NATIONAL QUALITY FORUM

+ + + + +

CARDIOVASCULAR MEASURE ENDORSEMENT PROJECT 2014 STANDING COMMITTEE

+ + + + +

THURSDAY
DECEMBER 4, 2014

+ + + + +

The 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
THOMAS KOTTKE, MD, MSPH, Co-Chair
SANA AL-KHATIB, MD, MHS, Duke University Medical Center
CAROL ALLRED, Immediate Past Chair, WomenHeart: The National Coalition of Women with Heart Disease
LINDA BRIGGS, DNP, George Washington University School of Nursing
LESLIE CHO, MD, Cleveland Clinic*
JOSEPH CLEVELAND, MD, University of Colorado-Denver
MICHAEL CROUCH, MD, MSPH, FAAFP, Texas A&M School of Medicine
ELIZABETH DeLONG, PhD, Duke University Medical Center
TED GIBBONS, MD, FACC, FACP, FASE, Harborview Medical Center; University of Washington Medical Center*
ELLEN HILLEGASS, PT, EdD, CCS, FAACVPR, FAPTA, American Physical Therapy Association
JUDD HOLLANDER, MD, FACEP, Sidney Kimmell

Medical College; Thomas Jefferson University
THOMAS JAMES, MD, AmeriHealth Caritas Family of Companies
JOEL MARRS, PharmD, FNLA, CLS, University of Colorado Anschutz Medical Campus; American Society of Health-System Pharmacists
GERARD MARTIN, MD, Children's National Health System
KRISTI MITCHELL, MPH, Avalere Health
GEORGE PHILIPPIDES, MD, Newton-Wellesley Hospital
JASON SPANGLER, MD, MPH, FACPM, Amgen, Inc.
HENRY TING, MD, MBA, New York-Presbyterian Hospital and Health System*
MLADEN VIDOVICH, MD, Jesse Brown VA Medical Center; University of Illinois at Chicago

NQF STAFF:
HELEN BURSTIN, MD, MPH, Chief Scientific Officer
SHARON HIBAY, RN, DNP, Senior Director
WUNMI ISIJOLA, MPH, Project Manager
KAREN JOHNSON, Senior Director
VY LUONG, Project Analyst

ALSO PRESENT:
JOSEPH ALLEN, American College of Cardiology
LISA BERGERSEN, MD, Boston Children's
KYLE CAMPBELL, PharmD, FMQAI*
STEPHEN CANTRILL, MD, Denver Health*
JENSEN CHIU, MHA, ACC/AHA
FRED KUSUMOTO, MD, Mayo Clinic
MICHAEL HO, MD, PhD, ACC/AHA
SAMANTHA TIERNEY, MPH, PCPI*

* present by teleconference

T-A-B-L-E O-F C-O-N-T-E-N-T-S

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

(9:05 a.m.)

MS. ISIJOLA: Good morning, everyone and thank you again for joining us. Today is day 1 of our Cardiovascular Phase II project. My name is Wunmi Isijola. I'm a senior project manager here. I also have Sharon Hibay, Karen Johnson and Helen Burstin, our chief scientific officer.

Would you like to say a few words?

DR. BURSTIN: Just add my welcome to everybody. Thanks so much for coming back again. I think you guys are one of our first standing committees, so I think part of what we'd like to see is how much easier is it to kind of do the second round having been through the first rounds, and is there a good steep learning curve? But also we also continuously try to improve on what we're doing, so please let us know if some of the tweaks we've made are helping or not. And I was just talking to Wunmi earlier about making sure we get you guys a survey as well to make

sure you can actually tell us what worked and what didn't.

But thank you so much for coming back again. And I think there are a couple process improvements including better clickers, which I'm told will not be the "could you push it one more -- one more time, please?" That should be over. And you'll know in fact whether you voted, so that will be very positive.

So with that, thank you so much.

MS. ISIJOLA: Okay. And also Vy Luong, our project analyst.

Before we get started I wanted to turn it over to our co-chairs to provide a brief introduction, Dr. Tom Kottke and Dr. Mary George.

CO-CHAIR KOTTKE: Sure. Tom Kottke, medical director for Population Health and Health Partners. Welcome and thank you for making the effort to come out. And I don't think I'll say any more right now.

CO-CHAIR GEORGE: Mary George from CDC. Just welcome. It's good to see you all

again and thank you for taking the time to do this important work.

MS. ISIJOLA: Thank you. Sharon, would you like to say a few words?

MS. HIBAY: Well, good morning. I think I introduced myself to mostly everyone. My name is Sharon Hibay. I'm the new senior director for the cardiovascular project. I've been with NQF for about two months, but I've been pretty involved in the cardiovascular space, the measure space for quite some time. Thank you for all your efforts. We appreciate them.

MS. ISIJOLA: And, Karen, would you like to say a few words?

MS. JOHNSON: I'll just say hello. I'm Karen. Nice to see you.

MS. ISIJOLA: Great. Thank you. So just some housekeeping rules. As you know, the restrooms are outside of this room. We will have several breaks throughout the day.

Please utilize the wireless network that we have available, and please make sure that

you're muting your cell phones and any other devices. We do ask also individuals on the line if you can mute your line and not place on hold.

And before we get started, I wanted to also make notice of Dr. Gerard Martin. He is one of our newer committee members.

So if you'd like to say a few words, a brief introduction of yourself?

MEMBER MARTIN: Gerard Martin. I'm a pediatric cardiologist at Children's National here in Washington, D.C., and I'm very pleased to join the Committee. Have been involved with quality efforts, both nationally and internationally, through our hospital and the American College of Cardiology for a number of years and excited to learn about your work.

MS. ISIJOLA: Great. Thank you. And now we'll turn it over to our general counsel, Ann Hammersmith, and she will go over our disclosure of interest policy.

MS. HAMMERSMITH: Thank you, Wunmi. As Wunmi said, we're going to combine

introductions with the disclosures of interest. For most of you it's your second year on the Committee, so I know you're expert disclosers, but I'll just go over a few of the ground rules to remind you.

I understand there are some committee members on the phone. I will call on you for your disclosure once we do the disclosures of the people in the room.

You received a fairly involved form from us where we asked you about your professional activities, community service, grants, so on and so forth. What we're looking for you to today is to disclose information that you believe is relevant to your service on the Committee. We don't want you to summarize your résumé. Please don't do that. We'll be here all day. But just reveal things that you believe are important, relevant to the work of the Committee that you want people to know and that people need to know.

So we are particularly interested in

grant activity, research support you may have received, Committee service and speaking. But only if it is relevant to what's before the Committee.

A few other reminders: You sit as an individual. You're here because you're an expert, because you have something special to bring to our table. You do not represent your employer. You do not represent the interests of anyone who many have nominated you or supported your nomination.

The other thing that I want to remind you of is that NQF is somewhat unique in our conflict of interest process in that we look at potential financial conflicts, but we also look at non-financial conflicts. And by that I mean you may have served on a committee for a professional society or somebody else where you looked at measures and that would be something that we would look for you to disclose but only if it's related to the subject matter before the Committee. Because of the nature of the work

that we do, we look very broadly to make sure that we cover the waterfront.

So I'll ask each of you to tell us who you are, who you're with and then if you have anything to disclose. And I always start with the co-chairs.

CO-CHAIR KOTTKE: Sure. Tom Kottke. Health Partners. I have no conflicts to declare.

CO-CHAIR GEORGE: Mary George at CDC and I have no conflicts of interest. Most of my major development work has been in stroke, cerebrovascular disease rather than cardiovascular.

MEMBER CLEVELAND: Good morning. I'm Joe Cleveland. I'm an adult cardiac surgeon at University of Colorado Health Sciences Center. I have participated in measure development through the Society of Thoracic Surgeons for surgical measures, but have no disclosures for any of these measures that are before us today.

MEMBER JAMES: Tom James. I'm an internist pediatrician with AmeriHealth Caritas,

a family of companies, a Medicaid-managed care company. I run the clinical policies. Formerly was on the AMA Cardiovascular Work Group, but that was some 10 years ago.

MEMBER MARTIN: I'm Gerard Martin. I'm representing the Children's Hospital Association. My disclosures are the American Board of Internal Medicine, where I'm on their sub-board for adult congenital heart disease and the American College of Cardiology, where I participated in the formation of the IMPACT Registry which measures quality income in cardiac catheterization and the quality network of the ACC. It's a set of measures to improve quality in pediatric cardiology practices.

MS. HAMMERSMITH: Just a reminder: You sit as individuals. You're not representing an organization or their interests. So I just want to highlight that.

So you may work for the ABC Hospital, but you're not representing their interests on the Committee.

MEMBER VIDOVICH: I'm Mladen Vidovich. I'm an interventional cardiologist, University of Illinois at Chicago and I also am chief of cardiology Jesse Brown VA in Chicago. I don't represent any of those, so not the Department of Veterans Affairs or University of Illinois. I am the upcoming governor-elect of the Department of Veterans Affairs for ACC, but again I don't represent any of those.

MEMBER HOLLANDER: Hi, Judd Hollander. I'm an emergency physician at Thomas Jefferson University. I got here through ACEP and I'm on their Quality and Performance Committee. I don't think there's any direct conflicts with anything.

MEMBER CROUCH: I'm Michael Crouch. I'm a family physician representing the American Academy of Family Physicians. I'm in a family medicine residency program. Don't have any conflicts of interest to report.

MS. HAMMERSMITH: I hate to beat a dead horse, but I'm going to because I'm a lawyer. You don't represent any organization

when you sit on this Committee. You're here as an expert, so you're not carrying anybody's water. You're here because of what you know as a professional, as an individual. So I just want to remind you.

CO-CHAIR KOTTKE: Yes, so don't use the R word.

MEMBER SPANGLER: Jason Spangler. I'm the Executive Director of Medical Policy and Quality Strategy at Amgen, and I don't have any conflicts.

MEMBER DeLONG: I'm Liz DeLong. I'm the chair of Biostatistics and Bioinfomatics at Duke. I have worked with respect to data analysis, both with the NCDR and STS, but I have no conflicts.

MEMBER ALLRED: I'm Carol Allred. I have just finished a term serving six years as chairman of the board of directors at WomenHeart. I come at this from a different perspective than anyone else in this room, I think; i.e., I'm a patient. I am very interested in many of the

measures because they affect me directly and many people I know.

MEMBER PHILIPPIDES: Good morning. My name is George Philippides. I'm a cardiologist and chief of cardiology at Newtown-Wellesley Hospital in Massachusetts. I'm the president-elect for the Founders Board of the American Heart Association, but I have nothing to disclose.

MEMBER MITCHELL: Good morning. I'm Kristi Mitchell. I'm a senior vice-president at Avalere Health. I don't have anything to disclose except for the fact that I worked 12 years at the American College of Cardiology.

MEMBER HILLEGASS: I'm Ellen Hillegass. I'm a physical therapist. I'm on faculty at Mercer University in Atlanta, Georgia and I have no conflicts to disclose.

MEMBER AL-KHATIB: Good morning. I'm Sana Al-Khatib. I'm an electrophysiologist at Duke University. I actually serve on the steering committee for the ACC NCDR, but I don't

have any conflicts in relation to that because I did not participate in their performance measures. I am a member of the board of trustees for the Heart Rhythm Society, and I co-chair their Measure Development Task Force, so I do have a direct conflict with two of the measures that they developed and I'll stay calm and quiet while you discuss them.

MEMBER BRIGGS: Hi, I'm Linda Briggs. I'm faculty at George Washington University here in D.C. I am particularly in the nurse practitioner group, and I am a nurse practitioner myself. I have no conflicts of interest.

MS. HAMMERSMITH: Okay. Thank you. I'm going to call the names of the people who are on the phone so that you can introduce yourselves and disclose anything you want to disclose. Leslie Cho?

MEMBER CHO: Hi, it's Leslie Cho. I'm from Cleveland Clinic and nothing to disclose.

MS. HAMMERSMITH: Okay. Thank you. Ted Gibbons? Is Ted Gibbons on the line?

(No response.)

MS. HAMMERSMITH: Henry Ting?

(No response.)

MS. HAMMERSMITH: Have a little feedback here. Is Henry Ting on the line?

OPERATOR: Sorry, Ann, I didn't mean to step on you. Just a reminder for those folks joining us by phone, please make sure you keep your computer speakers turned down or off to reduce feedback.

MS. HAMMERSMITH: Thank you. Joel Marrs?

MEMBER MARRS: Hi, I'm Joel Marrs. I'm on faculty at the University of Colorado and a clinical pharmacist. I have nothing to disclose.

MS. HAMMERSMITH: Thank you. Thank you for making those disclosures. Just a few final words before I go away and let you get on with your work. For any conflict of interest process to really work, everybody has to take some responsibility for it, including the

Committee members. So if while you're discussing/deliberating you think you have a conflict of interest, you think somebody on the Committee has an conflict of interest or is behaving in a way that's biased, we look to you to speak up, preferably in real time.

You can do that by bringing it up directly at any time. If you don't want to do that, you can go to your co-chairs, who will then go to NQF staff, or you can go directly to NQF staff so that we can deal with it on the spot. We don't want you sitting there in silence thinking, ooh, I think I may have a conflict, or, gee, that person seems really, really biased. We really want you to speak up.

So having said all that, do you have any questions of each other, anything you want to discuss, any questions of me?

(No response.)

MS. HAMMERSMITH: Okay. Thank you.

MS. ISIJOLA: Thank you, Ann. So just to kind of provide additional feedback based on

what Ann spoke to and just a refresher for your role as Standing Committee members, we do ask you not to represent your organization but yourselves. You are providing your expertise based on an NQF multi-stakeholder perspective.

And you all have received your two or three-year term. If you have any issues with that, please do let us know. And really your charge is to work with NQF and the goals of this project, which is essentially to review and endorse these measures based on our criteria.

And also, we do at times have responses from CSAC and for you, specifically the co-chairs, to respond to any requests based on the measures within this portfolio, but also oversee the portfolio of cardiovascular measures.

And as I mentioned, your role is to oversee the cardiovascular portfolio, and Sharon Hibay will review that with you shortly, but really consider the issues and the measures within this portfolio based on our criteria, but to identify if there are in fact any gaps,

measures that may potentially be relating or competing with various measures within our portfolio.

So with that being said, I will turn it over to Sharon Hibay and she'll speak more to some of the measures we have in this portfolio as it relates to our Phase I of this project and some of the measures that you've in fact endorsed.

Sharon?

MS. HIBAY: Next slide, please. Okay. So what you have in front of you here is the list of measures that we're going to be talking about in today's project. There are 16 of them that we'll be reviewing over the next couple of days.

Next slide, please. So of course we're all reasonably familiar, I hope so, with the National Quality Strategy and the National Quality Strategy priorities. Of course they're based upon the triple aim: better care, healthy people/healthy communities and affordable care. So much of the work that the whole health care

industry enterprise is doing is all around building healthier communities and having better outcomes.

So one of the National Quality Strategies very specifically calls out cardiovascular disease, promoting the most effective prevention and treatment practices for the leading causes of mortality, starting with cardiovascular disease. So it's the only disease specifically that was actually called out in the National Quality Strategy priorities. So we have some important work to continue to do.

Next slide, please. So I'm not going to go into great detail. There's a whole bunch of slides that are going to follow. We talked about this in Phase I. At that point, Reva Winkler provided a really nice overview of the work that we're doing that are reflective of the cardiovascular measures in our portfolio. Just in general they are coronary artery disease, AMI, heart failure and rhythm disorders. And then there are some other measures related to cardiac

cath, hypertension and cost and resource use.

Not all of the measures that are in the 70 measures within the CD portfolio will be reviewed by this Committee. They're reviewed by -- throughout all of the projects in different committees at NQF. So the preponderance are within this Committee, but they also are pertaining to some other committees as well.

Okay. So this is a framework that you saw, and it was I think also in the final report of Phase I, which is up and posted and I'm sure you've all taken a look at the good work that you have already done, and that now we continue. This is patient-focused episode of care framework for CAD and AMI. We're all very familiar with the framework, which is populations at risk. We talk about those. So it's either primary prevention, secondary prevention and with or without an AMI.

So we have focused on staying healthy, getting better, living with illness and disability and coping with end of life. We start

with your primary prevention. And if you have an event or an episode or a disease, you move to some sort of acute phase, post-acute/rehabilitation phase and then secondary prevention.

So what we've provided for you in the list of some of the subsequent slides are the measures that are within NQF's portfolio that represent these categories that I just kind of referenced. So we have these here for your reference. I'm going to go through some slides reasonably quick, but the way we have it formatted will help you identify the ones that we're going to be reviewing in these two days of discussions.

So starting with populations at risk, primary prevention. So we have a number of slides, and you can see three of the cardiac imaging ones we'll be discussing at this meeting.

Next populations at risk, secondary prevention. And one of those as well will be reviewed/bolded, you can see, at this meeting.

Now we move onto the acute phase AMI, and you can see we have one measure here as well.

Acute phase outcomes. We won't be reviewing any of these measures today, but again these slides are provided for your reference.

Next we talk about PCI. Again, not today.

And then for CABG patients these are in the surgery portfolio, but obviously very related to cardiovascular disease. These are not measures that we will be seeing.

And then post-acute/rehabilitation phase. And again, we will not be reviewing any of these today.

Populations at risk, secondary prevention for patients with CAD/AMI. And you can see a measure at the bottom.

Okay. So that is the CAD/AMI list. Now we kind of move on to heart failure. Excuse this slide; it's a little bit fuzzy. So basically we have populations at risk again for heart failure. Evaluation and ongoing

management. And then the acute phase and hospitalization. And then we move to outcomes. So we're going to show you the same format.

So populations at risk. We will not be looking at these measures today.

Evaluation and ongoing management. Oh, actually we're looking at one of the measures on there. It's 0083. So that one should be highlighted. Excuse me.

Acute phase hospitalization. We're looking at 2455 on this slide.

And these are now related to heart rhythm disorders. So A-fib, you can see we're looking at a couple of those.

And these are measures related to cardiac cath, and we'll be looking at one of those. That's a pediatric one.

And we have a measure for hypertension. Not today. Actually I think that's a population health measure as well, so I think that goes to another committee. I think. I don't know all of the measures yet. I'm on my

own steep learning curve; it's only the beginning of month three for me here at NQF.

And lastly is a cost and resource use, and that is also another committee as well. Okay?

Are there any questions on the portfolio or the information that has been presented so far?

(No response.)

MS. HIBAY: Okay. I'd like to move on to talk a little bit about updates to the process for measure evaluation. We did some work. Recently we talked a little bit about this again at the Q&A call that we hosted a couple weeks ago. Starting off in Phase I of the CD project, the staff did reviews of the measures. There was a call for multiple different approaches for us to provide some additional information, the staff to do some sort of look-see at the measure details.

We did some work with the consensus task for it which was approved by the Board. We

also had lots of our stakeholders, including members of this Committee as well, come to NQF and say, you know, this is a big bite to chew on. Can you have a little bit more predigested information for us?

We heard definitely from measure developers as well that they would like to have a little bit more supportive understanding of what will constitute the content of the conversation? And we also heard that people wanted what was perceived as a little more even playing field, that we all kind of started off on the same footing and made sure we talked about the measures in a consistent manner.

Lastly, we recognized that being an organization that convenes multi-approached stakeholders that we all come to the table with different gifts and different knowledge/ experience and that the idea of convening a stakeholder group brings all of that richness together, but that richness means that we have areas of strength and areas that we could learn

or need a little more understanding.

