18 percent moderate.

CO-CHAIR GEORGE: Specifications and reliability testing.

MEMBER RUGGIERO: So, for specifications, I think that everything is very clearly defined with the large of number codes that were added. And reliability testing, it was conducted at the level of the performance measure score. The primary analysis was conducted at the facility level, using two test with ability to identify statistical outliers, as well as signal to noise and they give a mean reliability score of about 43 percent, which they would get as being moderately reliable.

So, I think it is moderately reliable, based upon this.

CO-CHAIR GEORGE: Any comments on reliability or specifications? We'll vote.

MS. IBRAGIMOVA: Scientific Acceptability Measure Properties: 2a,

reliability. One, high; two, moderate; three, low; four, insufficient information.

Just waiting for one more vote.

The results are: 35 percent high; 65 percent moderate.

CO-CHAIR GEORGE: Validity.

MEMBER RUGGIERO: Face validity of the measure score and data elements were looked at through a seven-member Technical Expert Panel. And if you look at it, it had about 75 percent agreement of the 30-day window to look forward towards the surgery. They weren't able to reach consensus, however, based upon what clinical conditions should be excluded. However, those were based upon the AHA/ACC Guidelines. So, I think it was less important. So, I think it was valid.

CO-CHAIR GEORGE: George?

MEMBER PHILIPPIDES: So, I do have a question about the denominator exclusions. Three of the following: diabetes, renal insufficiency, stroke, heart failure and ischemic heart disease.

So, if you have got ischemic heart disease, you are excluded from this measure. So, in other words, somebody with ischemic heart disease going for a cataract, I send them for a nuclear test, which leads to a cath with you, and that wouldn't be a ding to me because they are excluded.

So, you are missing a whole lot of what I think are good things of really trying to take out the problem.

And the second issue, I have a patient -- we are going to turn it around the other way -- who is having chest pain that worries me, and they have risk factors and I get a nuclear test. And thank goodness it is okay, nothing wrong. A week or two weeks later, unbeknownst to me, they go for a cataract procedure. Then I do get dinged. Is that correct? Because I ordered -- no?

MS. ANDERSON: So, to respond to both of your questions, to the first point, if they only had one of those conditions, they wouldn't be excluded from the measure. They would have to meet

three or more of those criteria. So, suggesting some of the high-risk patients, rather than patients who just have one potentially complicating factor.

And to your second question, there are going to be incidental cases where you do perform a stress test and then the patient goes on to have a procedure, but we are doing it in a very low number of cases.

MEMBER PHILIPPIDES: Do you guys -- out of curiosity -- have that number? Were you able to ascertain that from your review?

MS. ANDERSON: We were not able to test that.

CO-CHAIR GEORGE: Other comments on validity? If not, we'll vote.

MS. IBRAGIMOVA: Scientific Acceptability of Measure Properties: 2b, validity. One, high; two, moderate; three, low; four, insufficient.

The results are: 6 percent high; 94 percent moderate.

CO-CHAIR GEORGE: Feasibility.

MEMBER RUGGIERO: Because the data source includes administrative claims using CMS hospital outpatient claims, it is very feasible, as they should be very easily collectable.

CO-CHAIR GEORGE: Any comments on feasibility?

CO-CHAIR KOTTKE: I just had a question. Why did ACC have such trouble using CMS claims for the Medicare with their measure, when this is feasible?

I mean you can't answer that; it is rhetorical.

MS. SLATTERY: So, as I mentioned, we are intending to report without a CMS contract in place. And so there are not regulations that exist currently that permit that. I suspect this one is being reported with a CMS contract in place. So, regulations are in place to permit it.

DR. BRUETMAN: Yes, this is under a CMS contract, so we have access to 100 percent of their data sets.

CO-CHAIR GEORGE: Any other comments on feasibility? We'll vote.

MS. IBRAGIMOVA: Feasibility. One, high; two, moderate; three, low; four, insufficient.

The results are: 71 percent high; 29 percent moderate.

CO-CHAIR GEORGE: Usability.

MEMBER RUGGIERO: So, it is already publicly reported on CMS's Hospital Outpatient Quality Reporting Program. And I think since it is already reported, number one, number two, it is going to give a tremendous amount of data going forward as far as identifying those outliers, I imagine at some point reimbursement, and so on.