So all of this together, we put together the staff review that I believe we did for all of the measures in Phase 1, and we went back after Phase 1 and kind of tweaked it. The staff reviews go through the five criteria themselves, and we provide kind of a pre-look from the staff perspective. And after we tweaked it, now the preliminary analysis which I understand is some alignment work with the MAP, that they also do some preliminary analysis there as well.

So we provided this information to the members of the Committee, and we also provided it to our developers. Then the Committee went through the process of reviewing the individual measures they were assigned as discussants. And then, as you all know, you added that information in your surveys so we could put this information, coordinate it one more time together. And the information that you have on the measures now includes the preliminary analysis, comments from

the staff members of NQF, and as well as your comments. And we also sent those back out again to the measure developers so they had an understanding of what we might be talking about today in relation to their measures individually.

Like everything that we do -- we all live in quality improvement, so we know that the journey doesn't end. So this is a process. And so we will be seeking your feedback, as Helen said earlier, as to how these changes are going, how the process is working. We're very interested in making sure we continue to move this along and be as efficient and helpful as possible.

We're also going to be after this Committee meeting reaching out; I will be personally reaching out to each one of the developers to see what did they think of this preliminary analysis process? So the idea of this is to expand opportunities for robust discussion.

I'm not going to read this slide here,

but the last thing I really want to enhance is that by no means is this staff review/preliminary analysis to replace the great work that the Committee members do. And I have to tell you -- so I spent a lot of hours last night looking at every single form. I'm very impressed with the work that they did; I'm quite pleased. It's quite obvious the detail and time that everyone has taken. So I thank you in advance for all of your efforts. But again, this is not supposed to replace; it's supposed to enhance and start as a springboard for conversation -- a consistent springboard, as you will.

So the other thing that I want to share with you that you all have access to, the Committee members, is the wonderful information provided to us by the measure developers. When they submit their measures, we go through what's called a completeness check. And we just want to know that all the documentations that we need to review for review and evaluation are available and then we compile that information. The

information you looked at was the measure work sheet, which had the preliminary analysis. If there was an e-measure that has that information as well, we did an e-measure review.

You have the pre-evaluation comments and you have pre- and public member comments. Just to put that out as a caveat, we actually for all of these measures received no comments from the public before the meeting.

We also have the measure information form and we have the evidence and testing attachments. And we tried to put it in an order that was easy for you to utilize. So we'll see how that goes as well. And we're very interested in hearing your feedback. We also have spreadsheets and any other additional documentation they may have provided.

So that's kind of the structure of what we did, why we did it and the background of this continued evolving work.

Are there any questions about the preliminary analysis at this point, or the

updates to the evaluation process?

Okay. Hearing no comments or questions, I'll turn it over to Wunmi for the next slide.

MS. ISIJOLA: Okay. Thank you, Sharon. So most of you are familiar with our process and how we will proceed with today's deliberations. We are expecting every committee member to have reviewed every single measure. I know you have been designated certain measures to provide your evaluations, but we want a robust discussion; we want to make sure that everyone is participating in the discussion, and obviously Dr. George and Dr. Kottke will facilitate that. And really to remain engaged during the discussion without distractions. I know everyone has their laptops and busy, but let's really remain engaged with the discussion for today.

And we do have discussants for each measure. We specifically did not identify a lead discussant because we want to make sure that everyone is engaged in the discussion. As you

know, you have your microphones. Please utilize them. Speak into them. If you're not speaking, please turn them off.

And the way we will proceed, and I'll turn it over to the co-chairs in a second, we will have the developers present a brief introduction of their measures, two to three minutes, and then we will turn it over to the discussants to provide their analyses of their measures. And we will open it up for discussion. Thereafter, we will then vote on each individual criteria of the measure and then an overall recommendation for endorsement.

And with that being said, I'll turn it over to the co-chairs to begin.

CO-CHAIR GEORGE: I would just add that it would be helpful if you want to make a comment to turn your name card vertical so we'll know to call on you. And try not to interrupt each other. For those on the phone, be sure and chime in when you have comments to make.

CO-CHAIR KOTTKE: So, now we get down

to business. Measure 0543, coronary artery disease and medication possession ratio for statin therapy. And the discussants are Sana and Kristi and Liz. And who's going to take the driver's seat? Sana?

MS. ISIJOLA: And, I'm sorry, are there any representatives from ACC with us today?

DR. CAMPBELL: Wunmi, this is Kyle Campbell from FMQAI. Can you hear me okay?

MS. ISIJOLA: Yes.

DR. CAMPBELL: Okay. Good. All right. Well, good morning. My name is Kyle Campbell and I'm a pharmacist and vice president for pharmacy and quality measurement at FMQAI. We are the measure developer and we're representing CMS today.

This particular measure was developed for CMS under the Medication Measure Special Innovation Project and it was originally fully endorsed in 2011. The measure has recently undergone a comprehensive review of underlying evidence and additional testing of the revised

specifications at several levels of analysis, including the accountable care organization level. The project was under the direction of a multi-disciplinary technical expert panel.

Major changes to highlight for you of the measure are the results of the process to align the eligible population with the 2013 ACC/AHA guidelines for the management of cholesterol. So, typically, rather than the denominator's eligible population being limited to those with coronary artery disease, the denominator has been expanded to include patients with atherosclerotic cardiovascular disease as defined in the guidelines.

In addition, the age criteria has also been modified to align with the guidelines. Therefore, the denominator is individuals of at least 21 years of age with cardiovascular disease presumed to be of atherosclerotic origin and at least two claims for statins during the measurement period.

And the numerator is individuals with

CVD who had at least two prescription drug claims for statins and have a proportion of days covered for statin medications of at least 0.8.

In terms of importance, this measure clearly addresses the National Quality Strategy goal of promoting the most effective treatment for leading causes of mortality, starting with cardiovascular disease. As you know, CVD affects large numbers, represents a leading cause of morbidity and mortality and has high resource use on the healthcare system.

Based on the data analyzed, the measure has ample room for improvement with mean rates of approximately 72 percent at each of the levels that we looked at. Additionally, with the new guidelines shifting away from the use of specific clinical targets, medication adherence to statins is an important concept for providers to understand and communicate to their patients.

The measure was found to be reliable and valid, and the method of measuring adherence as the proportion of days covered is harmonized

with the majority of adherence measures in the NQF portfolio.

We greatly appreciate your consideration of the measure today and look forward to answering any questions from the Committee concerning the measure submission. Thank you.

CO-CHAIR KOTTKE: Okay, now.

MEMBER AL-KHATIB: Okay. So, as we heard, in terms of description of the measure, introduction of the measure, it's the percentage of individuals with cardiovascular disease, including CAD, cerebrovascular disease, PAD presumed to be of atherosclerotic origin who are prescribed statin therapy that had a proportion of days covered for statin medications of at least .8 during the measurement period.

The level of analysis is the clinician, group practice, health plan, integrated delivery system, population, state. And as you heard, this actually was initially endorsed in 2009 and is being revised to

incorporate some of the emerging data and the guideline document that the developer referred to.

So this is a process measure and the hope is that this measure will help physicians identify patients who may benefit from a statin who are not adherent to it, may help providers develop communication and education tools to improve adherence to statins. And they clearly provide the argument that higher statin adherence rates are expected to result in lower rates of hyperlipidemia, cardiovascular events and mortality. And so they certainly state that adoption of this performance measure will improve the quality of care of patients and eventually improve outcomes.

In terms of the evidence that they provide, I think they did an excellent job providing this evidence referring us to the 2013 ACC/AHA guidelines, but also summarizing the results of a 2010 meta-analysis, which was a systematic review of 21 randomized control

trials. They also reported a summary of eight more recent studies of adherence to statins that consistently found high adherence to statins in persons with CAD was associated with lower all-cause mortality. So when it comes to the evidence, I think that the developer did an excellent job there.

I'm not sure if you want to move on to the other aspects or if you want to hear from other people?

CO-CHAIR KOTTKE: Anybody else want to comment at this time? Kristi or Liz?

MEMBER MITCHELL: I just have a question.

CO-CHAIR KOTTKE: Kristi? Yeah.

MEMBER MITCHELL: Yes, sorry. Again, being a non-clinician, I just need a point of clarification regarding the use of the term "cardiovascular" being inclusive of peripheral vascular disease or peripheral arterial disease. Because the documentation, I was a bit confused by that. I did appreciate the ASTVD reference.

I think that was very helpful, but I was just curious again as sort of a consumer as to whether or not PAD is inclusive within CVD or CAD.

CO-CHAIR KOTTKE: Yes.

MEMBER AL-KHATIB: Maybe the developer would want to comment. But, yes, so this is what the developer was talking about, that now the definition of cardiovascular disease includes either CAD or cerebrovascular disease or peripheral arterial disease because the pathophysiology is the same in terms of having atherosclerosis. But maybe the developer would want to add to that as well.

DR. CAMPBELL: Sure. Yeah, I think, just consistent with everything that's been said, one of the things that we looked at in the titling of the measure was to use the simple term of cardiovascular disease because we were concerned that if we used, like, arteriosclerotic cardiovascular disease, patients might be confused. And so that's why we limited the name of the title in the description.

CO-CHAIR KOTTKE: Thank you. Liz, do you want to --

MEMBER CHO: I'd at least consider, you know, that the guideline says it's coronary artery disease equivalent. Is that part of the population?

CO-CHAIR KOTTKE: Sana? The question was, was coronary artery disease equivalent part of the guideline? Diabetes, for example. I think the answer is no, that it's patients with established disease.

MEMBER AL-KHATIB: That's exactly right.

CO-CHAIR KOTTKE: Yeah, so it's not -- I assume -- was that you, Leslie?

MEMBER CHO: Yeah. Yeah, it's me.

CO-CHAIR KOTTKE: Yeah.

MEMBER CHO: Hi.

CO-CHAIR KOTTKE: Folks on the line, if you'd address yourselves until we get used to your voices, that will help.

It doesn't include CAD equivalents.

It's established disease?

MEMBER AL-KHATIB: Right.

MEMBER HOLLANDER: Yeah, I have sort of a philosophical question on this. And I can't tell from reading the measure, it's measured at what level? Like I see it's ACO and health plans, but it also appears to list just group practices. And so a lot of these adherence things, it seems to me that the individual clinician has very little control. So if you give somebody a prescription and you give them refills for six months and you're not supposed to see them for six months, how do you as an individual practice have any idea if they've filled their scripts at two months?

So I see the measure where it's at the health plan level and they have access to the data and can set up quality controls, but I frankly feel it's a totally inappropriate measure at the group practice level because it will just drive up costs. It will require them to see patients more often than they do it and they're

never going to develop a robust infrastructure to monitor refills between visits. And so I raise that question. I think this may be the only measure that we're seeing today that that's an issue in, but I know we discussed it last time.

MEMBER AL-KHATIB: I mean, this measure has been in use, so maybe the developer can share some information with us regarding how this has been done since 2009.

DR. CAMPBELL: Yeah, so this is Kyle Campbell again. The measure has been in use with the CMS QRUR Program where reports have been sent out to individual group practices to advise them of the measure rates related to adherence. The measure would be, as we specified, useful at all levels, but the physician groups would be limited to, based on the reliability, very large physician groups. And oftentimes those groups do have the resources to provide interventions for the patients in terms of education related to adherence. And so I think for that reason we would consider it appropriate at the physician

group level.

MEMBER AL-KHATIB: And I would also add, especially now that we have more widespread use of EMR, I think this is something that could potentially be accomplished within a health system where you can embed certain tools within the EMR to track, you know, at least that the name of the medication is on their list. Now, whether they're taking it or not is a different issue, and maybe that's what you were trying to get at.

DR. CAMPBELL: Yeah. No, just to clarify, I'm all good with large group practices in places that have the resources. I couldn't see that clearly delineated in here. I was not a primary reviewer. It may be in there. It would be nice to see a definition of large group practice, because if we're going to have it as a measure, somebody may consider it five and somebody may consider it 50 physicians. So I would say, going forward, that would be sort of a friendly amendment to the measure to define who

it applies to. I think it's very hard to have a measure without a clearer definition.

MEMBER DeLONG: I don't know if we're getting ahead of ourselves in terms of the criteria, but that was one of my concerns. If you look at the reliability specifications, there are several levels. And I can't remember the exact numbers, but group practices had to have a certain number in order to be considered reliable for this measure. But that creates a fair amount of bias in terms of who's being evaluated. I don't know if you want to get into that now, or I've got comments later.

CO-CHAIR KOTTKE: Tom?

MEMBER JAMES: Thank you. I had one comment, but since Judd, who practices right down the street from where I live, raised an issue, as a primary care physician, part of what I've been involved with in my practice is trying to improve the adherence rates, and there is a lot of techniques that have been done. I've discussed this with the ACP. They don't have any problem

with looking at the skill of physicians in being able to improve adherence rates.

But the second part, from working within health plans -- and I think, Tom, that you're very much aware of this -- is this whole idea of the units for measurement in terms of the pharmacy is now being challenged. It's not just the fill rate, but it's the utilization. And a lot of organizations are now looking at different methods, such as pill counters, to be able to get to this. We're not at that point in general, so this is fine as far as I'm concerned for now, but I suggest to the developer that they look at these newer techniques for the future.

CO-CHAIR KOTTKE: Thanks. Does anybody want to go back and address Liz' comment? Do you want to restate that, Liz?

MEMBER DeLONG: Well, the complete statement would be that there are a number of levels that they have dealt with, and as they analyzed the data, they found that I think over 50 percent of physician groups did not have

adequate numbers to be considered reliable. So they're not evaluating all physician groups. And I guess the group is fine with evaluating some and not others.

I think this measure has been in use since 2009, and they didn't produce any evidence that it has been effective, that it's improved care or adherence. It seems to be operating in a bit of a vacuum.

DR. CAMPBELL: This is Kyle Campbell for the measure developer. Just to address a couple of those comments, which I think we get to later in the evaluation, the minimum denominator threshold that we established for reliable measurement at the physician group level was at least 250 eligible patients. So it gets back to the original comment that, should the measure be publicly reported, it would be appropriate for large physician group practices.

CMS didn't implement the measure until, I believe, approximately 2011 in the QRUR program. And in the QRUR program, those measure

rates are not publicly reported. So they are provided to physician groups that are smaller than that, since the data are not comparative, but they're meant for internal quality improvement within the practices. We don't have yet any trend data from that program to indicate improvement that has occurred. We just don't have that data available to us yet.

CO-CHAIR KOTTKE: Thank you. This issue of the number of patients seen, I think this is a cross-cutting issue across all measures, that if you simply don't see enough patients in any particular area, you're not evaluated. And so, I mean, every single measure. So I mean, I'm comfortable with that. If you only have a few patients, we understand it's very unreliable and therefore you're out. And so the way to get out of the measure is not to see patients that apply to that measure. If that's what you want to do, you can do it.

Any other comments on evidence?

MEMBER DeLONG: Actually I think they

mentioned some disparities there. I don't know if that comes under --

(Off-microphone comment.)

MEMBER DeLONG: Okay.

MS. LUONG: So, for evidence, I'm going to start again with the ratings. One is for high, two is for moderate, three is for low, four is for insufficient evidence, and five for is insufficient evidence with exception. And the voting starts now.

(Voting.)

MEMBER CHO: For the people on the phone, how do we vote? It's Leslie Cho.

MS. ISIJOLA: Hi, Leslie. Could you provide your vote in the chatting tool?

MEMBER CHO: In the chatting tool?

MS. ISIJOLA: The chat box in the webinar. Or email.

MEMBER CHO: The chat box appears to be not working. It just says "welcome," and it won't let me talk. Oh, okay. Hold on.

MS. LUONG: Can everyone just point

towards me, just to make sure? Sorry, I should have reminded you. You should see your vote number. Thanks.

Okay. The ones that voted, you don't have to vote again, but I'm just going to reopen the polls so that -- do you have another one?

MEMBER CHO: Wunmi and Vy, did you get my vote?

MS. LUONG: Wunmi, did you get the vote? We are checking right now, Leslie.

MEMBER CHO: Okay. Can you put up the criteria on the webinar again?

MS. LUONG: Sure. So the criteria again, Leslie, is one for high, two for moderate, three for low, four for insufficient evidence, and five for insufficient evidence with exception. And this is for evidence.

Wait, let me do it again. I didn't know you were still voting. Go.

So we have 47 percent for high and 53 for moderate.

CO-CHAIR KOTTKE: Okay. Sana, do you

want to move on?

MEMBER AL-KHATIB: Yeah, so, for opportunity for improvement, the developer presented results from 10 states where they looked at 38 prescription drug plans and 434 physician groups and 31 ACOs. And as the developer stated early on, compared to the average measure result for all patients, they found the average to be 70.4 percent.

And since we're hoping that that will be much higher, I think there's certainly an opportunity for improvement. And here's where they presented the results on disparities, where they talked about the rates for African-Americans and Hispanics are lower, 58 percent and 60.4, respectively.

And, so, while I completely agree with Liz that it would be important to share data on the impact of the measure and whether it has led to improvement -- I would love to see that one day for all the measures. You know, I don't know that we should hold this developer to a higher

standard than all the other developers that we interacted with in Phase 1, where none of them, as best as I can recall, provided information about the impact of the measure.

In terms of priority, I think it's definitely very important, a very important disease, very prevalent, and the evidence supporting the use of statins in this patient population is very robust. So I don't have any concerns about the priority.

CO-CHAIR KOTTKE: Liz?

MEMBER DeLONG: I'm afraid I have to disagree with my colleague Sana. If a measure has been in use for several years, it seems there should be evidence of its impact. And as we accumulate more and more measures, I worry that we're going to flood the market with measures that haven't really proven out. So that's my caution.

I don't understand completely who has Part D insurance, but you have to have Part D in order to be evaluated with this measure. Is that

not a concern? I mean, I confess ignorance here.

CO-CHAIR KOTTKE: Does the measure developer wish to address the Part D question?

DR. CAMPBELL: Sure. This is Kyle Campbell again. So, this measure, the eligible population, based on administrative data that's available to us, includes the patients with Part D and that are fee-for-service. So this measure does not include patients that are in an MAPD or management care under Part D. This would be limited to fee-for-service Medicare patients with Part D coverage. And that limitation is simply the limitation of the administrative data source upon which the measure is based.

CO-CHAIR KOTTKE: Right. The short answer is there's data on fee-for-service Part D patients. There's no data otherwise. And the cost of collecting data otherwise is prohibitive.

DR. CAMPBELL: Correct.

MEMBER DeLONG: Well, when we talk about disparities, there is an implicit disparity in terms of requiring Part D insurance for this

measure, right?

CO-CHAIR KOTTKE: Absolutely. I think there's no question that probably somebody -- you know, an African-American who has no health insurance is less likely, but we can set the bar on measures too high and have nothing.

Any further questions? Do we vote or --

MEMBER JAMES: One clarification, because it notes that health plans are a part of the accountability structure, so that would take it beyond just fee-for-service Medicare.

CO-CHAIR KOTTKE: Okay.

MS. LUONG: So we are voting now on performance gap. You can vote one for high, two for moderate, three for low, and four for insufficient. If you can just point towards me and vote. Thank you.

(Voting.)

MS. LUONG: So for this performance gap criteria we have 37 percent high, 53 percent moderate, and 11 percent low.

CO-CHAIR KOTTKE: Sana?

MS. HIBAY: Excuse me. Can you provide the numbers, the actual count as well or --

MS. LUONG: We can provide the number of the actual count after the report is generated with this tool.

MS. HIBAY: Okay. Thank you.

MEMBER AL-KHATIB: Okay. So, moving onto scientific acceptability specifications. So the numerator is individuals with CVD who had at least two prescription drug claims for statins and have a PDC for statin medications of at least .8. Denominator is individuals at least 21 years of age, as of the beginning of the measurement period, with CVD, including CAD, cerebrovascular disease, PAD presumed to be of arteriosclerotic origin, and at least two claims for statins during the measurement period, in a 12 consecutive month period.

For denominator exclusions, they said not applicable. And here's where I had a

question for the developer. What about patients with contraindications? I mean, I realize this is using claims data, so it may not be easy to find those patients. But, I mean, I don't expect that this rate will ever be 100 percent, because you're going to have patients who have contraindications to statins or who have allergies, intolerance, what have you. How do you envision that being incorporated once this measure starts getting applied in other settings?

DR. CAMPBELL: Sure. Kyle Campbell again. Thank you for the question. Because the measure denominator actually requires at least two fills by the patient, we feel like that does confirm the physician's intent to continue the medication. So any severe allergic reaction or intolerance to a statin would be identified most likely before the second fill.

In terms of absolute contraindications, you know, statins are contraindicated in pregnancy, but in our particular data set, the prevalence of pregnancy

is extremely low, you know, much, much less than one percent. There are very few pregnancies among patients in this data set. So for that reason we didn't specify any exclusions for the measure.

MEMBER DeLONG: I'm a bit confused. I'm always confused. But maybe somebody can explain why the selection of patients claims to be 21 and over, but they have to have Medicare coverage. I thought you couldn't have Medicare coverage unless you were ESRD or a special population. So how does this cover those people between 21 and 65? Also you had to have two prescription refills. Well, doesn't that eliminate a number of people who got one filled right there at the pharmacy, at the hospital or something, and then never went back? I would think that would be a lack of adherence.

CO-CHAIR KOTTKE: Developer?

DR. CAMPBELL: Kyle Campbell again. Yeah, appreciate the question. So the first question is, in the Medicare population, patients

can actually be dual-eligible, and those patients are included in here. So they would be Medicaid/Medicare. And that represents, in our data set, approximately 20 percent of the population.

In terms of the two-fill requirement, I think it really was a trade-off in terms of harmonization and the concern that was brought up about intolerance to therapy. The majority of measures in the NQF portfolio that measure adherence use the two-fill requirement in the denominator. And, again, we didn't want to unnecessarily penalize folks if they had tried a statin and for whatever reason had an intolerance to it and the physician didn't decide to continue it.

CO-CHAIR KOTTKE: Right, so the denominator is smaller than all people who ought to be taking a statin, but it also is a nod to the problem that there are patients who simply don't tolerate them. So this is looking at people who ought to be taking -- you know,

there's good evidence they ought to be taking a statin. And the question is, are they? So, it's smaller than the totality of the population that ought to be on a statin and would tolerate a statin.

Further comments?

(No response.)

MS. LUONG: So we will vote now on high priority. Number one is for high, number two is for moderate, three is for low, and four is for insufficient. And the polling starts now.

(Voting.)

MS. LUONG: Fifty-six percent voted high, thirty-three voted moderate, and eleven voted low for high priority.

CO-CHAIR KOTTKE: Your mic is not on. There, now it's on.

MEMBER AL-KHATIB: All right. So we talked about the numerator, denominator and denominator exclusions with regard to the data source. So for this measure the data source is encounter and pharmacy claims. So they used ICD-

9 and ICD-10 codes, you know, provided to identify patients with these conditions under cerebrovascular disease, CAD and PAD.

The proportion of days covered. This is the PDC method that they used. It is a commonly used calculation, is what they said, of medication adherence, patient compliance. And this is calculated through pharmacy claims and they said that seven statin medications and several combinations are specified, and indeed they did that. And a calculation algorithm is provided in the document that they submitted as well.

In terms of any issues/concerns we may have about the specifications, definitions or coding, I mean, the only concern I would have is the accuracy of the coding. I'm not sure if any of these ICD-9 and ICD-10 codes have been validated in terms of their accuracy. But I suspect that the validity is on the high side, but if the developer has any information about validation of those codes, that would be helpful.

DR. CAMPBELL: Yeah, we don't have any additional information to provide with regard to the validity of the claim data other than to say that it is in keeping with other cardiovascular measures in the NQF portfolio in terms of harmonization of coding.

CO-CHAIR KOTTKE: Further comments?

(No response.)

MS. LUONG: We will be voting on Criteria 2(a)1 on reliability. One for high, two for moderate, three for low, and for four insufficient. And the polling starts now.

(Voting.)

MS. LUONG: Zero percent voted for high, seventy percent voted for moderate, sixteen voted for low, and five percent voted for insufficient for reliability.

CO-CHAIR KOTTKE: Okay. Sana?

MEMBER AL-KHATIB: I'm not sure if we want to go through the -- or if we needed to go through reliability testing before we voted on

the reliability issue, but I do want to cover that as well.

So they did that at the measure score level, and they talked about empiric reliability testing was performed on the measure score, as I indicated, using the data source and all levels of analysis as specified for the measure.

They talked about the results of a signal-to-noise analysis for 10 states, and the aggregate results for drug plans and physician groups were presented by them. And they showed that the reliability for the states ranged from .99 states. Mean results for drug plan, .71. Mean result for physician group was .72. And for the ACOs, 31 ACOs that they studied, that ranged from .69 to .98. And then they clarified that the signal-to-noise testing is a commonly used test of measure score reliability. And they tell us that measure score reliability varies between zero and one. A value of .7 is considered the minimum accepted threshold for reliability.

I don't know if, Liz, you want to

comment on this as a statistician.

CO-CHAIR KOTTKE: She sure does.

MEMBER DeLONG: From my perspective, there are two components of reliability, one of which is the one they addressed, which is that you could actually separate the signal from the noise. So you could say this group is statistically significantly on the low side, and this is on the high side. I think they did an adequate job of that.

They did not provide any assurance that -- for example, if they had randomly separated into two groups, would the same state have the same level of -- would that be a reliable measure? Could they repeat that measure reliably? And I didn't see that part.

CO-CHAIR KOTTKE: Judd?

MEMBER HOLLANDER: Sort of a statistics question, and maybe you could help me with this. I obviously don't want to get too drilled down in this. I don't understand signal-to-noise being a function of reliability, but yet

a lot of these measures do that. Is there like a two-sentence layperson's way you could describe that?

MEMBER DeLONG: I actually don't see that as reliability. I see it as discrimination.

MS. JOHNSON: So maybe I can help a little bit here. Here at NQF we allow two different ways of looking at reliability; one at the data element level and one at the score level. So at the score level what you're trying to do is be able to say that you can actually distinguish providers, which is what you'd like to do if you're doing something with accountability. So that's what this reliability, this signal-to-noise, actually does. It tells you how well you're going to be able to distinguish providers in their quality. So does that help?

MEMBER DeLONG: It's really a discrimination measure.

MS. JOHNSON: Yes, it is. Yes.

(Simultaneous speaking)

MS. JOHNSON: Yes. Yes. And to get to your point, Liz, too, our criteria for testing is not -- we don't really have a really high bar for testing. Although some may disagree with that. But you're right that they could have done some additional testing, but that is not a requirement. So the fact that they did at the level they did would be fine.

CO-CHAIR KOTTKE: So I have been informed that we need to re-vote on reliability since we did not discuss reliability testing. Yes, we failed to be reliable.

MS. JOHNSON: Let me just tell you why we're doing that, just in case you're unsure. Both specifications and testing is what you're -- both of those roll up together for your vote for reliability. So that's why you're re-voting here. This is thinking about your conversation on specs and thinking about your conversation on testing.

MS. LUONG: Polling starts for voting on 2(a), reliability. One for high, two for

moderate, three for low, and four for insufficient. And this includes 2(a)1 and 2(a)2.

(Voting.)

MS. LUONG: So the voting results for reliability concludes with 74 percent for moderate and 26 percent for low.

CO-CHAIR KOTTKE: Okay. Let's do validity and validity testing together.

MEMBER AL-KHATIB: So, in terms of the validity testing, the first question is whether the specifications align with the evidence. And I think they do in terms of whether the measure was tested for validity at the data element level or the measure score level. And they did it at the measure score level.

And then the question of whether the testing demonstrates the measure is valid, my answer is actually yes. The way they did it is they did convergent validity by comparing the measure results to similar NQF-endorsed measures for adherence to medications. And they said that they found that the measure results are in the

same range of 70 to 76 percent for this measure as three other measures of adherence for ACO plans, groups and states, with correlation coefficients of greater than .9 for states, but lower correlations for drug plans and lower still for physician groups.

In terms of the inaccuracy of the coding, that certainly remains a question for me, but that's true of all claims data.

And for missing data, they said that they identified that as a possible threat to the validity of this measure, and so they ran an empirical assessment of this potential threat. And so a potential bias they said may exist if day supply within the prescription drug event data is missing, which is a required data element to calculate medical adherence.

In order to evaluate this scenario they analyzed the number as a percentage of beneficiaries in the measure denominator with one or more claims that had missing days. And they presented the results on pages 46 and 47 of the

document that we have, which is the preliminary analysis document. And based on what they showed, I think their argument is reasonable. I don't have major concerns about that.

CO-CHAIR KOTTKE: Any other comments on validity?

MEMBER DeLONG: I think it was good that they correlated this with other measures, but in order to buy into that you have to buy into the whole definition and the denominators and whatever. So they're consistently applying their methodology and getting similar results.

CO-CHAIR KOTTKE: Judd, are you -- oh. Gerard?

MEMBER MARTIN: So, just a question about -- this is, I guess, a process that they're doing. And the question is now that this has been in use -- and it goes back to one of the previous comments was, if it's valid that they are able to show that some people are more compliant than others, wouldn't at some point the natural thing be to see that cholesterol values

that were changing in that population? Is that part of the -- or does that come down the road?

CO-CHAIR KOTTKE: That would be nice to show. The cholesterol values are not accessible in administrative data. So the developer doesn't have access to cholesterol values. That's the short answer.

MEMBER BRIGGS: Just in response to that, because of the new cholesterol guidelines, there are no longer any hard targets for cholesterol levels, so trying to get at that data would be very difficult. Because you could look at -- I mean, in the past we've used 100 as a value of LDL that we would like people to get to, or 70 for people that are high risk. But that's no longer the case. So did it decrease by 10 percent, 15 percent? You'd have to know where the patient actually started at and then know when they were tested next, which gets a little bit more difficult.

CO-CHAIR KOTTKE: Way down on the end and then Liz.

MEMBER SPANGLER: Yeah, I was just going to comment. There is no targets but there is lots of talk about percent reduction. So maybe that's what you're asking, you know, are there any results around that, or is that something we should be considering? Because they do talk about that in the guidelines.

CO-CHAIR KOTTKE: Of course then you'd have to have LDL levels. I mean, you have to have access to the actual --

MEMBER DeLONG: I just agree that there should be some impact to some of these measures, and we haven't seen impact. And especially for a measure that's been in use, I think we need to start seeing impact. Maybe we don't require it now, but to be useful it has to have some impact.

CO-CHAIR KOTTKE: Just one last comment. I think Kaiser is probably the only organization in the country that could do that, because they have virtual identity between membership and care -- sorry, service delivery.

Even at Health Partners we have a very large discrepancy between membership in our health plan and service delivery.

Tom?

MEMBER JAMES: I was just going to say, one of the things about measures is that they're also useful tools. So while we have not seen the kind of change we'd like to see, I think it's more attention. The health plans particularly are feeling the pressure to go about and do something. That's what's going to be measured.

CO-CHAIR KOTTKE: Yes, Liz.

MEMBER DeLONG: I will stop, but I always worry about unintended consequences. For example, what Judd brought up. Now we're going to make people come back for extra visits to make sure they're filling their prescriptions, and all they have to do is really fill it. I just worry that we impose overhead on some measures when we're not sure they're working.

CO-CHAIR KOTTKE: Judd?

MEMBER HOLLANDER: Obviously I'm going to agree with her agreeing with me, but I'm going to raise an issue that I don't think we should settle here, but I think gets more important as we go on and was raised in the last round of us doing this, which is what is the bar that it really needs to get over in order to implement?

So one could put together a hypothetical scenario where there's 16 measures that are measurable, that are reliable and valid, that aren't really going to improve patient care, but are going to take a lot of time and money to implement. And it seems to me -- and maybe I'm interpreting this wrong, and I know that no one will confirm this is true -- that the default here is that everything goes through and that everything that already existed was good enough to exist, so it should continue to exist.

And I know we sunsetted, or whatever we called it, a couple measures last time and we had a lot of discussion around it. And I think it would be really good if there was a bar that

if a measure existed it had to do something. And if it was really successful, it should go away because it was really successful. If it wasn't really successful, it should go away because it didn't do what it should do. But if it's on the path to getting it done, then that's a good measure.

And I think it's probably a discussion that's an NQF or CMS discussion rather than around this table, but I raise that because I think it is important feedback to go back. I don't want to go home saying, oh, we approved 16 other measures, I have no idea if they're going to help anybody.

CO-CHAIR KOTTKE: You have obviously raised the bar for several people on the other side of the room. We'll start with Ellen and then --

MEMBER HILLEGASS: Well, I want to go along with Judd because I have some measures that I have a lot of problems with, and so this is speaking towards tomorrow as well. And my

concern is, is what is the bar? If this has already been out there and this is acceptable and we don't have data, then we should I guess be continuing to accept other measures that are looking at things like, did they fill their prescriptions, rather than did their prescriptions work, those kind of things.

So I guess the question is, what is the bar at NQF? Do you want us to accept measures that really don't have a lot of evidence but could do good? Or do you want us to look at measures that in the future may make a difference in practice? So there's a difference here. And I guess I'm confused, as a person on the Committee, as to where we're going with this. And if this measure comes back again in a couple years, do we accept it without valid data that it's worked?

So where are we going with these measures? Because some measures are just starting. They don't have any evidence, but they might do well. But then there's this other one

that doesn't have evidence that it's done well. We believe that it's done well, but do we continue to keep it? So where should we function as members of this group and where should we cut off and say, no, this shouldn't be accepted and, yes, this should? Is it based on evidence? Is it based on we think it's going to do well for the consumer? Which is important for Carol.

CO-CHAIR KOTTKE: Yes, I think we're the bar-setters. I mean, I think that's why we're in this room.

MS. JOHNSON: I can also take a shot at your question. That's actually why we have the criteria that we have and why it is set up in the way that it is. So we really want you to -- to the extent you can, because we're all humans and it is hard, but to try to adhere to the criteria to the extent that you can.

And also remember that some of our criteria are what we call must pass, and those are the ones that we really want you to pay attention to. That's the evidence -- well, it's

the importance and the scientific acceptability criterion. So those are the ones that are really -- if they don't quite make those, then you should seriously consider you're not recommending for endorsement.

The conversations that you just had in terms of seeing improvements and that sort of thing, that comes under actually the usability and use criterion. And while it is extremely important, it is not a must pass. So that's where you have to weigh basically your own feelings about the utility of the measures.

So I'm not sure that I really answered your question, but again the criteria are hierarchical in a way that we did on purpose.

MEMBER AL-KHATIB: I guess what we're trying to say is that we hope to see potentially a change to the process, if that's possible. I mean, because we all agree that we -- and at some point the measure has to prove that it has led to improved patient outcomes. And so why not make this, for existing measures that have been in use

for five years, why don't you make that a requirement that the developers should prove that there was some impact, some improvement in patient outcomes?

MS. JOHNSON: Yes, so I don't think anybody at NQF would disagree that that's what we want. We want measures that most quickly drive improvement.

I think there's a couple things. One is, as Tom said earlier, it is sometimes very hard for the developers to get that data, so it can be very hard to show. I think the other thing that NQF struggles with is the bar, as Tom mentioned, and some people feel that our bar is already too high. So that would really be setting a very high bar, so we have not enforced that sort of thing. But you are right that we would prefer to see more information, more data, et cetera for measures that are coming back for maintenance review, especially if they've been in use.

CO-CHAIR KOTTKE: Carol and then

George and then Judd.

MEMBER ALLRED: Okay. As I've gone through a number of the measures, as we're studying them, I'm struck by the lack of patient responsibility. We're putting a major burden on health care providers, increasing cost there, but we're really not looking at it's up to the patient in the bottom line to be compliant or not compliant, and no one can make that patient do that. So are we adding to the cost of health care and making it less efficient, and how do we get at the bottom line of what is the patient's responsibility?

CO-CHAIR KOTTKE: George?

MEMBER PHILIPPIDES: So, I agree with Judd, who agreed with Liz, who agreed with Judd --

(Laughter.)

MEMBER PHILIPPIDES: -- that we need to be very careful about adding on too many measurements, too many metrics that are difficult to pull off given how busy the whole health care

system is.

Having said that, I'd like to now take the opposite side of some of the things mentioned. Yes, I think we do want to see outcomes after these are performed, but this is a process measure. This is a change in practice that helps us get to a platform that will later get to assess outcomes.

I don't think that we expect the developers here to create something that is going to prove that cholesterol management is good for cardiovascular outcomes. That's been done. I think we know that if you take statins things go well. This is really looking at something very basic that I think is a broad theme, which is can we change adherence? Okay? And you can take that concept in the cholesterol realm, in hypertension.

What we're looking at is given the reality that people are imperfect and they oftentimes are irresponsible, I think you have to take that as a given. That's not going to

change. Is there a way that we can create a system to improve health care outcomes for those people? To me that's what this is getting at. And I actually think that that's an important thing. Whether or not you think it's the patient's fault -- I actually don't like the word fault in this, it is what it is. Humans are humans. What we're looking at is can we change adherence?

So I think this should be measured, not on whether or not the cholesterol levels go down or that they've had fewer strokes. We know if you take the statins that will happen. We should measure this on can we get people to take statins for longer? And I actually think that that's not an inconsequential endeavor.

So I'm going to take sort of the opposite side and say if we look at this as a process measure not an outcome measure, I think there's some validity to doing this. I'm not sure I believe that, but I wanted to throw that out there.

CO-CHAIR KOTTKE: Okay. Thank you. Leslie has a question or a comment on line.

MEMBER CHO: Hi. So, two things: One is that even at Cleveland Clinic our employees -- when given a prescription for statins, only 50 percent of those people/employees refill their prescription a second time. And I think that when, you know and that's us closely monitoring our employee population. And I think that as NQF moves towards these measures that are more nebulous than did you get your aspirin within however many minutes when you came with a STEMI, or did you get your EKG, I really think it's important for re-endorsement process to have some kind of an effect, that you show some kind of an effect. Without that, to keep on re-endorsing these sort of more nebulous measures, I don't know what it does.

MEMBER BRIGGS: So I would kind of echo that. This is a CMS measure and I would hope that CMS has some resources to measure their effect across years. So I would expect if this

measure has been in place for three to four years, that there would be at least a year's worth of analyzable data that they could come back to us and say, okay, it was the compliance rate was 74 percent among ACOs during this time period and now it's 54 percent, or maybe it's 80 percent. So it got better, and whether that better is significant or not. But there should be some change in it, otherwise the measure really didn't do what it was intended to do.

The idea in a process measure is to get people to use the process. To do what they can to make patients, through education and reminders and things like that, to improve the number of prescriptions filled and hopefully taken. Medicare should see a change in that total amount of prescriptions filled over time. So if they come back to us for renewal, there should be some data to say that the process actually did improve or didn't improve.

CO-CHAIR KOTTKE: Judd, you had a

question?