CO-CHAIR GEORGE: Any comments on usability? We'll vote.

MS. IBRAGIMOVA: Usability and use. One, high; two, moderate; three, low; four, insufficient information.

The results are: 88 percent high; 12 percent moderate.

CO-CHAIR GEORGE: Any final comments before we vote on the measure? We'll vote on the measure.

MS. IBRAGIMOVA: Overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? One, yes; two, no.

Just missing one vote.

The results are: 100 percent yes; 0 percent no.

CO-CHAIR KOTTKE: Okay, thank you. The last two measures of the day, 0229 and 0230, we are going to do together. To give Liz a little bit of a break, I would suggest that we ask Sana to be the primary and Kristi to the be the secondary. And Leslie can poke them if they talk too long.

And then if Liz or Judd want to --

MEMBER DELONG: Judd wants to.

CO-CHAIR KOTTKE: Judd, do you want to do it? Yes but we are doing it together. So, we are just -- because the only difference is is whether you were hospitalized for heart failure or

MI. It is 30-day all-cause RSMR following for patients 18 and older. So, they are nearly identical except for the reason for admission. So, the science should be the same.

MEMBER AL-KHATIB: I can proceed, if everybody is okay. I can start.

CO-CHAIR KOTTKE: Okay, Sana will start and anybody else can --

MS. WILBON: So, we will go through the importance and scientific acceptability of each measure separately. I believe the scientific acceptability is going to be probably slightly different because of the condition. The evidence, obviously, is going to be different for importance. And then scientific acceptability may be slightly different.

And then we will vote on usability and feasibility one time for both measures, just because they are so similarly constructed, that those votes, we will just duplicate those votes in the report for both measures.

And then we will vote separately for

endorsement for each measure.

CO-CHAIR GEORGE: Okay.

MS. WILBON: Sorry. And recommendation for suitability for endorsement for both measures.

I'm not sure I understood what you said but yes. If there is another way that makes it a little bit easier logistically, we can do that, too.

MS. HERRING: Sorry, everyone. So, we are going to go through the first measure that we are going to review and we will go through all the votes on that one. And then for the second measure, we will just vote on the first two portions and then we will manually enter the second two. That will just be the same that we all did together, if that's okay.

CO-CHAIR KOTTKE: Okay, Sana -- no. Well, then we are going to start with 0230, I guess, and that is Liz.

MS. BERNHEIM: Okay. Hi, I'm Susannah Bernheim. I'm the Director of Quality Measurement

at the Yale CORE Group and I have with me Jeph Herrin, who is a statistician from Yale on these measures.

I am aware, first, that I am the only thing between you and dinner. So, I will try to be brief. And also that I now don't know which measure I am presenting. They are very similar but I'm not sure which we are on. Am I talking about AMI or heart failure? AMI. Okay, thank you.

So, I have a set of slides. They are not up. I can go without them. They were there a second ago. But I will just present without them -- oh, great. Okay.

So, I don't need to spend a lot of time with this committee talking about the importance of measuring AMI and mortality. The only comment I will make is that we have seen a lot of improvement since this measure started being reported but we are still seeing substantial variation across hospitals, which is a sign of sort of ongoing importance.

This AMI 30-day hospital mortality

measure has been publicly reported as part of CMS's inpatient quality reporting program since 2007. So, it is a measure that is coming back for endorsement maintenance. It has been endorsed since 2007 as well. It was included in CMS's hospital value-based purchasing program in 2013.

They note that there was a full medical record validation done when the original measure was developed. And we, at Yale CORE, annually review the measure to see if there any updates.

And I just wanted to flag for this committee the updates that have occurred since this measure was lastly endorsed by NQF. They are not major but we have included the VA hospitals. I think that was in 2011. Don't quote me on that date.

We have incorporated new formatting for claims and annual updates to the map that is used to bring the respecters into the map.

I'm going to do a really quick overview of the measures. We look at patients discharged with a principle diagnosis of AMI who are 65 or

older. Although the measure has been tested to be used as an all-payer measure, it is currently reported in the over 65 Medicare Fee-for-Service population and VA beneficiaries. We include patients who have not been transferred. The first hospitalization in the transfer is considered the hospital responsible for the 30-day outcome. And we only include patients who have Part A and Part B for the 12 months prior to admission, in order to have risk adjustment.