MEMBER HOLLANDER: I'm just going to make a proposal, because I think we're all on the same page in this room. And so, my proposal, which is way above the scope of this Committee, is that somewhere in the document there is a specific question and a vote is there evidence that this measure has changed anything in the world? Better language than that.

And it should actually be recorded so it goes forward and, if NQF and whoever else wants to use it as a measure, they can. We're not saying you can't use it as a measure, but we're saying 20 experts around the table have looked at the data and found it did squat or it was great. And so, for any measure that's been in existence longer than X time, three or four years, or whatever everybody thinks is appropriate, that should be a line item that we vote on, that we record, and it probably should happen across everything.

CO-CHAIR GEORGE: I think we've talked

about this recommendation that we all are sort of feeling we're giving back to NQF for quite a while, and I think a lot of us are in agreement with that concept that if you've been using it, show us where you were at the beginning, show us where you are now. But I think we need to stick to validity right now, that part of the discussion.

MEMBER DeLONG: Could I just add to Judd's proposal that we elevate usability and feasibility under these circumstances? If it comes back, that should be a priority.

MS. JOHNSON: We'll certainly take it to our governing body. So thank you for that feedback, it is very good and we are often thinking about our criteria and they have evolved over the years and we will expect that they will continue to evolve.

So, and just to make sure that you're not confused, what you're talking about is improvement. And you will be talking about that under the usability and use criteria. So as you

vote on validity, try to think about the conversation of improvement, try to move that over in your mind. Vote only on validity and then bring it back when you're ready to vote on usability and use.

CO-CHAIR KOTTKE: Right. So we're ready to vote, but don't vote on what I'm about to say.

Health Partners, about ten years ago came up with a composite measure for diabetes, and much to our chagrin, six percent of our patients met the composite. Today 50 percent do. And so for like the D-5 composite measure we'd be able to show when you really put your mind to it, you can change things. I think it's quite valid to say, you know, does a measure change things over time? But that's a different conversation.

Okay. We're ready to vote on validity. And do we have anything about threats to validity that you wanted to say?

Okay. Sana says no.

MEMBER AL-KHATIB: I actually

mentioned that when I talked about -- they pointed out the issue with missing data and then they did that additional analysis, and as I said, I don't have any major concerns about that.

CO-CHAIR KOTTKE: Okay. We're ready to vote on validity.

MS. LUONG: For validity, one you can vote for high, two for moderate, three for low, and four for insufficient. And polling starts now.

For validity, five percent voted high, 74 voted moderate, and 21 voted low. So validity passes.

CO-CHAIR KOTTKE: Usability and use?

MEMBER AL-KHATIB: So, the data source is encounter and pharmacy claims. Costs and burden are low. All data elements are in defined fields in electronic claims. So I think it's actually feasible.

MS. HIBAY: Yes, we're talking about feasibility right now, just to make sure everyone's clear.

MEMBER AL-KHATIB: That's correct.

CO-CHAIR KOTTKE: Okay. Oh, I'm sorry.

MS. HIBAY: Okay.

MEMBER AL-KHATIB: That's what you were talking about.

MS. HIBAY: Yes. Very good. Thank you. I just was clarifying. Thank you.

CO-CHAIR KOTTKE: Go ahead.

MEMBER AL-KHATIB: That's all I have.

CO-CHAIR KOTTKE: Okay.

MEMBER AL-KHATIB: I mean, as I said, the data exists, I think it can be used. I think it's feasible. That's all I said.

CO-CHAIR KOTTKE: That was so short, I missed it.

MS. LUONG: So the polling already started. You can vote for feasibility with one for high, two for moderate, three for low, and four for insufficient.

For feasibility it passes with 47 percent for high and 53 percent for moderate.

CO-CHAIR KOTTKE: Okay. Usability and use.

MEMBER AL-KHATIB: Okay. So this measure is used in the CMS' physician feedback quality and resource use report with benchmarks. Though it's not publicly reported or presented, it has been submitted through the measures under consideration process for the CMS ACO Shared Savings Program.

And here's the question that we've been battling with: Indicate whether there's any information on improvement over time. Unfortunately, there is not -- and so that remains a concern that we can discuss it as a group.

CO-CHAIR KOTTKE: Any comments on usability and use? Judd, are you still --

MEMBER HOLLANDER: I echo my comments from before, and I guess this is the area based on direction that should let that impact our voting.

CO-CHAIR KOTTKE: Time to vote.

MS. LUONG: Polling starts now. Oh --

DR. CAMPBELL: This is the measure developer. I wondered if I could make a comment as it related to the prior conversation on usability?

CO-CHAIR KOTTKE: Sure, go ahead.

DR. CAMPBELL: So I think, while this measure was originally time-limited endorsed in 2009, the full endorsement didn't happen until 2011. And if you look at the data that we have available, the data were for calendar year 2011 as it was reported to physicians and those data were not available in a summary report until 2012.

So I just wanted to make the Committee aware, as far as this measure goes and maybe this overall conversation, that the implementation process and rulemaking process can be fairly lengthy. And then there is a lag in terms of actually getting data from the program that can be used for analysis and trend analysis. And so, in the case of this particular measure, although

it has been endorsed since 2011, we really only have data from the program since 2012. And so, I just wanted to make you aware of that issue.

CO-CHAIR KOTTKE: Yes, that is a good point, that it takes time to get time data.

Are we ready to vote on usability and use?

MS. LUONG: Polling starts now for usability and use. One for high, two for moderate, three for low, and four for insufficient information.

For usability and use we have 16 for high, 26 for moderate, 32 for low, and 26 for insufficient information.

CO-CHAIR KOTTKE: I heard that we had a lot of competing measures.

MEMBER AL-KHATIB: Yes, there a lot of competing measures. And if you haven't had a chance to look at them, you can actually see them on pages 53 and 54 of the preliminary analysis.

I just want to highlight certain points, because we don't have time to delve into

all the details. Basically what they described is that this measure also includes -- in addition to CAD, includes cerebrovascular disease, peripheral artery disease, and those were not included in the previous measures.

They also point out that the age range for this particular measure is different. They use 21 years of age and older as a cut-off compared with the previous measures. And then they also talk about the entire 12-month measurement period for this particular measure, unlike other measures that don't cover the entire 12-month measurement period.

Those were the main differences that I saw, but certainly if other people noticed other differences, please bring them up.

CO-CHAIR KOTTKE: Yes, at the end?

MEMBER SPANGLER: I just had a question for the developer, because I think the developer is the same. And I didn't know whether this is a discussion we should have here, but obviously the MAP process is included. I don't

know how many have seen the MUC list yet. There's a brand new statin measure that's being proposed. And I'm just wondering, that measure seems to be very, very similar to this. Actually a little well, a little more encompassing I would say because it's not adherence, it's actually initiation and adherence. And I just wondered what the -- because I think the developer, that is CMS as well, and what they think about -- is that measure going to replace this current measure that we're thinking about or do they think that's going to be in conjunction with this measure? I'm just wondering to get thoughts.

MS. HIBAY: I can address that. So we kind of set up a standard for those measures that are going to be coming in the future, that if it does become a measure that would be competing or related, we would have that conversation when that measure presented.

I think we all know that there's yet one more happy phase -- many more phases on this

project. And so we do anticipate at our next phase with the measure -- call for measure is ending June 30th, Wunmi? We do anticipate some statin measures coming our way. So if they do, and we anticipate that they will, we will review them at that next Committee meeting.

MEMBER SPANGLER: Okay. So I guess it's a process question for me.

We know there are measures coming. We've seen one of them already. I think some of us have an idea of what the other ones are going to be, but that shouldn't affect what we're looking at this one right now.

MS. HIBAY: That should not affect -- only what you have in front of you. If we were going to do a competing and related discussion about this measure at this time, yes, but not at this time. Only what's in front of you at this time. Thank you.

Are there questions about that?

CO-CHAIR KOTTKE: Okay. We need an overall vote.

MS. LUONG: So the polling starts now for overall suitability for endorsement for Measure 0543. One is for yes, and two is for no.

So for Measure 0453, for overall suitability for endorsement, 79 percent voted yes and 21 percent voted no. And that concludes the polling for this measure.

CO-CHAIR KOTTKE: So time for a break? Okay. So obviously everybody has thought long and hard about these things and taken the job very seriously. Liz has one more comment, or not.

Oh, do you?

MEMBER SPANGLER: I have one more. I had a question. I'm sorry. Again a process question. I know for reliability and validity there's a threshold percentage that we have to reach. Is that not true for usability?

(Off-microphone comment.)

MEMBER SPANGLER: Okay. It's not. Got it. Okay. Thanks. That's what I thought.

CO-CHAIR KOTTKE: So we'll be back at

11:00.

(Whereupon, the above-entitled matter went off the record at 10:45 a.m. and resumed at 11:02 a.m.)

MS. ISIJOLA: I think we are going to go ahead and get started.

CO-CHAIR GEORGE: We're on 0670?

MS. ISIJOLA: Yes.

CO-CHAIR GEORGE: Okay. The next set of measures that we're going to take up are related to cardiac stress imaging. We begin with 0670. Discussants are Joe and Sana.

MEMBER CLEVELAND: Hi, good morning. And I think Sana and I agree that since she presented the bulk of the last one, she'll chime in, but I'll take the lead on presenting this.

CO-CHAIR GEORGE: We'll have the measure developers give us a little overview.

MEMBER CLEVELAND: Yes, please.

MR. ALLEN: I'm Joe Allen from the American College of Cardiology, and I've been working on the appropriate use criteria

development and implementation for the past, oh, I guess ten years or so, and so these measures are derivative of those efforts, and you know, I'll just say a few words to set the context because these are quite different measures than most people are used to.

They are looking at avoiding things rather than doing things. As well, they are not necessarily what you might typically think as a process measure. I know it says in your notes that, you know, these may be thought of as a process measure, but they're not an action in that you're not looking at giving a beta blocker or filling a prescription. You're really -- it's a culmination of looking at resource and clinical use and then assigning a value to that.

And so we really look at them as efficiency measures and, in part, outcome measures because they are -- when we go through the process of developing what should be done and what should not be done, we're looking at its clinical value to the patient and to the

population, as well as the resource use for that population. And so, it is quite different from what you might see in other measures. And again, these are inverse measures. So lower is better, that's another difference than, you know, you see in a lot of measures.

The testing and validation that we took on has spread over a number of years. I did include a number of publications in the Reliability and Validity section. There wasn't enough space to include all the detail from the various studies that we've conducted.

We've gone from back in 2008 looking at whether or not we could reliably collect the information to, could we actually put this out in practice and have labs contribute the information? Could we produce change? And then, what were the outcomes? Were there unintended consequences?

And you know, you might think outcomes, typically again, are thought of, are people getting better? Here, we are trying to

avoid unintended consequences of not testing someone that might really need testing, and we want to make sure that if we are saying don't do something, that it truly shouldn't be done. And we have changed the terminology to be rarely appropriate rather than inappropriate. You may be familiar with those terms.

So currently there are over 1500 institutions that we've collected data on nationwide in our various activities, and greater than 31,000 cases. We actually have a statewide project in Delaware, as well as a partial state project in Pennsylvania with a large private payer that has been going on for the past three years, and so unfortunately, I wasn't able to include that in this data set because it is not completed. But I just wanted to, given all the usability discussion in the last measure, to say we have lots of data from that experience and some really interesting things that I hope to publish shortly, but we won't be able to get those published and out there, but I can speak,

you know, to some generalities if we get into that section.

It is widely used for private payers, as I said, lab accreditation, PQRS, QCDR. And in 2017, CMS will require clinical decision support for any advanced imaging such as nuclear imaging, that's covered here, CT and MR.

So despite some concerns about ability to do this, we are working hard to make it easy for folks to do it in 30 to 90 seconds. You are really just ordering a test and saying why you are ordering it with a few data variables, and we've actually streamlined that over time, and so we can talk about that as questions come up related to that.

And then lastly, I'll just say that this is a population review. There are a lot of things that you can do on an individual referral to, you know, ordering physicians to give feedback back. This measure looks at an aggregate population and how often testing is done in various populations. Once you have that

population measure, you can go back and say, what are the specific actions I want to do on individual cases?

But what we really try to do is emphasize a population because the private payer approach before these measures, and still continued today, is what is known as prior authorization. And so we really wanted to provide an alternative that was more based on quality improvement and helping people change their practice patterns over time, where prior authorization looks at individual cases and says yes, no.

And that's not what we're really trying to do here. We're not trying to say don't ever do this. We're just trying to look at how often it is done, and then over time, reduce the use in populations where it does have low value. There may be exceptions to the rule, as people brought up in some of the comments.

And in general, I'd also want to say that these pick out three particular indications

that were highest amongst our populations that we tested of rarely-appropriate use of these types of procedures, but it is in a universe of about sixty different other things that we do collect as a part of that data registry collection effort, and so you may think of other things, like incomplete revascularization on some of the later measures that we'll want to talk about. There are indications for that, and they aren't included in that particular measure, but there's an ability to say another reason -- it's just not captured in that particular measure because it's only looking at that particular reason.

So I'll close with that, and I'm sure we'll have other questions that come up as we go through the measure. Thanks.

MEMBER CLEVELAND: Thank you. I'll introduce the measure. Just as Joe mentioned, this is a family of three measures. This is Measure 0670, which is Cardiac Stress Imaging not Meeting Appropriate Use Criteria, Pre-Operative Evaluation in Low-Risk Surgery Patients. As we

have heard, the measure steward is from the American College of Cardiology Foundation.

So a brief description of this measure is the percentage of stress SPECT MPI, stress echo, CCTA, or CMR performed in low-risk surgery patients for pre-operative evaluation. So as we've heard, this is I guess a little bit of a new look for us. At least in our first meeting, I don't think we had any appropriate use criteria.

So really, this is kind of starting to look into this area where trying to find when to do, when not to do a given procedure or a test in the environment, and as such, I think that's something that is certainly novel here.

If it's okay, this is -- the data source for this, obviously registry data, is that the level analyst or the clinician group practice in facility. And I think primarily targeted at the facility as I read this, I think that's correct.

If I can jump into the evidence -- and

I think this is probably where we'll spend a fair amount of time because, again, when one looks at the evidence provided, certainly there are a lot of guideline specifications that were provided. To try to look at those guidelines and map them specifically to this is difficult.

The developer references the RAND Delphi process that was used in -- I've done that once. And obviously for people that may or may not be familiar with that, that's basically convening, if you will, a group of experts who then sit around a table and vote from a score of one to nine whether something is inappropriate, unknown, or appropriate, given various patient scenarios. And I know when we did this for a coronary bypass, it was over 90 different scenarios and things like that.

So at the end of the day, I guess to jump into the -- at least our measure review document and answer some of these questions. I think that, you know, as I say, I think the question for the Committee, is the evidence

directly applicable to the process of care being measured? I think so, as we have it, but I think we need to have some discussion about how we feel about the RAND process.

I think the process is proximal and directly related to desired outcomes. And then, I think there are -- is there evidence of systematic assessment? Expert opinion. Beyond those, I don't think there is a lot of evidence for systematic evidence beyond expert opinion, and so I think we'll have to figure out if we really think this is a level of evidence we can support without other evidence.

CO-CHAIR KOTTKE: Sana.

MEMBER AL-KHATIB: The first thing that I want to bring up is the whole issue of the level of the measure, in terms of the level of analysis.

You know, I think I am struggling with making the imaging facility the level for this analysis because for the most part, I mean yes, they need to be looking to see if the indication

for the procedure is appropriate or not appropriate. That decision and determination are largely made by the ordering physician, not the imaging facility.

So I am struggling with that concept, of holding the imaging facility responsible for something -- I mean, unless it's dangerous, I think they're going to do it, but are they really the responsible entity for the appropriateness of the test? It's really the person who ordered the test, not the imaging facility. So I'll start with that.

CO-CHAIR GEORGE: Do you want to address that?

MR. ALLEN: Sure. We did look at both referring physicians and measuring at the facility level, especially at the imaging facility, and we find that it is a partnership between both sides. In our statewide Delaware project, it is the onus on the imaging lab to look at this, and lab accreditation also requires labs to look at it.

I know it is a different approach, though what we found when we measured at the individual ordering physician or even group level, often they didn't have enough cases to reliably give a pattern that we could give community feedback on.

And we found, in our partnership with the referring -- or the centers performing imaging in the cases where it's done, including the studies that are cited here. They found, surprisingly so, that the referring physicians really do want that guidance and partnership, and despite, you know, initial thoughts that referral centers would push back and say, well I won't send my imaging to you if you talk about this with me.

In general, if it's not about individual cases -- remember, they are having to go through prior authorization anyway through private firms that they really hate. This is not that. This is about a population, and so when you give feedback, it's not on an individual

case. It's on, we're seeing this generally in our community, and we'd like to remind you of different information that -- other alternative tests that could be done in these situations.

Or in the case of a pre-op, that here's a great evidence base from the guidelines, from actually randomized trials that have looked at this issue, and we don't need testing. You can be reassured, many of the protocols are set at the facility level.

And so we felt this particular measure as well as the other two, it is a partnership. And you can look at the other side for quality improvement, but for accountability, there weren't enough cases to hold individual ordering physicians accountable.

CO-CHAIR GEORGE: Liz and then Judd.

MEMBER DELONG: I guess I would question the rationale of we don't have enough cases for the appropriate victims, so to speak, so we'll move it up a level to a different entity. And I do wonder about the interaction

there, when -- and you guys all know more about this than I would -- when a physician refers to a facility and the facility pushes back. You sort of addressed that, but I would think it would create a little bit of tension.

MEMBER HOLLANDER: So that's one of the points I was going to make. And from my point of view -- now I am an emergency physician, but at every place I've ever practiced radiology has no right of refusal of tests. The clinician who knows the patient decides.

So to me that's a nonstarter and I totally, this time I will echo Liz's comments, that you made it very clear there was a place you wanted to measure, but there wasn't enough data. In my world that means stop and let's do something else, rather than just lay it on the imaging center. So to me, that's like a critical, critical thing.

The second thing that I have real issues with is I am not sure that I want measures being developed by a Delphi process that is just

coming forward. Now, I recognize that ACC/AHA guidelines are some sort of a modified Delphi process anyway, but they should be guidelines from a society, not from a group of people putting together the measures. They should be published, they should be vetted, they should be signed off on the appropriate boards, and everybody else, and then they are guidelines that we should be holding people to.

My fear is that, you know, you get any ten people in a room, say you went through a process, and say now there's expert support for it. I have real problems with that. So those are my major two items at a 10,000 foot view for this.

CO-CHAIR GEORGE: Tom Kottke, Tom James, and Gerard.

CO-CHAIR KOTTKE: Yes, thanks. We -- when I was down at Mayo for 17 years, we would get a call from the radiologists, and they'd say, you know there's a better test. And we get the same thing in the Twin Cities.

And I have on my screen right now, since 2006, we have had a -- basically what you described as it's just a decision in support of, you know, why you are doing this. And we've flattened the curve. I mean, basically since 2006, there's been no growth in high-tech diagnostic imaging in our market, where it was just going through the study itself, if somebody wants to put it up afterwards I will show it to you.

The system works. I think you can create this partnership, and it's part of choosing wisely that, you know, you talk. Doctors talk to each other from time to time about, you know, is this the best test? Because typically the radiologists, or the nuclear cardiologists, know a little better about what -- you know, and if you have shared decision-making about do you really need this patient who is going to have an extraction of a cataract, do they really need a pharmacological stress test of their myocardium? And the answer is no.

MEMBER JAMES: Couple different things. First, I would consider this to be more of a resource use measure rather than a true cardiology measure, just for the reasons about who is accountable. In which case, it may be better held in another venue.

Secondarily, and I do have admit that I was on the PCPI work group under Joe Drozda that looked at some other prior-to-surgery measures, and I provided data from Humana on that. But one of the things that I learned about that is that you have to look at the timeframe.

This is a 60 day window, as I read this. As a practicing -- part of my life as a practicing internist, is being in a hospital. The typical scenario that at least that I used to see a lot in a couple of different states -- I can't hold a job and stay in one place. But what I used to see was a patient would come in with chest pain, be evaluated, have an acute MI, and would be followed if they're low risk for a while, and then have a decision for surgery.

That imaging test was done way back early on, and so it may be outside of the timeframe. So I've got that particular concern.

So those are my points. Thank you.

MEMBER MARTIN: So I just -- maybe it's beating a dead horse, but this partnership thing is I think a critical one. We've attacked this in pediatrics, and you know, at our hospital, you can't order an echocardiogram. Pediatricians can't do that.

And the idea that you can send a patient to radiology for a cardiac test and they will just do it, I think there is a flaw in that. I think that there should be a partnership, and I think we -- particularly where imaging was so drastically out of control in the cost curve, you know, I think it does take the person with expertise in the disease working together with the community physician to limit overuse.

CO-CHAIR GEORGE: Thank you. Sana and then Linda?

MEMBER AL-KHATIB: So I mean I

completely agree that in an ideal world, that's what should take place. That you know, as an ordering physician, you talk to the imaging person, and then you discuss and decide regarding the best test to order, or if it's appropriate even, to order it in the first place.

But that's not what takes place everywhere. And I can tell you, at least where I practice, nobody asks me anything. I order a test, it gets done. So maybe there is variability in practice, and so again that would raise a concern for me, if that's not the model that is being used in different places.

But in terms of the evidence, just a couple of questions. I was one of the reviewers assigned to it. I mean, I agree that ordering these tests that are not necessary is not a good thing. I think we need to see more evidence that it's actually, you know, harming people, hurting people.

I hear you about the application of the AUC, but I am not aware of anything that has

shown that if you actually follow the AUC to the letter, that actually patients do better. I haven't seen those data. So if you have any data like that and you can share it, even like in general terms, I think that would help me focus my thinking about this measure.

MEMBER BRIGGS: So in the discussion about facility and whether facilities should be held accountable. For this particular measure, because it's a pre-operative measure, many pre-operative evaluations are actually done at the facility level.

They have nurse practitioners, PAs, other people actually that are doing those pre-op exams and potentially ordering those tests. So the facility has some involvement at that level, and many of these facilities also have large cardiology groups that work within them.

And again, so there could be a level of responsibility of the facility related to that. So I'd -- in some areas where primary care physicians might be ordering, maybe the facility

isn't as responsible. But in a lot of metropolitan areas in particular, you get a lot of pre-operative evaluation actually done at the facility level.

CO-CHAIR GEORGE: George and then Michael, did you have a comment? I'll take yours after George.

MEMBER CROUCH: Another comment from the non-ideal world. As a primary care physician, I was very surprised to see data on primary care physicians ordering these tests. I never order these tests. I send them to a cardiologist and may suggest I would like for them to have to go to a stress test, but that's as far as it goes for me as a primary care physician in the Houston area, suburban Houston. And my strong sense is that, if the test gets ordered, it gets done most places. I think it would be nice to have the partnership where you get the feedback.

And I do have, interestingly, other imaging people call me with some regularity and

say Dr. Crouch, do you really want this test? I think that you may perhaps want this other test. I say, fine, because I know they know more about which is the better test for neurology imaging, or whatever. But that isn't happening in our area in cardiology stress testing.

MEMBER PHILIPPIDES: In our facility, if I order a nuclear test, it gets done. But it might not be a bad idea for there to be somebody on the other end who says, you know George, I was looking at this, would you consider doing this or not doing it, because four months ago they had the same test and it was okay? So in a weird way, adopting this kind of measure might actually create a system which is facility-wide and not based on one individual, and I think that would be working better.

In regards to data about the harm here, I don't know of any direct data, but there is now a lot of indirect data suggesting that all of these nuclear tests that we do cause cancers, right? So if you look, you know, nationwide at

the number of SPECTs, CCTA as well, that's a lot of radiation. And we will cause X number of cancers, this has been sort of calculated, by doing nuclear tests at least, that are perhaps not necessary. And I have seen people going for cataract procedures who go and get a nuclear stress test. And if that were my mom or my spouse, I would be very unhappy about that.

So I think there probably is some downside. Again, I am not sure this measure gets there perfectly, but I think there's reason for us to start to look at these kind of things.

CO-CHAIR GEORGE: So keep in mind, we are talking about the evidence. Comments from Tom, Mladen, and Judd.

MEMBER VIDOVICH: I just have one brief question for the developers. Is -- the measure is supposed to be lower is better, right, you know? And that's a continuum of lower reaching zero. How low is acceptable, meaning is there a target you are looking for? Meaning are we looking at ten percent, 20 percent, or should

we have maybe a grading scale? Maybe there will be like, I don't know, less than 20 percent, less than 30 percent, or something?

It seems that lower is better than nobody should ever get a stress test for any reason for low pre-op surgery, which is unlikely to happen. So people don't get penalized for, I don't know, not reaching zero, or one, whichever number you pick.

MR. ALLEN: Great question. The goal of this measure is not to get it to zero. It measures what generally should not be done. On this particular one, there's a lot of evidence that you shouldn't, but because of -- you know, like cataract surgery or whatnot. And so this particular one, we do see people driving closer to zero.

But we generally look at these measures as a collection, and when you look across the different rarely appropriates, especially the ones that are captured in these measures, folks generally start at about

somewhere between 15 and 30 percent rarely appropriate cases, and if they have a systematic process in place like has been talked about today through decision support, and again, remember this is about a population, not individual cases -- understand the sensitivity of referral and ordering physicians. Some people may feel great about doing that, other folks may rather talk about general, like, let's have a faculty meeting and talk about how do we work as a facility on this.

Different approaches. That doesn't mean -- the measure is to capture information. What you do with that information across the population is really meant to help facilitate the conversation, either individual or group, so.

And it's not, again, to go to zero. Generally, we start at 15 to 30 and we go down to five to eight percent rarely appropriate, and almost nobody gets to zero.

CO-CHAIR KOTTKE: I think we have to be very careful about conflating community

standard with quality care. And if it were, we wouldn't be sitting here.

MEMBER CLEVELAND: I guess a comment and a question. I guess, in -- my question is for, I guess, our NQF staff. Joe alluded to some data that, you know, is coming out on some things that's not in the packet. Is that permissible to share with us here? Because I think that's part of what we're struggling with.

I mean the process of evidence for this is just, this is a little bit outside the bounds of what we've had to look at. And I like it because there is some tension with the appropriate use criteria, and I tend to be an AUC advocate where I think we need to look at this. But I think that's where some of the -- I guess where I've struggled in terms of trying to put this into what our current evidence statements are and, you know I mean, this is like one of these motherhood and apple pies. It's like yes, nobody should get a stress test for a cataract. I get that, but how do we get there with what we

have?

MS. JOHNSON: Right, so in this particular instance, I think the developer can certainly -- and anybody else in the room that understands the evidence and would be able to share that verbally, that would be totally appropriate and you could act on that.

We would probably ask, and take it a little bit further, we may actually ask the developer if there is more evidence that they could actually put into the form so that it is there for posterity. We might ask them to come back and actually put some of that verbal stuff into the form. Does that help?

CO-CHAIR GEORGE: Judd?

MEMBER HOLLANDER: I am having a little trouble with a couple things on this that maybe the developer can help clarify.

One is, you know you mentioned you may be publishing stuff. We have not limited the evidence base here to published materials. In fact, most of these things are totally

unpublished materials. So I would ask you to comment on the data which is maybe in press, which, you know, in the outside world can't be shared publically, but in this world, we can't make a decision unless we see. So keeping things from us, you know, to protect publication rights, I don't think does the public service. Which is what we're in this room to do.

The other thing is, and maybe I interpreted this wrong because nobody followed up on my comment before. Are these appropriate use criteria, are these widely accepted? Are these ACC/AHA appropriate use criteria? Because it doesn't say that in the documents that I see, but if they are, then I would think of it very differently than if a group of people developing a measure got together and decided what appropriate use was.

And then I'll kick back on my colleague George over there, in that I don't know that there are actually harms associated with this. There's a lot of modeling. Like if you

blow up a Chernobyl or Three Mile Island, radiation is bad. But there is certainly a lot of debate, and two broad camps.

One is, you know, less radiation is always better than more radiation, and I could agree with that. But the other is that there is no evidence that medical radiation in an adult population does anything outside of make for good modeling papers, and we don't know the answer to that. So I think it's actually really important to know what is the evidence and how did these criteria come to be, and will the world accept them?

MR. ALLEN: Great question. I apologize for any confusion. In the measure packet you'll see a number of publications. They are through a society joint process following a rigorous process that is derivative and based on guidelines, but takes it one step further and looks at particular clinical scenarios, as was discussed.

But it is through a rigorous process.

It's not just some random ten people in a room, and in fact, you know, we get into debates all the time, whether it be imagers or cardiologists or interventionalists or practicing physicians, how do we do this?

I will speak just briefly to it. We do go through a multi-step process. We have a writing group that's independent from the rating panel, so when you describe the scenarios, the writing panel can think of all the things that they want to ask, but they can't dictate what the scores will be at the end result. So there's separation between those that are writing the scenarios and looking at the evidence and the people that are actually saying is this reliable and should it be done?

And then there's a rule that less than 50 percent of the panel can represent any particular treatment or diagnostic choice on the panel, so we're not just putting a bunch of imaging folks in the room and saying would you like to do more imaging? We are, you know, we

have a strict process that requires only 50 percent or less the imagers and then put it through a whole review process, and then it gets endorsed and looked at by societies.

And so it is very similar to the guidelines. We just don't do the systematic reviews because the guidelines do that for us, and so we map every clinical indication to those systematic reviews. So it's a lengthy process, can't describe it fully here, but that gives you an insight.

You asked about harms as well. You know, there is an article in the packet that does look at the predictive value of these tasks within the particular areas that we're looking at for rarely appropriate and shows that, you know, compared to appropriate or maybe appropriate studies where you actually get value that tells what you should be doing with a patient, in these rarely appropriate circumstances collectively, they don't actually contribute to the decision-making process.

And so that's a 2013 publication. There are a couple other abstracts that have been in process that I haven't seen actually published, but there is a published paper in circulation on that particular issue of -- you know, it doesn't help, it's taking resource use and as has been discussed, although any one test would not necessarily contribute to radiation, we shouldn't be exposing people to radiation if there's not a clinical value to it, so both from the financial and the no clinical benefit, that's where that's coming from, and I think we have pretty strong evidence in that case as well.

On the unpublished data, I wish I could share it with you. It is a partnership with a private health plan, and so that is their data, and it's a part of a partnership with them, and so even to discuss, you know, specifics on that, I would be violating both some HIPAA and business associate arrangements to give you any specifics or to provide it to NQF where it may be then posted later on.

And so that's why I am talking in generalities about it and available data that we have, because later on, you'll talk about whether or not this can be collected and done a regular basis. I can tell you we've been doing it for three years, and it is done on a regular basis, and so I am not trying to withhold, and my chapters and my members actually beat me up on a weekly basis about why can't we release the data -- in part they wanted to get three years of solid data so that they would know that the trend is reliable.

They didn't want to just release it and say hey, we're doing better, and then, you know, have some relapse as you often see on these types of observational studies, so.

CO-CHAIR GEORGE: Yes, Sana?

MEMBER AL-KHATIB: So I have a comment and a question. The comment has to do with your question about well are the AUC accepted, you know, within the cardiology world?

And first of all, I have to say that

overall, I am supportive of the AUC, but I have some reservations about them. I could use a couple of examples from the ICD, the implantable cardioverter defibrillators AUC document. They had more than 250 different scenarios that that document summarizes, so it's actually not very user-friendly for the average clinician who doesn't have time to go through all the whole document.

A couple of their criteria went against the guidelines. So I think that those things really need to be better aligned, in my opinion.

And I still think, as somebody who believes in the evidence, that we need to see data on the association between applying the AUC and good outcomes. And I still, as I said, see that this is a missing piece, I think in relation to all the AUC criteria, but please correct me if I am wrong.

The question that I have to you is when can we expect to see some of these results?

I mean, are we looking that these will come out within six months, within a year, within five years? I mean, just kind of getting a sense of when they'll be out.

MR. ALLEN: Sure. Well first I want to say, you know, despite the Highmark data not being available, you know, in your publication materials you have several articles over, you know, ten years of data, and this has been looked at in several publications. And so we don't have the particulars on the statewide mandatory use type thing, which is the broadest application, but we have a lot of single center as well as multi-center looks at imaging-appropriate use.

And I know there may be differences around defibrillators or stents and things like that, and those are newer things, we are looking at them, and you know, the imaging criteria, I have to say, we didn't even develop measures on them until the second round of appropriate use because we knew there were some limitations and we wanted to get those single-center studies done

to better refine and give us feedback, when was there value, when was there not value?

We also work a lot on shortening. Even though we might publish 250 different things, we work on decision support, and especially in imaging, to break it up. And so you might see 60 scenarios in our criteria, but then we break it up into one to five questions in one particular area that any physician has to answer at the time that they're ordering.

So they're not looking up 60 different things, we're saying, well why are you ordering? Well, it's for pre-op. Okay, so what type of surgery? Low-risk surgery. Okay, well you don't really need a test because it matches this particular indication.

It's as quick as that. It's a one to two question survey at the time of ordering that then matches that you're not trying to order for all 250 things, you're trying to order for a particular patient at that particular time, and then we report across the population on those 250

or 60 things to help you understand what's your patient mix at your facility, to help you then reflect upon that and work within your facility however you see fit to do the quality improvement around that.

So -- and, you know, looking at outcomes I guess, whether it be imaging or stenting or defibrillators, I mean there are randomized trials that try to look at these issues. I mean the question, a lot of the questions that we're facing on imaging are related to, you know, ischemic outcomes in stable populations where most of the trials have not shown a benefit to even doing what you would do after the imaging.

There's symptom relief and other things quicker, and you know, I am very supportive, even in our stenting criteria, not limiting people's ability to do that. But if you're having asymptomatic patients, like in this particular case where they're just coming in for a pre-op evaluation, even if you got an imaging

test that was positive, what would you do with it?

You're not relieving symptoms at that point. You might randomly find something that shows a mild defect in a woman who, you know, has an attenuation artifact on a nuclear study, and then there are case studies published in the literature showing that you then go on to cath and then have a perforation. I mean, these are not -- there are adverse events in these things, and they're only case studies because nobody sets out for a $30 million randomized trial to show we shouldn't do something.

CO-CHAIR GEORGE: So we've had quite a bit of discussion here. Are we ready to move on and vote on the scientific evidence?

CO-CHAIR KOTTKE: Now this may make everybody uncomfortable as hell, but you know, the problem isn't the harm to the individual patient, it's the harm to society.

And I've got a graphic right here. Massachusetts, FY01 to FY14, growth adjusted for

inflation of consumer price index, health care 81 percent, early childhood education and care minus 27 percent.

And so what we're doing when we're doing these useless tests is we're taking education out of the brains of little kids. We may not be able to process that here --

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

Are there any comments on the phone?

MEMBER CHO: Yeah, I just wanted to sort of validate Tom's point. And I think that for the next three measures, the evidence is quite low.

I think that we all agree as, you know, as physicians that there's over-testing in America, and if this is a small way for us to start the talk about over-testing in America, then I think it's okay to pass measures like this that have low evidence, but the intention I think all of us are in perfect alignment with.

MS. LUONG: Voting begins now for

evidence. One for high, two for moderate, three for low, four for insufficient evidence, and five for insufficient evidence with exception.

For evidence, 39 percent voted moderate, 50 percent voted low, 6 percent voted insufficient evidence, and 6 percent voted insufficient evidence with exception, so the measure does not pass.

MS. JOHNSON: Okay, just to remind everybody, we were looking for at least a 40 percent to 60 percent gray zone area in either high or moderate together to make it pass, so it actually just came in under the level, so at this point we're not going to continue the discussion of the measure.

And we will be talking with the developer a little bit later just to see if there's something else, maybe, that the Committee was not aware of, and if there is, he would have the opportunity to bring it back after public comment and see if there's, you know, something else that he may want to bring forward.

So that's where we are now. And it looks like Liz has a question, and --

MS. DELONG: Why wouldn't this measure be transferred over to possibly resource or cost division? Because what our problem is, there isn't evidence for some of this harm, et cetera, but there is evidence about cost.

MS. JOHNSON: Well for one thing, it's not so much that the evidence is about cost. The cost and resource use groups are a little bit more technical group that is looking at much more technical measures than this, so this one actually, we think, fits with the CV. And Tom, you can --

CO-CHAIR KOTTKE: I was just going to say that I voted low because I followed the directions. I mean, you know, if you take the algorithm down, it makes you vote low because there's no QQC or meta-analysis.

And you know, I mean, I think it's a good measure, but I just followed the directions.

MS. JOHNSON: And what we can do there

is work with the developer because as I was reading it, and the AACs were new to me as well, but my understanding of those, and maybe I'm incorrect, my understanding is they were based at least on an evidence review, a systematic review.

That didn't necessarily come through in the submission, and I also know that you cited several articles but didn't really summarize them. So that may have affected people's viewpoint, is how they went through the algorithm, so.

MEMBER DELONG: So Judd made a proposal earlier, and I would like to make another one. I think it's a shame that the evidence killed this, given that the comparison between this measure and what its potential is relative to the previous measure, it seems that this measure has a lot more potential for an impact.

MEMBER VIDOVICH: I mean, I just want to say something I thought initially I wouldn't say, but let me just briefly discuss this.

And I completely agree with you. This measure has great potential. But it contains so much paramedical stuff that it's hard to measure, such as medical legal considerations, such as billing considerations which are well above and beyond what we can figure out.

Yes, this measure will address this, but again, that is well above and beyond what I can even possibly imagine anybody fixing in our lifetime. So again, and I don't know how to incorporate it, and it's very similar to the next measure, 0671, that we'll discuss. And it does contain it somehow implicitly in there.

CO-CHAIR GEORGE: So we will move on -- oh, I am sorry, I'll just take you in order. Joe, Tom, and Judd?

MEMBER CLEVELAND: I thought the insufficient evidence with exception, can that not fall into kind of putting us into the moderate range? Where again, because on the body of evidence, I agree with Tom, it's low, and it's even -- to quote him, but it's such a good

measure, it's like well based on the evidence alone I can't really say, again, within the structure that we have, that I can stamp this and say good.

I like it. I kind of thought that might be a vote that would also be one that would not count against the evidence, but I guess that's just -- what does exception then mean? If that is exception, that we could move on unless that's a majority, is what you're saying.

MS. JOHNSON: Right, but we can't add exception with evidence with moderate to get above your 40 to go forward.

MEMBER CLEVELAND: Got you.

MS. JOHNSON: So, I mean, if enough people are uncomfortable and they would have said, you know, knowing that they may have switched their vote, I mean we could talk about potentially re-voting. That's a little bit irregular. So --

MEMBER CLEVELAND: I certainly don't want to be the lone rebel -- as a heart surgeon,

we're in the minority enough here, so.