Key exclusions, this is the next slide, are patients who are discharged alive on the day of admission or the following day and have not been transferred to another facility. And then some very, very small number for demographic issues. We exclude patients who are in hospice in the 12 months prior and the day of admission. And we exclude patients who are discharged against medical advice.

The measure adjusts for age, gender, and comorbidities. We use ICD-9 codes for inpatient and outpatient claims for the 12 months

prior to and including the index admission. They are grouped with those CMS condition categories and we don't include any that could be complications of care.

And we use -- we look at all-cause mortality within 30 days from the date of the index admission. We use a statistical modeling that is a hierarchical generalized linear model to account for the in-hospital correlation. And it is reported as a predicted to expected ratio that has been multiplied by the national observed mortality rate.

This is results from this year's reporting. The only two things I think are worth looking at is the last three separate years and then the full three-year group together. You will note at the bottom the c-statistic for this measure is 0.72.

The rates for AMI have been going down steadily. So, if you look across the mean for just these last three years of reporting, we have gone from 14.8 to 14.6 to 13.3.

And I think that may be all. Let's see what the next slide is. Okay, I have slides on SES if it comes up but we don't need to go there. I'm happy to but we don't need to talk about it now.

CO-CHAIR KOTTKE: Okay. Liz -- oh, Judd.

MEMBER HOLLANDER: So, I did this.

CO-CHAIR KOTTKE: Okay.

MEMBER HOLLANDER: So, thank you for doing my job so beautifully. This is really easy.

MS. BERNHEIM: I'm sorry. I didn't know it was your job.

MEMBER HOLLANDER: No, this is great! I think all the measure developers should do the same thing.

So, I think the evidence is pretty good. It is really hard to argue with this. And you know they presented it very nicely, the rationale that at a facility level you can make a difference and you will see some of the data. So, I don't have a lot to add with respect to the evidence.

CO-CHAIR KOTTKE: Anybody else?

Leslie says let's vote. So, why don't we vote?

MEMBER DELONG: I don't have -- was this measure developed actually on California data several years ago?

MS. BERNHEIM: Yes, I am sorry for the confusion. I saw in one of the early comments there was confusion.

So, the measure was developed in claims data ten years ago. It was validated with a national sample but we were asked to look if it would work in an all-payer population. And so we used California data and looked at whether, within California, where we had all-payer data, it would work.

So, the California data was just to validate that the measure worked in an all-payer population. It's not currently used that way but it would be.

MEMBER DELONG: But the model itself was not developed exclusively.

MS. BERNHEIM: No, the model was developed in national Medicare data.

CO-CHAIR KOTTKE: Okay, let's vote.

MS. IBRAGIMOVA: Importance to Measure and Report: 1a evidence, health outcome, or PRO. One, yes; two, no.

We are just missing one.

The results are 94 percent yes, 6 percent no. And this only applies to the 0230 measure.

CO-CHAIR KOTTKE: So, we are doing opportunities for improvement for 0230. Right?

Judd, can we do better?

MEMBER HOLLANDER: So, I think we saw the performance gap on the data that you showed. So, I have nothing to add.

CO-CHAIR KOTTKE: Let's vote.

MS. IBRAGIMOVA: Importance to Measure and Report: 1b, performance gap. One, high; two, moderate; three, low; four, insufficient.

MS. VICALE: Tom, please enter your vote in the chat.

MEMBER JAMES: I'll do it again.

MS. VICALE: Tom, can you just text the

vote? We are having a little technical issue with the chat box. Thanks.

MS. IBRAGIMOVA: The results are 69 percent high, 31 percent moderate.

CO-CHAIR KOTTKE: Specifications and reliability testing.

MEMBER HOLLANDER: So here, this is actually kind of curious to me. So, they did the reliability testing and you might need to explain this, but they looked at nearly 500,000 admissions over a three-year period, split them in half into two samples, and then did the risk-standardized mortality rates at each hospital and then looked at the intra-class correlation coefficient and it was only 0.41.

And I don't see why -- and that is actually listed here as moderate. I consider that pretty bad, actually. It is not kappa, it is an intra-class correlation coefficient. And frankly, I can't explain it.