MEMBER JAMES: I'd like to go with Tom and Ellen that this really is a resource measure as far as I'm concerned, but if we're going to be looking at it strictly from the scientific point of view, the use of Delphi technique -- and you started to address it, but it's not written very well.

The Delphi technique is part of -- is on the AHRQ evidence, hierarchy of evidence, is way down there. And being able to use more in the way of good, clinical studies would raise the level of evidence, and that would be one suggestion that I would be making on this thing.

The other thing is looking at an episode of care as opposed to this arbitrary 60 day period, if that's possible. It is with health plans, but I don't know if with other data sources.

MEMBER HOLLANDER: Yeah, I was going to say, I think what you're hearing is we all love the idea. It's things within the measure

that could be tweaked.

I for one would rather see it at a health-plan-wide or institutional-wide rather than the imaging facility. I mean, I hear what Tom's saying, and frankly I agree with it. I love if I am calling for a test and they say this is a better test. I hate if they say no, you can't have the test you want, and there's no alternative.

But regardless, it's the system as a whole that needs to work it out, and that's what everybody said while we were debating imaging facility or doctor. So make it for the system as a whole, and then, you know, maybe hopefully soon you'll have more of the evidence to get us over the hump to do that.

But I'm just trying to express encouragement because I think everybody is saying we went down the list and this is where it falls out, and yet you're sitting on a pot of gold, and I understand the reasons you can't disclose it. But when that's disclosed, that would probably

change the boxes in here and get to the next thing.

But it really does give you a unique ability because you now have time to change the measure based on the feedback you're getting here, so when it does come back with more evidence, it might actually work well.

And then my final comment is if the world really does move to many more ACO type models, this will become sort of irrelevant because it will take care of itself. And so focusing at the system-wide level will set up institutions that play into an ACO better rather than focusing just at an imaging center level.

CO-CHAIR GEORGE: Liz, did you have a comment?

MEMBER DELONG: No.

MS. JOHNSON: Go ahead, and since the developer has asked, let's hear what he has to say real quick --

MR. ALLEN: Yes, so I am a little confused on the conversation because we are

talking about two types of evidence: one which is evidence for the measure and its impact on care, and we have quite a bit that we can measure it and that it does have an impact in changing practice behavior.

The other is the evidence for not doing something. And although we have some data, and actually the peri-op probably has the best data on it, I didn't quote it in the measure form because it's all in the actual publications that we provided, although, you know, we can pull some of that out in the future -- but I mean, there have been randomized trials looking at whether imaging improves outcomes related to cardiovascular and the surgical outcomes, and yet everybody voted that it had low evidence. I am a little confused by that.

And I know that it wasn't directly, you know, brought out because we relied on folks knowing the process, and the last time this was reviewed, it was reviewed in the Resource, so maybe the level was a little bit different, but

you know there are randomized trials showing that imaging does not contribute to better outcomes, and we are saying don't do it, and CMS has approved this for QCDR. They actually have a parallel measure that measures this on claims.

And it's going to be mandated in a couple of years. And so it seems odd that NQF would step back from this measure as it's about to be mandated nationwide. Just -- so and that applies to any of the other two measures as well.

So we'll see how we go through on the other two, but just want to say that.

CO-CHAIR KOTTKE: Yeah, I mean, and maybe I misinterpreted the algorithm. But I'd ask, you know, is there either a meta-analysis or a discussion of quality, quantity, and -- what's the C? Consistency.

You know, and I guess -- and maybe I missed that.

MEMBER BRIGGS: So I'm thinking that a lot of it has to do with just what data actually was presented in the document because

while the Delphi process was your ultimate use for -- determination for the appropriate use criteria, that was supported, as you said, by guidelines, which are supported then usually by multiple, multiple studies.

Perhaps pulling out some of those key studies and citing them within this, the structure for the evidence here, would be helpful for people, because again, trying to dig too deep in multiple, multiple documents is difficult at times.

I mean, you live these things and live and breathe them, and I can understand why you think oh, they should know. But not everybody at the table may have read every supporting study that was involved in creating the scenarios that you're talking about.

CO-CHAIR GEORGE: Michael?

MEMBER CROUCH: I just want to reiterate that the strongest measure applications that we've had have included all of the relevant evidence summarized succinctly, and that wasn't

done here, and so we're missing a lot of information.

I didn't have time to read all the articles that were attached to all the measures. Even the ones I was assigned to, I barely had time to skim. So if you want to maximize the chances of it doing well, you need to give us all the bullets you've got, lay them out there in the application.

CO-CHAIR GEORGE: Sana?

MEMBER AL-KHATIB: Also, so the question here that you posed to us is that briefly explain the evidence presented by the developer that supports the relationship to outcomes.

And so in thinking about that, I mean, correct me if I'm wrong. I mean, I am an electrophysiologist, I am not an imaging person. But I am not aware that there are well-designed and conducted studies, not necessarily randomized clinical trials, but even, you know, observational studies, that show this association

between if you order a lot of these tests, the patients end up getting harmed because of this and that.

Now, of course, we know of people where, you know, you had false positive results and people ended up having invasive procedures and had a complication. Of course, we do know of those cases. But what is the magnitude of that problem? I, you know, I don't know.

And then the other thing about the association with radiation and those bad outcomes, you know, give us more information that yes, here's the association there. If you can't do that, you know, even if this is to be revisited, if you cannot do that, I still think that your best bet is with the cost and resource, because there the association is very clear. And that's why I'm not surprised that CMS approved this measure. Of course they want to save money.

But -- and I'm all for that, you know, from the societal perspective. But as a clinician, I need more information to support the

issue of yes, people are getting harmed.

MS. JOHNSON: Yes, so we will definitely be working with Dr. Allen to see what we can do with his submission. So he may be able to bring it back, if not by -- I am not sure if he would be able to do it in our post-meeting call which we have scheduled a couple weeks down the road. If not, then perhaps after the public comment period.

CO-CHAIR GEORGE: Okay. I think we'll move on to the next measure, which is 0671. Discussants are Leslie, who is on the phone, Tom Kottke, and Mladen.

MEMBER CHO: Hi. Is the measure developer joining us?

CO-CHAIR KOTTKE: Yeah, he's right here. But Leslie, Tom here, if I can have a question. I mean, the foundation for this measure is exactly the same as the foundation for the prior measure, as is the foundation for the next measure. And so is there any sense in discussing it?

MEMBER CHO: Yeah, so I am just going to make a pitch for this. And bear with me while I just go through this thought.

I totally agree that the evidence for this measure is very low, as is the evidence for other measures in this group.

However, I think that given the over-testing that's prevalent in this country and given the fact that these are measures that were previously endorsed by the Committee -- by the Cardiovascular Committee, I still think that these are worthy measures for us to vote on.

And even though the evidence might be low, it really is, if you think about it, it has very high priority. And because of the high priority, while the evidence might be not as good, I still think these measures should be re-endorsed.

CO-CHAIR GEORGE: Ellen?

MEMBER HILLEGASS: For clarification, was this measure and the previous measure endorsed by this committee or by the Resource

Committee? Because I am reading that they were endorsed, but my understanding is the last measure was endorsed by the Resource Committee.

So was this one also endorsed by the Resource Committee? Yes.

MR. ALLEN: Yes, all these measures came through the cardiac efficiency call for measures. It got wrapped up into the CV update.

We agree we're not sure that it's the right place for this discussion. You know, I really do appreciate the, you know, the feedback on, you know, the clinical information. This is the exact same information we provided to the Resource Committee, and they understood the reason why in a Resource Committee you don't have clinical evidence for not doing something in general.

You know, I can't come back before the next call. I could pull up the studies, you know, when I get back to the office that show that peri-op testing has been shown not to have evidence, and I can give you, you know, multiple

citations there. And this one, testing on asymptomatic patients does not generally contribute to anything else.

And you know, I think harm is a higher bar -- I know, Sana, you asked for that, but I don't know that you have to harm patients not to do something to them when there's no clinical value. I mean, I think most of our appropriate use, especially around imaging, is about if you're not getting clinical value for the patient, then I don't have to harm the patient in order to say I shouldn't do it. I mean, it's my ethical responsibility not to do something to a patient that has no clinical value.

And I think all these measures on the evidence would pass from they do not provide clinical value to the decision-making process. And as I said, you know, there's been outcome studies now, more so than most of the measures that we'll see before you today, showing that they actually align.

I mean, we re-proved that these do not

contribute to clinical value in outcomes. And so most of your other measures are derivative measures on process things that don't have direct impact on outcomes. We went through the trouble of actually showing that they do not provide clinical value, which is the bar that we thought -- at least the Efficiency Measure Group wanted us to show, and now today we're not passing them based on other considerations.

So you know, again, I will just say that, you know, I will make the pitch for this one again that there is no clinical value that has been shown. We have outcome studies stratifying appropriate versus rarely appropriate, and they're not contributing to the decision-making process.

We have studies showing that people have implemented these measures and reduced resource use dramatically, and avoided doing things to patients that they don't need to have.

And now, I recognize we can come back and provide additional things in the forms that

may have not been apparent, and bring those out based on these questions, but you know, I guess, you know, like others have said, we needed to understand what bar we're passing. If it's the harm bar, I don't think we'll pass the harm bar because I'm not going to argue whether radiation in an individual patient on an individual case when you're 80 years old is going to ever result in cancer. I mean, I think those arguments are kind of funny to have.

And you know, I think the 40 year olds on the asymptomatic, low-risk patient that, you know, we see a lot of 40 year old women get nuclear tests for whatever reason and we're not sure that's the best -- they should have an echo or a stress treadmill, but you know, those are the types of things that we're really looking at.

But so harm is hard to show, and I'll never come back with evidence that shows, at least on these, that we're harming patients by doing things, but they're not benefitting, and that's the bar that we set.

MEMBER CHO: I -- this is Leslie, sorry to jump in.

For the measure developers, since this was endorsed many years ago, do you have any data that there has been improvement? Measure 0671?

MR. ALLEN: Yes, we have seen improvement, and there are two studies cited, as well as when we get to those places, that we pulled them into the measurement form. I mean, most folks have improved, you know, on -- between 30 and 50 percent reduction in rarely appropriate tests, which means you're cutting a significant number of these tests, and especially, you know, as you look at these particular measures both pre-op and routine use after PCI, centers have done a lot of effort to avoid that.

And it used to be an annual thing, just like a dentist appointment, you'd come back in for your nuclear tests. And we've reset that expectation over the last ten years. Now there are still pockets where people do that, but, you know, it's because of measures like these that

we've reset expectations, and you know, it's been endorsed under Choosing Wisely.

MEMBER CHO: Why doesn't the measure 0671 include bypass surgery? Why does it only include PCI?

MR. ALLEN: We saw that it was the more frequent use of kind of the routine testing. We didn't see a lot of folks routinely testing folks after CABG surgery.

And there were some questions, as you got further out from CABG, about its use, and so, you know, this was the one that we saw repeatedly in our studies that came up as a routine thing that we wanted to focus people on. It's not that CABG wasn't sometimes an issue, but it was less frequent, so we decided to focus on this.

CO-CHAIR GEORGE: Tom and then Gerard.

MEMBER MARTIN: So I guess since Tom spoke about how he voted, I guess what I would say is kind of where you said you followed the rules for the level of evidence, it's really funny because I sat there as a pediatric

cardiologist and said we're always complaining that we don't have appropriate use documents in pediatric cardiology because we don't have enough evidence.

You don't become an appropriate use paper or guideline paper within ACC or AHA without a lot of evidence. And so I sat there and said, they've got evidence. And I know that because you don't become a manuscript without the evidence.

And I do know, looking at over the last several years the cost curve for imaging, and that not only that cost curve but actually some of the cost curve has been bent over -- since there's been greater awareness, the curve is being bent.

So I think it is, probably there is evidence, and it becomes a question of just kind of the declaration of the evidence and for people to know that. And there is risk to families with excess testing because now that cost is being -- you know, most of the cost that's now going into

the system has now been shared with the patient.

So if there is an unnecessary test done, that patient is going to be paying for it.

CO-CHAIR KOTTKE: If I could jump in that. Yes, I would agree. And we need to think of harm more broadly then a patient getting cancer or whatever, we need to think about social harm in this year's costs or next year's.

I mean, you know, they don't go just into space somewhere. They come -- every dollar spent in healthcare comes out of the patient's pocket somehow or other. And we need to think of harm broadly.

MEMBER JAMES: From working within several different health plans, this particular measure is really a tip of an iceberg. I would -- what I have seen is the wide variation in the use of this technology preoperatively, pre-procedure, as opposed to post-procedure.

And I wondered if that had not been addressed, the importance level would be really great.

MR. ALLEN: Again, the reason why we put these particular ones forward, they were the most common ones. And so we did see a lot of you know, in particular hospital-based settings where the facility was an imaging center based around a hospital that these ended up -- pre-op ended up being the issue in the outpatient cardiac space where you know, somebody was in Florida and coming as a snowbird and getting their annual treadmill with a nuclear scan, we saw it as a common issue.

They're slightly different settings of imaging facilities of where these are done and where those issues came up. And again, we picked the top ones that showed up in our studies.

MEMBER VIDOVICH: So for this measure 671, I was one of the reviewers. So I have some comments and I just wanted to see what the group thinks about it.

And I don't want to sound as a stickler. But I was bothered a little bit about the unintended consequences of this measure.

So I'm an interventionalist and I completely agree that tests doing a nuclear once a year after a PCI is wrong. And I don't do it. So I think the measure captures the gist what it's supposed to do.

But I think the problem is with some of the details, and usually the devil's in the details. First, the title routine. What is routine? How do we define routine?

What is routine? Once a year? Twice a year? Or does routine -- is this a surrogate word for asymptomatic? Right, that's one thing that I would like a better understanding here.

Next thing is all PCIs are considered the same, right. Well, they are not the same, right. Because some may be intentionally incomplete then you had to bring them back to see if there's ischemia. Some may be after a STEMI, some may be for stable angina, you -- there -- it's a big heterogeneity, which is all lumped in one category.

So sometimes you actually do have to

stress asymptomatic patients to find out what's going on. Another concern I have is well, asymptomatic may mean no symptoms, but some of these patients presented with newly diagnosed low ejection fraction after PCI for whatever reason, right. Which may not be captured in the measure.

And then you do have to further tease it out. Some patients have asymptomatic ECG changes. Some patients present with arrhythmia, which may be asymptomatic, which we may have to tease out. This is not captured in the measure.

So as an interventionalist, I have a little trouble about just the definitions in the title. And then another thing is that I may need some statistical help to understand, is about the numerator and denominator.

It says numerator, number of stress tests for asymptomatic patients within two years of most recent PCI. So that's fine. Denominator, number of stress tests. So that's number of stress tests within the facility? Or number of stress tests in patients who had

received the PCI? Number of what stress tests? That is what I don't understand.

So is this all the stress tests in a stress lab? That hospital stress labs? That interventionalist? Anyway, I'm being a bit long.

And then where does the two year come from? I would imagine from our cardiology data that PCI has a two year warranty, CABG has a five year warranty.

You know, that word on the street is, but you know, the evidence is somewhat limited. How does this end up in the title of the measure? Anyway, too long, I'm sorry.

CO-CHAIR GEORGE: Before you respond we have a comment on the phone. So I think we'll take that and then. Leslie, did you have a comment?

MEMBER CHO: Oh, no. I already said my peace.

CO-CHAIR GEORGE: All right. Thank you.

MR. ALLEN: Okay, so for the time

frame, the two years, that is the you know, what is considered routine within two years. And so you know, the measured title is a generic description. Of course the specifications actually come out and define what the routine is.

And so of course, you know, you can use different words in general titles. And we picked this one to try to communicate you know, in general. But when people drill down, it is the two years.

The two years did come from -- there were studies that we provided within the appropriate use criteria development that looked at when you might want to look at this. And there were periods around three years where some people were looking, maybe you need to bring them back for various reasons.

And so we didn't want to put it out at three years because it was starting to show that there was some reason to bring them back. We set it at two years because that is where most of the studies you know, started to show a change after

two years. And so we put it at two years trying to avoid the annual testing.

The incomplete revascularization are other reasons for bringing the patient back related to stenting. We do have -- remember this is on a clinical registry data sheet where we're looking at a universe of potential indications for why the test is done. This is not like a claims based measure where we're relying to other reasons why patients might be getting these tests.

So the physician has the opportunity to indicate those other types of reasons. Whether they be related to complications of acute coronary syndromes, a staged procedure. We actually have a specific indication on incomplete revascularization and wanting to look at.

So this measure is from a clinical registry that looks at a universe of things. And then picks out -- if you haven't selected all those other things and you're just saying you're doing it for this reason, then it is rarely

appropriate to do.

And so we are covering that. We can't cover it in the universe of measures. If we measured all 60 different indications, we'd have 60 measures in front of you.

And so picked only the ones where people explicitly said that they were doing it for this reason and this reason alone. And they had the opportunity to indicate other clinical reasons as you just said.

The systematic definition as well, covers a lot of different things. And so asymptomatic is the absence of those things. That's not just typical chest pain, oh my gosh, you know, I feel this crushing pain. It's a universe of symptoms that may be ischemic in relation. And so some of the things that you talked about would be covered there.

And you know, the denominator is facility -- is for the imaging facility. And so it is all the stress imaging orders at that particular facility to give us a broad dominator

so that we can get that population view and have the facility discuss that.

And if we put it at a facility level, it wouldn't be all that different from you have to measure it somewhere. And just measuring who got PCI or who got surgery, even CMS originally proposed looking at all patients that got surgery and then who got testing.

You end up with this huge denominator and very small number of patients. But it doesn't mean that those patients weren't meaningful in the universe of the people that got imaging. And so we defined it at the facility level.

CO-CHAIR GEORGE: Liz?

MEMBER DeLONG: I guess I'm seeing this as another example of switching the population to meet the numbers rather than the rationale. Suppose you've got -- didn't you mention a scenario in Miami where you're sending people for routine stress imaging. And the facility throws back most of those as

inappropriate.

So they end up with very few in the denominator that they actually do. And of those, maybe several are inappropriate. They get a low score -- or a high score, which is low because they really appropriately sent back most of them.

MR. ALLEN: So you know, the universe of patients that we're looking at remember, is all imaging tests that are ordered at that facility. And what you're looking at is the number that were done for reasons that don't have a good clinical rationale about them.

And so when you reduce that number of folks that are coming back for that routine, you are getting a lower score in this measure, which is better performance in this case. And so it is an inverse measure.

You're not affecting the denominator by not doing that per se. Other then you are -- well, so you are shrinking the number of patients that are getting imaging.

MEMBER DeLONG: You're not giving

credit for the ones that get sent back. Right?

MR. ALLEN: True. But we can capture the absence of an action.

MEMBER DeLONG: Right.

MEMBER VIDOVICH: Correct me if I'm wrong. Shouldn't the denominator be of all the PCIs that you've done, how many inappropriate stress tests have you done? Rather than within an image facility, which can have a wide variety of imaging tests required, right?

Because that's -- is that a quality measure of the intervention -- or the physician ordering or of the imaging facility? That's what I'm wondering.

MR. ALLEN: Again, this is accountability and who you'd want to have the discussion, I know we had the discussion in the last measure, would you want to hold accountable the imaging lab? I think it would be even more tenuous to hold accountable the surgeon or the interventionalist for a procedure that was ordered prior to them or after they performed.

And so we didn't believe accountability wise that it was appropriate to, not just for a numbers sake, but to tie it to somebody that was not even a part of the discussion about whether or not the test was done.

At least with the imaging facility, generally, especially now a days, these are large systems where the imaging facility is related to the folks that are ordering. And the accountability level is at least at a place where they could have a direct input.

I'm not sure that you would want to have a conversation about why PCIs had imaging done before or after.

MEMBER VIDOVICH: So as written, as I understand it, this looks at the quality of the imaging facility, not of the person ordering the test after PCI, right?

MR. ALLEN: Right.

MEMBER VIDOVICH: That, yes? Okay. All right.

CO-CHAIR GEORGE: Judd?

MEMBER HOLLANDER: So, who typically orders these tests? I guess it's getting a little bit of what Liz was talking and a little bit what we're saying here is the imaging facility we know doesn't order it. We know the interventionalist probably is not caring for that patient anymore.

And so we haven't figure out -- what's that?

CO-CHAIR KOTTKE: Yes they are.

MEMBER HOLLANDER: Are they?

CO-CHAIR KOTTKE: As a practicing cardiologist, I think there's a lot of that that's going on.

MEMBER HOLLANDER: Okay, so I'm just say --

CO-CHAIR KOTTKE: A routine, two months after you get the angioplasty they order it and they get you know.

MEMBER HOLLANDER: Right. Okay, so my question is actually, do we have any insight into

who orders the majority of these tests? Is it the cardiologist that referred them to the interventionalist? Is it the primary care provider? Is it the interventionalist?

You know, and so I'm trying to figure out sort of the you know, in the court, the chain of evidence or chain of responsibility. And where would be the right person, and maybe it's all over the map and it should be a health system issue.

But I think we should figure out who's likely to fix it and who's responsible for it.

CO-CHAIR KOTTKE: Well, if I can answer, a couple of questions. Liz, you know if you have 20 percent -- say you have 100 patients that are sent, 20 of them are inappropriate. If you send back 19 of them your score goes from .2 to 0.12.

So in fact your score gets better by sending. Unless you have a very high proportion of inappropriates.

If you're doing 100 tests, 20 percent

are inappropriate. You send 19 of them back so you're only doing 81 tests and one inappropriate, your score is .012. Your score is better by.

MEMBER DeLONG: Yes, but you're not capturing the fact that --

CO-CHAIR KOTTKE: That we're being good guys?

MEMBER DeLONG: You know, if it gets large, you're really not capturing --

CO-CHAIR KOTTKE: Yes, they're not -- people don't do it -- I hope 80 percent inappropriate tests, you know.

MEMBER DeLONG: Yes.

MEMBER HOLLANDER: The fact that it's not being caught, and not -- if you were measuring it on the interventionalist and talking about what percent of his caths subsequently got an inappropriate stress test, the numbers you're saying are right. And you're then attributing it effectively by the denominator as a relationship to the interventionalist.

But the denominator here is the number

of stress fact --

CO-CHAIR KOTTKE: Right. No, I'm talking about the number of tests done by a testing facility, not a cath.

MEMBER HOLLANDER: Right.

CO-CHAIR KOTTKE: But, you know, we have a bit of problem here because if we accept this one and don't accept the last one, what do we say, well we just made a mistake the first time and we didn't.

And so what I would sort of as a -- what?

CO-CHAIR GEORGE: I think there's been some other concerns raised on this one that we didn't raise on the last one.

MS. HIBAY: This is really just to answer Liz's question. I think that you were posing how many were ordered and then how many were completed.

And so -- I think that's the variation of the two -- the two data points that Tom was referring to. So I think what you're suggesting

is, is you know, so did they start off on the wrong foot, someone caught them. And they went back. And that's not being completed.

But this one is how many patients got the test as opposed to how many were inappropriately ordered, but the system caught it and then they went back. I think that's the difference between those two populations if I understand the question correctly.

CO-CHAIR GEORGE: Mladen?

MEMBER VIDOVICH: I'm being difficult. But there should be some allowance for -- adjustment for the baseline differences between the imaging facilities. Or different imaging facilities will have different mix of patients, right?

And this measure will not capture this, right. You know, so if you are maybe, I don't know, facility A, maybe your percentage will be different than the others. And then how do you account for that, right?

You know, because some may be

penalized because they get more of one and less of the other. Because it takes the aggregate of all stress tests of variety of the patients that come in, right?

And so the different test facilities will have different tests. I mean, I know that myself. At the VA, I have a complete different population at the University. And they're just a half a block away, right.

And then you couldn't call them to the same standard I think, without some adjustment or based on differences.

CO-CHAIRMAN GEORGE: So, is that concern I would -- just, is that concern related to the evidence or related to the specifications?

Okay. Any other comments on the evidence for this measure.

MEMBER CHO: So, it's Leslie, and my final comment is this. If we follow the algorithm, the evidence is low. But I still think based on the intention of the thing, my feeling is to let this measure pass on the

evidence.

Now, here's the thing, if this measure was initially approved by the Resources Committee, then maybe it should go back there for their re-endorsement. But honestly, if this is the final stop for this measure, I would be kind of sad to see this measure fall.

I mean, the evidence is low. I mean, there's no way to get around. By the NQF algorithm, there's no way to get around that. The evidence is low.

But I still think the intention of this measure is so good that it should go to someplace else to have it be re-endorsed again.

CO-CHAIR GEORGE: We are trying to find out whether we can get any more clarity on that issue. We haven't -- don't have an answer yet.

MS. JOHNSON: I will remind people that you do have the option of insufficient evidence with exception. So if that is something that you feel strongly about, you could vote

option five here, and it would go on for additional discussion.

Again, we also have the option of potentially having the developer come back and add to his submission to beef up that evidence if you think it exists. Which you know, I think I heard the developer say that it does.

So it's a little hard for me, I'm not clinical at all. It's really hard for me to understand. Was it a -- just the submission itself is not quite where it needed to be for you guys to be able to vote a higher thing on evidence.

In terms of turning it over to the Resource -- cost and resource use project, that's certainly not something that I could say we would do right now. And I think at least until Helen gets back in the room, I think we have to assume that it's going to stay with CV.

The other thing that I will point out is that we talked about this earlier, our criteria and our guidance. And how we ask you to

apply the criteria evolve over time.

So it was passed earlier in a different Committee, but that was in I think 2011. And our criteria and guidance have changed.

So the idea that something went through before and doesn't now, should not necessarily be surprising because of the evolution of the criteria and the guidance.

MEMBER DeLONG: But I would say if we were to vote on this as insufficient with exception that perhaps there's other people who would have changed their mind in the room on the previous measure.

MS. JOHNSON: And if that is the sentiment, then we could potentially go back and do a revote.

CO-CHAIR KOTTKE: I mean, I guess, I for one didn't understand the implications of insufficient with exception. I thought that was just you know, like insufficient and here's some things.

But I didn't realize things would go forward on an insufficient with exception. And I don't know if anybody else made that same interpretive error. But I didn't.

MS. JOHNSON: Right. So let me explain. This is a must pass criteria. So the way that it would must -- the way that it would pass would be if it gets at least 40 percent and either high or moderate. Or if it gets that level in the insufficient with exception.

So what you're saying is I think that there is not -- if you vote for number five, what you're saying is, I do not believe that there is sufficient evidence to pass this. However, I believe that the benefits versus the harms is such that we would be willing to offer an exception to this measure.

MEMBER DeLONG: Could I just say, I think it is a shame if we killed the first one and pass this one. The first one had a lot more clarity to it.

For this particular one, I would like

to emphasize what Mladen said. Because the denominator really does depend on the patient mix and what they're being sent for.

And I think Tom's example was a good one when you have small numbers. But when you have big numbers and a lot of them are for other reasons, and they're in the denominator and you send back your 19, you're not going to get credit for sending those back.

CO-CHAIR GEORGE: I'm going to take the prerogative of seeing if you would like to do a show of hands of going back and revoting on the first measure given Karen's explanation of insufficient with exception.

MEMBER AL-KHATIB: I have a question about the explanation that you gave though. Because I think we certainly need more guidance from you as to what the exception refers to.

You talked about you know, the net benefit and harm. But that's actually pretty vague in my mind. Like how much benefit are we talking?

I mean, does that ratio have to be really high for us to make an exception? Or even if the potential benefit might outweigh the potential harm, we're okay to use that response?

MEMBER HOLLANDER: Can I add before you answer, because it will be a coupled answer to Sana's comments. My interpretation to what we said in the first measure is nothing is dead if it doesn't pass. That they could come back on the conference call a couple of weeks later.

And I think we're all acting like that vote means it's dead and we can never talk about it again. And at least what I thought I heard is they could come back, revamp the measure in a couple of weeks. We could talk about it on the conference call and pass it at that time.

So I'm wondering if this is actually a big deal? Because right now we would be passing something we don't think is perfect. And I frankly like the idea of making them retool it, come back and have a three week delay. And it would still be out there in the real world on the

exact same date if they did that.

Am I hearing that correctly?

MS. JOHNSON: You are hearing that correctly that we could ask them to bring back and just beef up the evidence section.

So I think the question for you guys is, do you think that he can beef up the evidence section enough, knowing what you know, so that in three weeks, you would either vote moderate or high. Or you would still go exception with evidence. I think that's really the question for you.

In terms of the guidance about exception, the exception option is something that we hope is rare. Because what we would really like to have is evidence based measures.

But with that said, you know, we don't give any kind of a threshold. So this one really is your gut feeling about whether you think it is -- it deserves an exception if there isn't adequate evidence.

CO-CHAIR GEORGE: And Sana, to address

the question about rate it as insufficient with exception, I'll just read you from the algorithm.

Does the Steering Committee agree that it is okay or beneficial to hold providers accountable in the absence of empirical evidence of benefit to patients. Consider potential detriments to endorsing the measure, focus attention away from more impractical practices, more costly without benefit. Divert resources from developing more impactful measures. If the answer to that is yes, then rate as insufficient with exception.

MEMBER HILLEGASS: Could I also ask that when might we find out if this could be transferred to a different group? Or this would possibly be transferred to Resource?

Because that would make a decision -- my decision on how I vote. If we got an answer today, then we could table this until later to vote on.

MS. JOHNSON: Well, I think Helen would be the one that would make that decision.

And I think she is coming back today. I have emailed her. We've been -- but she hasn't gotten back to me. She had a meeting today.

So to be honest with you, I don't know when she would be able to make that determination. I will tell you that the cost and resource use projects really are much more technical in nature then these are.

These are the groups that will get episode groupers. And they're looking at kind of a different flavor of measures than what this is.

So again, I don't know what Helen's decision would be. But that's what I can tell you about that.

CO-CHAIR GEORGE: Well, I'm going to go back to my original question about a show of hands of people that would like to revote on the evidence for the previous measure, which was 670.

(Show of hands.)

CO-CHAIR GEORGE: And any from the phone? Wunmi?

MEMBER CHO: I mean, I'll go with the

majority.

(Laughter.)

MEMBER CHO: Am I the deciding vote? Oh God. No? What is the majority saying?

CO-CHAIR GEORGE: So that was seven to -- we're split on this.

MEMBER CHO: Oh, my God no. So here's the thing, it is my gut feeling tells me I want to kind of revote. But I mean -- I don't know.

MEMBER AL-KHATIB: But we will revote after we have more information from the developer and from you know, answers to the questions. So we're not saying we're not going to revote. We're asking just to delay the revote if that's possible.

MEMBER CHO: Okay.

CO-CHAIR GEORGE: So given that, we -- I think we're ready to vote on the evidence for 671. Carol?

MEMBER ALLRED: Yes, I still have a question. I guess in my mind I'm not clear. What are we setting the standards at? And do we

have a consistent standard that we're setting here?

Because if there is evidence out there, I would like to hear his additional evidence so that we had a clear -- a clear thing we're working on. Right now it doesn't appear that we're clear cut.

We didn't hear the evidence. We hope it's there. What are we voting for? Can he come back and bring us the evidence?

MR. ALLEN: Yes, let me ask again, what the evidence that we're really asking for? You know, we heard we wanted a harm evidence. I'm not going to be able to provide that.

If that's going to be indirect evidence that there is no benefit, but the trials that I would present to you would be on where it would benefit patients. And in many of those trials they also looked at populations similar to these that there would be a lack of benefit.

For this particular one, most of the evidence that appropriate use cites, is you know,

follow up studies that are observational. That you know, where is the kind of the cut point.

I don't know that I'm going to come back. People don't set out to do trials to avoid things. This is a very different type of measure. You're asking questions about evidence of things that generally would benefit patients.

And I feel like the evidence review process for NQF isn't set up well, at least in this particular circumstance to judge these measures. Because nobody is going to fund a trial to show a lack of benefit for something.

In general, I can only point to places where it does provide benefit. And a few studies that show that there's a lack of clinical benefit, as I provided already in the packet.

And so you know, I have some evidence, but I won't have a whole lot of evidence. And you know, there are other issues that came up around you know, whether that particular part of -- like this measure on PCI, you know, is it the best one to go after.

But you know, just on the general evidence, I don't know that I'm going to come back with something that's going to be so compelling versus what we've discussed today.

MEMBER VIDOVICH: I would just say, if you can just perform a cost effectiveness study that would look into cost as outcome measures, that would solve that issue, right. Because there's costs associated with each. I mean that's an easily measurable discrete granular measure.

MR. ALLEN: Right. And the information that's provided that we would talk about later in usability if we got past this evidence, would talk about the impact that we've had on resource use and the ability to change this based on this measure.

I mean, that's the application of the measure once you put it forward and you say there's enough evidence, I can show that it reduces costs. It changes how people are ordering tests.

Patient mix I know was asked. That's the whole point of this set of measures is to look at your patient mix. And so if you have a patient mix that you're ordering a lot of tests for appropriate reasons, great. That is going to show up in these measures because you're not going to be doing it for really appropriate.

These measures are again, population-based measures and telling you what your patient mix is. And in these three, they're saying you have patients that are potentially in the mix that aren't going to get clinical benefit based on what we know of who would get clinical benefit.

Asymptomatic patients do not benefit from subsequent procedures in this, except in the circumstances we already talked about, in complete revascularization, instability and other -- and those are already captured in the registry.

So I can bring back information. I just don't know that it's going to be

incrementally, you know, hugely more then what's in here.

CO-CHAIR GEORGE: Right. So our choices right now as I understand it, are we can vote on the evidence. We can choose to delay our vote. And delaying that vote could depend on what Helen has to inform us about if she's able to come back. Or could delay our vote until the developers come back with more evidence.

So delaying a vote could end up in one of those two scenarios. Or the third option is to vote on the evidence now. Any more?

MEMBER CROUCH: Yes, I'd just like to hear some data. You keep talking generalities about -- I'm all in favor of reducing unnecessary tests, don't get me wrong.

But I'd like to hear some specific data about how much -- how much money or how many tests you were -- the impact that this has had. And I'm happy to listen to data.

But I'm not convinced by you're saying we've got data and you should have read the

article, it's there. You've got to give us some facts please, to work with.

MR. ALLEN: We didn't get to the section that we would talk about those things because we haven't gotten past the scientific evidence. There are data tables in the presentation showing the things that we're talking about.

But we're not going to be able to have that conversation if we don't get past the science, so.

MEMBER BRIGGS: So I think one of the things that the developer is trying to say, and maybe we just need to say it a little bit different way. Is that ethically you don't do studies that have no useful impact.

You can't say I want to do this test to see if it would harm somebody. You just can't ethically do those tests. You wouldn't get past your IRB, the Institutional Review Board to do that.

So there's certain things that you

just are not going to have a randomized control trial for. I mean it's really hard to do randomized control trials on patients for example that are coming in in full arrest to a facility. Because who gives informed consent for that patient?

I mean there are some ways around things like that. But there are -- the randomized control trial, while we hold it in very high regard, is not always the be-all and end-all of all evidence.

So that looking at with exception might make sense for these types of indicators. Because again, the evidence is, is there more evidence that this might be beneficial than harmful? And then can get into the cost and all of those kinds of things down the line.

CO-CHAIR KOTTKE: But there are -- there are randomized trials that fail. I mean, for example, a very recent in the last week, a report of a randomized trial doing coronary CTs in patients with diabetes. Does it help? Does

it help manage -- reduce events? No, it does not.

And so -- but I think it was just you know, re -- I mean, just restating, stating a little more clearly. I mean, I think we're going to have a split vote right now.

Come back in a couple of weeks, I think the Committee would probably be fairly favorable just have some -- just a restating of the evidence, not looking for new evidence.

MEMBER AL-KHATIB: Yes, can I also just make one final comment if I may? I reviewed the first measure, the 670, not the other ones. I mean I looked at them, but not that closely.

But you talked about how you provided information about the cost under usability and use. And I actually couldn't find this information.

So I mean, I echo what was said by my colleague here, that if this information needs to be really readily available to us as we're reading. Because we are delving into like a lot

of documents to review. So if you can make that clear.

And if costs is what we're going to focus on, I'm all for it. I'm the last one who you know, would want us to order tests that patients don't benefit from them. But if we're going to focus on costs, we would want to see the data for that.

CO-CHAIR GEORGE: Okay, so Judd and then --

MEMBER HOLLANDER: I just want to say you know, this is supposed to be about the evidence. And I think we have an obligation to follow the algorithm, however imperfect the algorithm is. That's our charge to do that.

And I think you know, I think Ellen sort of said this before, you know, we don't want to layer more work on the people unless it meets the criteria.

I think this is a great thing. I think all these appropriate use things and driving down test results are critical.

But I don't frankly feel ethically right sitting around the table looking at a work around. If there's a process, we should follow the process. And a no vote on this measure now doesn't prevent it from going forward and hitting the public at exactly the same time it would.

I realize it's a little extra work for all of us and the measure developer to get to that point. But I just personally would feel bad about it from the work around.

Either the evidence makes it now, or the evidence doesn't make it now. And then there's a next step. It's not dead in the water.

There's no -- there's no reason why we should want to compromise on changing the process the way it's laid out.

CO-CHAIR GEORGE: So once again, we have a choice of halting our debate and coming back at a later point in time with the developer. Or we could vote on the evidence now.

MEMBER CHO: Are we still talking about 670?

CO-CHAIR GEORGE: No, 671.

(Laughter.)

MEMBER CHO: I'm like, God bless, okay. Let's go for it.

CO-CHAIR GEORGE: How many are in favor of voting on the evidence at this point in time?

(Show of hands.)

MEMBER CHO: Me.

CO-CHAIR GEORGE: And how many are in favor of delaying our vote at this time and coming back at a later time with further evidence from the developer?

(Show of hands.)

CO-CHAIR GEORGE: Did everybody vote? So we will come back at a later point in time.

MEMBER CHO: Oh, I just have one comment I want to make to the developer before they bring the 0671 back.

You know, when I was looking through the measure, the biggest problem I had was the numerator of the -- I'm sorry, the denominator of

the measure of 671.

So the numerator is I understand all the people who get a PC -- who get a stress test without symptoms who got PCIs. But really the denominator should be all the patients who had PCI without symptoms.

That actually should be the denominator. But that's not the denominator. And that is my biggest problem with measure 671. Do you understand what I'm trying to get at?

So, what we're trying to answer is this. All the patients who had PCI, let's say that's 100. 90 percent of -- 90 patients out of 100 didn't have symptoms. That should be your denominator. Not the denominator that's currently listed as number of stress spec -- MPI stress spec that go CCTA and CMR. That should not be a denominator.

MR. ALLEN: I understand what you're recommending. Our use of this measure, now nationally for many years, which has shown significant improvement and we'll talk about that

when we come back, is at the imaging facility level and in giving us feedback.