And then they wrote in here that oh, that is just they split it into a year and a half

each and if they were larger numbers, it would be better. It is sort of like reading the paper that it wouldn't be a power problem if there were twice as many, it would be significant.

So, I don't actually buy that explanation but on the other hand, it doesn't make sense to me that it should be that far off in the testing because that is not any of our experience. So, I don't know if you can comment on that.

DR. HERRIN: Well, I guess first I would say that the measure of reliability we used in the test is not the same metric that a lot of measures report. From the beginning, there is a measure reliability signal to noise ratio which a lot of people use, and this is one that sort of gets to the same thing but it is different in that the signal to noise ratio is looking at how much the measure actually distinguishes different entities you are measuring.

We haven't been as concerned with that because we construct confidence levels for our measures and we use those to identify whether there

are outliers.

So, if we construct confidence intervals and we see that if there is a certain number of hospitals, that are at a higher level, then we are confident that we are seeing a signal. And that is what we are seeing.

So, this reliability measure we report is really just a measure of how consistently a hospital would be measured if we measured it multiple times.

So, I want to make it clear this is a different kind of reliability. The fact that it is 40 percent, yes, we would all like it to be higher. But what we have seen consistently with these measures over the last five years is in the range 40 to 60 percent.

So, I think we are content with it. We would like to see it higher but then we also don't know how it actually -- we don't know how to compare it to other kinds of liability measures.

CO-CHAIR KOTTKE: George.

MEMBER PHILIPPIDES: A question about

the exclusions. You excluded people who were discharged alive on the first day or the next day of admission, even though they weren't transferred to an outside hospital. I'm just not clear why.

MS. BERNHEIM: So, when the measure was first developed, this exclusion was put in and the thought was it was a way to ensure that we were really getting medically significant AMIs. So, if somebody went home the same day they showed up, did they really have an AMI?

And we actually thought very seriously about reconsidering this exclusion in this measure. And to tell you the truth, we didn't change it because the CMS payment policy around people who only stay one night is in flux and these patients are basically going to all end up out of the measure because they are not going to be considered inpatient admissions any more. And before we made a change, we decided we would just leave it as-is, leave that three-day payment -- not the three-day payment -- the to midnight rule -- I just confused to CMS rules. I apologize -- the

to midnight rule settle down and see where we land.

We think where we are going to land is that these patients are just going to be excluded by payment policy.

So, that was the original theory and why we didn't make the change in this round of the evaluation.

CO-CHAIR KOTTKE: Sana.

MEMBER AL-KHATIB: So, actually my comment has to do more with the heart failure one but since we are talking about the specifications, I would like to bring it up now.

One of your exclusions was patients who are enrolled in Medicare hospice program within 12 months before the encounter. What about people who get discharged to hospice? I mean you are expecting those people to die. So, if they die within a week or two weeks after discharge, you were expecting that. So, why are we dinging the physicians for that?

MS. BERNHEIM: So, this has been a hard question for the mortality measures all along.

And what we wish we had was a marker that you were enrolled in hospice based on your clinical status when you showed up at the hospital because those patients we would want to capture.

So, if you get enrolled the first day, which is rare, we exclude you. The problem with enrolling based on hospice status at discharge is we are trying very hard not to risk adjust or exclude based on things that have happened during your clinical care. And the concern that has been raised is if you have a complication that should have been prevented and that leads to your condition deteriorating and then you go into hospice, if I take those patients out, I am losing part of your quality signal.

I will say it is not perfect. It has always frustrated us and it is something we really are hoping the EHRs are going to help us fix. But when we have talked repeatedly to Technical Expert Panels, the support has been for doing it this way as the best solution with the data that we have.

But that is the concept, that you might

be getting rid of complications that you might not have.

CO-CHAIR KOTTKE: Liz and then Leslie.

MEMBER DELONG: To enlighten Judd, this morning we actually had an ICC of less than three percent. And I think that was the result of randomly dividing into two groups.

This one seems to be dividing on the basis of time. And you might expect less intra-class correlation from time to time because hospitals might be improving more than others.

What concerns me is that you are also seeing an increase in coding of comorbidities and that could account for some of that fluctuation if some hospitals are up-coding more than others.

DR. HERRIN: We are not dividing them on the basis of time. We are using the same time period.

CO-CHAIR KOTTKE: Leslie.

MEMBER CHO: Well, are transfer patients lost?

MS. BERNHEIM: No, sorry. It is

always a hard thing to explain.

No, if you are transferred, if you are admitted to Hospital A and transferred to Hospital B, we follow you starting with the index admission to Hospital A and then mortality is associated with Hospital A. And Hospital B is not counted as the index in that case.

CO-CHAIR KOTTKE: Okay, ready to vote? Let's vote.

MS. IBRAGIMOVA: Scientific Acceptability of Measure Properties: 2a, reliability. One, high; two, moderate; three, low; four, insufficient.

The result are 18 percent high, 76 percent moderate, 6 percent low.

CO-CHAIR KOTTKE: Validity.

MEMBER HOLLANDER: So, this is great. They did a correlation between the claims-based mortality rates and record review based rates and the correlation was 0.91. So, you can't do a lot better than that.

CO-CHAIR KOTTKE: Liz.

MEMBER DELONG: Well, the statistician can explain this but a correlation doesn't mean an exact correspondence. As a matter of fact, I did a little bit of a simulation and for 1,000 hospitals, if you have a correlation of 0.9, you could still have three percent of hospitals ranked in the top ten percent by one metric versus not in the other metric, if you see what I am saying.

Three percent of hospitals could be misclassified and we are talking about claims data versus clinical EHR data. One presumably might be the gold standard and we are missing three percent of the time on one end and three percent of the time on the other end.

So, if there is a reward for being in the top ten percent, those who would have been in the top ten percent by clinical data, three percent of them might miss that reward. And likewise, the penalty.

CO-CHAIR KOTTKE: So, you want to make sure you are in the top five percent.

MEMBER DELONG: Well, I just question

the validity of using claims data for these purposes.

CO-CHAIR KOTTKE: SDS?

MEMBER HOLLANDER: Yes, so they actually did a really, really extensive analysis to come up with their risk adjustment model and then looked at SDS and found that race and dual eligible status were related. But when they plugged that into the model, it just didn't make enough of a difference to want to keep it in the model.

And I think through the two pages of explanation here, they tell a good enough compelling story and I am not going to argue it.

CO-CHAIR KOTTKE: Liz.

MEMBER DELONG: I agree that they did a great analysis. I don't know why they didn't do that for the claims versus clinical.

CO-CHAIR KOTTKE: Okay, any other action? Seeing no further action, let's vote for validity.

MS. IBRAGIMOVA: Scientific Acceptability of Measure Properties: 2b,

validity. One, high; two, moderate; three, low; four, insufficient.

Missing one vote.

CO-CHAIR KOTTKE: Kristi is leaving early.

MS. IBRAGIMOVA: Tom, did you vote?

MEMBER JAMES: I voted twice.

MS. VICALE: Thanks, Tom.

I think we are still missing one vote in the room. If everyone would just hit their clicker one more time. Thank you.

MS. IBRAGIMOVA: The results are 29 percent high, 65 percent moderate, 6 percent low.

CO-CHAIR KOTTKE: Feasibility.

MEMBER HOLLANDER: Very feasible.

CO-CHAIR KOTTKE: Liz? Liz says yes. Full vote.

MS. IBRAGIMOVA: Just missing one vote.

The results are 88 percent high, 12 percent moderate.

CO-CHAIR KOTTKE: Usability and use.

MEMBER HOLLANDER: Very useable. I mean actually this is obviously publicly reported and very, very likely used.

CO-CHAIR KOTTKE: Anybody? Let's vote.

MS. IBRAGIMOVA: Usability and use. One, high; two, moderate; three, low; four, insufficient information.

The results are 88 percent high and 12 percent moderate.

CO-CHAIR KOTTKE: Okay, unless somebody has something to say, let's vote on the final measure. Recommend, yes; or --

MS. IBRAGIMOVA: Overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? One, yes; two, no.

The results are 100 percent yes, zero percent no.

CO-CHAIR KOTTKE: So, now our stewards get to present 0229, briefly.

MS. BERNHEIM: I am going to be very brief. Almost everything is the same with these

two measures. Obviously, the heart failure measure looks at heart failure.

The one important thing to note is that there is one additional change that we have made to this measure which is in the next year of public reporting, patients who are given an LVAD or a transplant during an index hospitalization or in the year prior would be excluded from the measure.

I think that is really -- and there hasn't been, unfortunately, as much improvement as there has been in AMI but there is some.

I think, otherwise, clearly, as you said it is the same.

CO-CHAIR KOTTKE: Thank you. Sana.

MEMBER AL-KHATIB: Nothing really much to add in terms of the evidence. Clearly, the evidence is very strong in support of the kind of measure for mortality in patients with heart failure. So, I think the evidence is very strong.

CO-CHAIR KOTTKE: Kristi has nothing to add. So, seeing nothing -- no other movement, let's vote.

MS. IBRAGIMOVA: Importance to Measure and Report: 1a, evidence, health outcome, support, PRO. One, yes; two, no.

MS. MARINELARENA: Before we get any further, do you all feel comfortable with the previous votes for the other measure or do you want to vote on evidence?

MEMBER HILLEGAS: Could you read the voting results for the evidence for the AMI measure and if people feel like it probably would be the same, we can carry over. Well just for evidence. For evidence, for 1a.

MS. IBRAGIMOVA: So for evidence for 0230, it was 94 percent yes, 6 percent no.

MEMBER AL-KHATIB: I would argue that actually those need to be voted separately. Those are two different clinical conditions. I think we need to have a separate vote.

MS. WILBON: Okay, that's fair.

MS. IBRAGIMOVA: The results are 100 percent yes, zero percent no.

CO-CHAIR KOTTKE: Opportunity for

improvement.

MEMBER AL-KHATIB: Well, yes, definitely, they provide some interesting data regarding the average 30-day risk-standardized heart failure mortality. They said the weight is about 11.7 percent during the measurement period between 2011 and 2014. And they said that there is a range of 7 percent to 19.3 percent. I felt like that was not a narrow range. I think there is a gap in here.

CO-CHAIR KOTTKE: Thank you. Any -- Mladen, but really quickly.

MEMBER VIDOVICH: Do the developers differentiate between the heart failure deserved and reduced at all?

MEMBER AL-KHATIB: I didn't see that distinction.

MS. BERNHEIM: They are both included in the measure.

CO-CHAIR KOTTKE: Okay. Opportunity for improvement. Let's vote.

MS. IBRAGIMOVA: Importance to Measure

and Report: 1b, performance gap. One, high; two, moderate; three, low; four, insufficient.

The results are 76 percent high, 24 percent moderate.

CO-CHAIR KOTTKE: So, now we accept that the science is all the same and then jump to the --

MS. WILBON: If the committee feels that the discussion would be duplicative in any way and they are comfortable, we can read the votes from science and reliability and validity from the previous measure, if everybody is okay with that.

MEMBER AL-KHATIB: Yes, I am certainly comfortable with that in terms of the methodology because it looked identical.

CO-CHAIR KOTTKE: Yes. And seeing no opposition here. So, we go to the final vote to endorse.

MS. IBRAGIMOVA: So just to read for the record, reliability is 18 percent high, 76 percent moderate, 6 percent low.

Validity is 29 percent high, 65 percent

moderate, 6 percent low.

Feasibility is 88 percent high, 12 percent moderate.

Usability and use is 88 percent high and 12 percent moderate.

So, now we will go to an overall vote. Overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? One, yes; two, no.

The results are 100 percent yes, zero percent no.

MS. VICALE: Thank you very much, measure developers.

At this time, we are going to open up the call and the meeting for member and public comment. Operator, if you could open up the line for those comments.

We are also noting that we are about almost 15 minutes past the original for member and public comment. So, we want to allow those folks to plenty of time to provide their comments.

OPERATOR: Okay, at this time, if you

would like to make a comment, please press * then the number 1 on your telephone keypad.

MS. VICALE: And if there are any comments in the room, we ask you to please come to the microphone and make your comments.

MS. IBRAGIMOVA: Are there any comments over the phone?

OPERATOR: No, ma'am, there are no comments at this time.

MS. VICALE: Additionally, we have received some comments throughout the afternoon via the chat window through the web platform. And at this time, we will go ahead and read off those comments. And just so you know, those are for all of the measure that we have reviewed for Phase 3 throughout the day.

MS. HERRING: Okay, all of the comments that we have received via chat were in regards to 2764, which was the measure about the fixed dose combination.

And our first comment was from David Mann, who said, so does this measure, to be in

compliance, require the prescription of BiDil product explicitly? Has any other NQF metric ever required the use of a specific proprietary product or a specific medication dosage?

He also said requiring the use of a single proprietary product is the most one-size-fits-all decision that could be made. It removes dose titration and dose modification for side effect control from what the metric will allow as quality care.

David also said, I don't object to trying to promote combination therapy in this group but I think saying that any other prescription except BiDil is not quality may go too far. I would like someone to raise this issue in the discussion later. Use of aspirin for anti-platelet effects is off-label use, isn't it?

And then I have some comments from Adolph Falcon, who said that I would like to state that I think the discussion of cost of best quality of care is not an argument against this measure. This committee should examine the evidence of what

is best quality care. If this measure demonstrates the best quality of care for African Americans, the cost discussion then belongs in cross-policy debates on the Hill and CMS to eliminate disparity of care.

We have another comment from David Mann, who says the precedent this will set is that if a drug company does a trial on its proprietary product, then quality metrics will require use of only proprietary targets. He said, does BiDil change outcomes compared to generic equivalents is the relevant question.

Sorry, we have quite a few comments in here.

He also said does the metric actually measure a concept that is essential for quality use of a specific proprietary product.

Back to Adolph Falcon, who said I think this process needs reconsideration from a meaningful stakeholder comment that could inform the consideration. It would be useful to have limited comments and questions during the

consideration of individual measures.

And that is all I see at this time but I can go back to it, if need be.

MS. WILBON: We can send those out via email to the committee as well or post them on the SharePoint page so you guys can access them.

DR. BURSTIN: And just since that issue was raised a couple of times, now, I did ask Reva Winkler, who has been on the call today, and how has been around NQF for the longest. There has been one other example of this in the early days of some of the early cardiac measures where it specifically referenced using clopidogrel when it was really only a single drug, no generic available yet but the evidence suggested it was the best at the time. Now, obviously, that has changed a lot over the years, there are generic forms available.

So, this is an issue we will continue to kind of vet and consider with all of you but there is, at least, a precedent.

MS. VICALE: Okay, thank you. We have a comment in the room.

DR. MASOUDI: Hi, Fred Masoudi. I am making this comment as a member of the public, not as representing anyone.

My only concern about that specific issue is not the expense of the fixed dose combination but rather the fact that the current heart failure guidelines do not make a distinction between the use of the fixed dose combination and the individual component drugs, irrespective of what the FDA has approved.

So, the current heart failure guidelines do not specify that hydralazine and nitrates need to be provided in a fixed dose combination to conform to the guideline recommendation, which I think is an important consideration.

MS. VICALE: Thank you. We have one more comment via the chat window.

MS. HERRING: This comment is from Paul Heidenreich. The ACC/AHA Guideline writing group had the option of limiting their Class I recommendation to the fixed dose and, instead, also

recommended use of generic hydralazine isosorbide, creating a performance measure that penalized providers for using generics as if they had used no therapy was a significant concern by members of the ACC/AHA performance measures task force.

MS. VICALE: Are there any other comments on the phone or in the room?

OPERATOR: There are not comments at this time.

MS. VICALE: Okay, thank you operator. So, quickly, before we do adjourn for the date, I would like to note, thank you all for the efficiency that we worked in this afternoon. We were able to end, actually, on time.

And I did want to note that tomorrow morning, we will be beginning at 8:00 a.m. for continental breakfast and the meeting will begin promptly at 8:30. And we hope to keep everything on schedule for tomorrow as well.

And if the committee would just remain for five quick minutes after, that will be it.

And I would like to just ask Mary

George, and Tom Kottke for their closing remarks of the day.

CO-CHAIR KOTTKE: Well thank you. And when and where is dinner?

Thanks. Good job.

CO-CHAIR GEORGE: Yes, thank you for hanging in there today.

MS. VICALE: The operator can end the webinar for the day. Thank you very much.

(Whereupon, the above-entitled matter went off the record at 5:05 p.m.)