CMS failed miserably when they tried the peri-op measure by putting the denominator as all surgeries, even if they were just low risk surgeries. And the number of patients that received imaging as the numerator. It got to be such ridiculously small numbers, it was hard to differentiate.

And this particular measure, the way it's structured, again looks at case mix across all your imaging tests. And looks at three particular ones that are high frequency issues for rarely appropriate.

And so this should be looked at in general as a set. And it is looking at all imaging tests that are performed. And of those, what's your patient mix receiving it for rarely appropriate clinically without value tests.

And so I understand what people are saying. You know, could we put the PCIs underneath, could we put the surgeries

underneath, could we put all asymptomatic patients underneath?

I mean on the last measure, which I -- you know, we're deferring I guess all three, you know, putting all asymptomatic patients that show up in any health system as the denominator and then how many people got imaging tests, you have you know, in some health system, a couple of hundred thousand patients. And you know, maybe a handful you know, 100 or 200 that might receive it in that category.

The percentages would be meaningless for action and for improvement. We're trying to get to a collection of information that a facility and an imaging facility can use to work with their referring providers to actually impact change.

And a number of centers have done that. And so, we'll bring back the evidence. You know, I'll bring back as much as I can because we've touched on everything from cost, which is implementation of the measures, to harm,

which I don't know that I'll have. To clinical evidence around you know, is this valuable.

And you know, I guess I'll just bring it all back. Because you know, we don't get to the next section to talk about that if we don't get past that. And then you can use it as you see fit through the rest of the section.

CO-CHAIR GEORGE: Okay, I think we're going to break for lunch and come back refreshed for more discussion.

MEMBER CHO: What measure are we going to talk about when we come back? I'm sorry, which measure?

MS. HIBAY: 672. We're still going to have some conversation about 672 that's specific to the measure itself.

So when we come back at the post-call meeting, which is December 19th, we will have had some conversation about the measure. Not starting off from fresh.

(Whereupon, the above-entitled matter went off the record at 12:43 p.m. and resumed at

1:17 p.m.)

CO-CHAIR KOTTKE: Okay, so Sharon is going to lead us through what the order of worship is going to be this afternoon.

MS. HIBAY: Okay, so I think you saw us all feverishly discussing next steps for these three measures over the lunch break. So, I just want to kind of give you a status report on kind of where we are.

For measure 0670, that was the first measure, that was not recommended, based upon evidence. That will go out for public comment and we have asked the measure developer to come back with some additional information and we will reconsider that after the post-comment period. And then, as a group, we will revisit it on a post-comment call. Okay?

For 0671, the committee has recommended to defer that for discussion. Excuse me. Defer the voting until the post-meeting call. We have decided what we think we would like to do is move that also to the post-comment

call so we can talk about that measure. You know they all are in tandem. And we would like not to vote, the group has decided not to vote on 0671 today.

On 0672, we would like to have a vote for the same recommendation as 0671. A show of hands and show of voice on the phone to see if you are in agreement with deferring, and that would be the discussion and voting for 0672 until the post-comment call.

And Wunmi, I don't know if you were able to capture that date. March 18th is when would be the post-comment call. So, that gives our developer some time to bring back the information and represent this and also for us to capture more comments from our stakeholders as well.

So, just in general, are you okay with the plan for re-voting on deferring 0672 to the post-comment call as well? You are okay? Okay.

Show of hands of how many, please, and show of voices, are okay about deferring the

voting and discussion for 0672. Are we unanimous? Okay, and on the phone, please? Are you still at lunch or are you on mute? You can't answer if you are still at lunch, I recognize.

MEMBER GIBBONS: It's okay by me, Ted Gibbons.

MS. HIBAY: Okay, very good. And is Leslie on the call? She just left. Okay.

Okay, so by unanimous vote we have decided to defer 0672 discussion and vote to the post-comment call on March 19th. Okay, very good. Okay. Okay, very good. That's where we are.

MR. ALLEN: Thank you everyone.

CO-CHAIR KOTTKE: Thanks. Okay, so we are at 0900: Electrocardiogram Performed for Non-Traumatic Chest Pain. The discussants are Gerard, Jason and Judd. Who is going to lead off?

Oh, the developer will lead off.

MS. HIBAY: Is the developer on the line?

DR. CANTRILL: This is Dr. Steve Cantrill from Denver. I'm on the line. I think I was going to give the initial presentation.

CO-CHAIR KOTTKE: Okay, go ahead. The floor is yours.

DR. CANTRILL: Thank you very much. Chest pain, as you all know, is a real problem for us in emergency medicine. We see more than five million patients a year that present to our EDs. They present with non-traumatic chest pain. And more than 1.4 million of those patients end up being hospitalized for ST segment elevation in terms of having an actual heart attack, an MI.

So, this continues to be a very important issue for us. The EKG is instrumental in terms of determining whether a patient with chest pain in fact might be suffering an ST elevation MI. And this continues to be a very important issue.

We realize there is a minor performance gap with our latest data. We have many, many physicians that are still reporting on

this. They feel it is an important measure and we would like to have it continue be at the reserve status, if at all possible.

I will take any questions.

MEMBER HOLLANDER: So, this is actually right in my wheelhouse ad I am actually one the authors on one of the references on this. And I think it is obviously a critically important thing.

So, I think it is a critically important thing but I am just wondering if the focus of this, which has been out there, might be a little off and want to open that for discussion.

So, the numbers that were presented are actually really old numbers and it is really different now but they are focusing on people with disease. So, right now, although there is five to eight million chest pain patients who come to the ED with chest pain, it turns out the likelihood of having an MI, STEMI or non-STEMI is actually closer to the five percent range. It is

a much smaller number and many more of those are NSTEMI than STEMI, as compared to 15 years ago when the majority was STEMI. Now, it is actually the minority are STEMI by a decent margin. And the whole idea of the EKG is to identify the only time-sensitive thing we do in emergency medicine, which is reperfusion for STEMI patients. But yet this measure looks at people discharged from the ED and whether they got an EKG. One of the competing measure, 0289 looks at time or median time to EKG. So, there is great data on time to EKG. There is guidelines that say it should be done within 10 minutes of ED arrival. That is a critically important process measure that has been shown to correlate very nicely with outcome in early reperfusion. The earlier you get the EKG, the earlier you get your reperfusion.

This is the other end of the spectrum. These are the people nobody thinks have a STEMI or NSTEMI and they are going home. And although they don't present the data, there is a lot of data that circulates out there that says we miss

two to five percent of MI patients but that is way old data. And with the advent of observation units and everything else over the last 10 to 15 years since we realized we are missing tons of MI patients, it is really relatively rare. You know they show up as single digit numbers in malpractice cases.

So, the evidence is if you have a STEMI, you need to find it out fast. But the measure is nobody thinks you have a STEMI and did you get an EKG before somebody sent you home. And they are not the same thing.

So, I think measures on time to EKG for the patients that have cardiac disease are critical. I am not sure we are accomplishing much by doing this and the evidence really, again, speaks to the importance of detecting the STEMI patient and not the importance of missing the MI patient, which is really what this is focused on.

DR. CANTRILL: Well, you make some very good points. And one of the thoughts of

this measure is that, as you pointed out, that years ago it was a five percent number of missed MIs in patients that we sent home from the Emergency Department, which is terrible when you think about it. We have gotten better.

But still, the at-risk population here are those that where you don't think they are having an MI and in fact you missed it. And granted, those numbers are small. If you are the one that gets missed, the numbers aren't so small.

But it really is the at-risk population where people think oh, that chest pain couldn't be having a STEMI and we are not going to get an EKG. So, those are still the at-risk people. And again, we feel it is important and looking at even at close claims data that still does represent a problem for us in emergency medicine.

But I thank you for your comments. Those are very good comments.

CO-CHAIR KOTTKE: Other discussants?

Gerard, did you want to say anything?

MEMBER MARTIN: I did not have anything else to add on that.

MEMBER SPANGLER: Tom?

CO-CHAIR KOTTKE: Yes.

MEMBER SPANGLER: And I am not an expert like the other two gentlemen but there doesn't seem to be much of a performance gap here, based on this data. I mean it seems like this has done pretty well. I mean so I don't know if that is because of the measure or not. So, that was one concern that I had just because we are at the 95 plus percentile for this performance rate and even higher in the aggregate. So, that was one thing that I had a concern about.

MEMBER HOLLANDER: I will speak to that. The performance numbers are the 50th percentile is at 100 percent. The 25th percentile is 96 percent and change. The 10th percentile is less. It is 88 percent. But I think I would agree, there is little performance

gap.

I mean, that being said, you know I do med mal consulting and I see patients where people went home without an EKG every once in a while. And you know somebody may have missed an MI that day because something bad happened a couple of days later. So, there is clearly single digits, numbers of these cases going around. It is an e-measure. It is not terribly hard to measure. It is something that is important. I would agree the performance gap is relatively little.

DR. CANTRILL: I would agree with that as well, if I could respond. And again, your points are excellent.

Just looking at the numbers historically, again, how much of this is the Hawthorne effect, the effect of measurement on the measure we may never know. But our numbers have improved over the last five years that we have data for.

So, I think part of that is the impact

of the measure and we would hate to give that up because we think it does have a very positive impact in terms of the health and safety of our patients.

MS. TIERNEY: This is Sam Tierney with the PCPI. If I could just add to Dr. Cantrill's comment on the gap, I would also say that the PQRS program, which is the program from which most of the data came, most recently was a voluntary reporting program with minimal rates of participation. So, we would say that the rates suggested her, the performance rates are probably not nationally representative. And this measure does have a high reporting rate, I think around 60 percent of emergency physicians report on this measure but that is still not 100 percent.

I would also add that the measure, the PQRS program focuses on the Medicare population and the measure focuses on patients who are aged 40 and older. So, it is a little broader.

So, again, just to put the data that you see in a little bit more context.

DR. CANTRILL: Thank you, Sam, for those comments.

CO-CHAIR KOTTKE: Okay, Judd, do you want to -- so, this is evidence. Do you want to give us some guidance on --

MEMBER HOLLANDER: I think that the EKG is important. It is a little different interpretation. We have had a lot of conversations about did you show a medical, which is sort of a paradox. It just seems sort of better on the back.

CO-CHAIR KOTTKE: Right, so you have moved into opportunity for improvement.

MEMBER HOLLANDER: What?

CO-CHAIR KOTTKE: You have moved into opportunity for improvement.

MEMBER HOLLANDER: So, I think there are minimal opportunities.

CO-CHAIR KOTTKE: Yes, but we need to go back and vote on evidence.

MEMBER HOLLANDER: Oh, okay. Yes, so I think that that, personally, if I was deciding

for myself, it would be much more about EKG, than did they just get one. But I guess we can get an EKG.

CO-CHAIR KOTTKE: So, are we ready to vote on evidence?

MS. LUONG: Polling for evidence start now for Measure 0090. You can have 1 for high, 2 for moderate, 3 for low, 4 for insufficient, and 5 for insufficient evidence with exception.

The evidence criteria passed with 59 percent voting for high, 35 voting for moderate, and 6 voting for low.

CO-CHAIR KOTTKE: So, opportunity for improvement. Anything else you want to say, Judd?

MEMBER HOLLANDER: Twelve percent rates the 10th percentile and --

CO-CHAIR KOTTKE: Of a population of ER docs who are voluntarily reporting. Sixty percent of ER docs report this?

MEMBER HOLLANDER: Yes, I think I would ask the developer. I think it is actually

probably group practice reporting rather than individual physicians but it is 69,000 providers are reporting. So, it is a pretty robust number.

CO-CHAIR KOTTKE: Other comments? So, can I ask the proposer, there was a hint that this is a biased estimate, performance is biased upwards. Is that correct or not?

DR. CANTRILL: Sam, do you want to take that?

MS. TIERNEY: Yes, sure. So yes, I think we have seen with other measures in the PQRS program that the performance rates are quite high and oftentimes not consistent with the rates for the medical literature. I think that largely stems from the nature of the program being voluntarily, up until recently. And all the data we have is from when the program was a voluntary reporting program.

So, I think that tends to include people who are already performing well on these types of -- on this aspect of care. And as a result, I do think the performance is skewed

upward.

CO-CHAIR KOTTKE: Any other discussion or should we vote on opportunity for improvement? Linda.

MEMBER BRIGGS: Are we talking about disparities, too, here, in this section?

CO-CHAIR KOTTKE: Yes.

MEMBER BRIGGS: Okay because one of the things that is in the bullet points is that there were no data related to disparities here. And I think one of the areas that is of most concern is groups that are listed, it would be good to have some data if there is data available about older people, females, in particular, and non-white patients and exactly what does happen in those groups? Because those are the people who get missed and then there are consequences because of that.

CO-CHAIR KOTTKE: Other comments? Are we ready to vote? Let's vote.

MS. LUONG: Voting starts now for performance gap: 1 for high, 2 for moderate, 3

for low, and 4 for insufficient.

For performance gap, we have 59 for moderate and 41 for low. We'll keep going.

CO-CHAIR KOTTKE: Okay, priority.

MEMBER HOLLANDER: So this is where I think I was a little tougher on this one because we have done the improvement. So, identifying patients with STEMI is really important. Identifying patients to make sure they had an EKG when you think they nothing and presumably had some other explanation before you send them home, which is who this measures, is ED discharges, I don't believe now has the same national priority and I don't believe they compare to ten years ago the missed MI rate is the same. I mean troponins are identifying patients and pretty much everybody gets an EKG and a troponin no matter why they are there these days.

And so I think the missed rate, there is no real data but I think it is now very, very low. And so I think the degree of priority this may have had when it was approved is probably

less now.

CO-CHAIR KOTTKE: This is an e-measure. So, does that change it up?

MEMBER HOLLANDER: Well, I think it makes it relatively easy to measure. I mean there is one or two things we will talk about logistics later when it comes up. But I think if it is a well-done e-measure and it is not really hard to collect and could report back, then I think there is no harm in doing it and it is not unduly burdensome. But I still think the priority itself is not high.

CO-CHAIR KOTTKE: Anybody else care to weigh in?

CO-CHAIR GEORGE: I notice that says this is a National Quality Strategy priority. Is that correct?

CO-CHAIR KOTTKE: They all say that.

MEMBER HOLLANDER: I'm not going to comment on the specifics but I think the specifics are to identify the acute MI patients.

And you know just because there is

obviously a lot of cardiologists in this room, one good example is like identifying hypertension is usually predictive of someone having an MI and following them years later. Knowing somebody with chest pain in the ED, has hypertension has zero predictive value for whether or not they are having an MI now.

So, not everything that is really important in the outpatient setting or for the specialist is important in the ED. So, knowing that, identifying patients with STEMI early is usually important, everybody agrees, looking at the EKGs on the people you think probably don't have STEMI in the small percentage you are missing. It probably doesn't have the same importance.

So, you have to separate out which end of the spectrum you are looking at a little bit, as you think about this.

MEMBER SPANGLER: Mary, I was going, to address your question, I feel like almost all of these measures, when it comes to the National

Quality Strategy, says it is a priority because of the one bullet that says starting with cardiovascular disease. So, these are all cardiovascular disease, they all say it is an NQS priority.

So, I think Judd saying we need to dig a little deeper, is it really a priority or not just because it mentions cardiovascular disease and the National Quality Strategy and all of these are cardiovascular measures doesn't mean they are all high priority.

CO-CHAIR KOTTKE: Right, everybody loves their mother.

Okay, well we looked at trying to use ECGs to identify non-STEMIs and off the readings, it is nearly impossible because they are read hours or a day later.

So, are we ready to vote on priority?

MS. KAYE: This is Toni with AMA PCPI.

CO-CHAIR KOTTKE: Yes.

PARTICIPANT: I guess we just wanted to comment in terms of looking at the questions

in terms of priority for the committee, you know does this address a significant health problem, either high prevalence, high severity, high cost. So, I think we would make the point that chest pain is a very high prevalence issue and the severity, if it is missed, if you miss an MI, even though it may not be that common, it is the type of event that has, we think, large consequences, if it is missed and so we kind of view this almost in light of like a never event.

And so we make that as our case for the high priority.

CO-CHAIR KOTTKE: Thank you. Mladen.

MEMBER VIDOVICH: I mean it is not the most specific or sensitive test. As you will notice, you mentioned in the emergency department, everybody gets a battery of tests, no matter whether they need it or not and probably ultra-sensitive troponin will be out soon in the United States or in the world and that will probably be a better test than everything that walks with chest pain gets an EKG before they are

discharged. I think that is what you somehow alluded. I maybe have not used the most refined language but I mean it is a cheap test. There is not risk.

CO-CHAIR KOTTKE: Right.

MEMBER HOLLANDER: I just have to say this on priority. It is not meant to be funny but it might actually seem funny. The appropriate use discussions we have had for the last 12 hours, they pick up a higher percentage of people with disease than an EKG at the time of discharge than somebody you think is wrong is going to pick up a missed MI.

So, missing an MI is bad. I am not trying to downplay it. But the missed MI rates 15 years ago were two to five percent. Now, we are trying to look at people that someone thinks has pneumonia or has a clear shoulder sprain or something and get an EKG on them. It should actually be a one percent event rate. That's fine. It is a cheap, it is an easy test. It is harmless. I have never seen anybody get hurt

from it. It might lead to additional testing if you find some other things. It is not without unintended consequences. And so I just want to frame it that you could do things to find rare events. This is one of them but so are all the other things that we tried to do away with over the last two or three hours.

MS. TIERNEY: This is Sam Tierney with the AMA PCPI. If I could just add a clarifying comment. I think it was Dr. Chiu, I think that might have been you who was just speaking. I think you mentioned that the measure looks at whether they got an EKG on discharge. And that is not, I just wanted to clarify that is not the focus of the measure. The denominator is, in terms of sort of looking at this retrospectively, we are looking at patients who had an ED discharge diagnosis of non-traumatic chest pain and then seeing whether or not they had an EKG performed before that but not necessarily at diagnosis -- I mean at discharge. We would expect it to happen more at

presentation.

So, I just wanted to clarify the timing around the measure and that is actually how it is constructed.

MEMBER HOLLANDER: Okay, no, I understood that. But let me ask you one clarifying question. When you say ED discharge, does that include patients admitted to the hospital or is it just people actually physically discharged home?

For example --

MS. TIERNEY: We would have to look into that. I think it includes patients discharged home and to the hospital but we would have to look at that more in our specifications.

MEMBER HOLLANDER: Yes, because I can't tell that and that would actually be an important point to me because most chest pains admitted to the hospital, they are obviously getting an EKG. So, the way I read this, and maybe it is incorrect was just for ED discharges, not for patients within the ED stay, which would

include observation, admission, and discharge home. It would be helpful to know that.

MS. TIERNEY: We could certainly look more closely at our specifications and add that clarification.

CO-CHAIR KOTTKE: Other comments? Mladen, you have commented?

MEMBER VIDOVICH: Yes.

Okay, it looks like we are ready to vote on priority.

MS. LUONG: Polling for high priority starts now; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

For high priority, 6 percent voted for high; 47 percent voted for moderate; 47 for low.

So, can we keep going?

MS. HIBAY: So, this measure has not met 60 percent but it is in the 40 to 60 percent, which it means it is in the gray zone. So, what we do is we just keep moving forward with the measure discussion.

CO-CHAIR KOTTKE: Thank you.

MS. HIBAY: Is that correct? Yes. Sharon is showing she is a bit of a novice.

CO-CHAIR KOTTKE: Scientific acceptability.

MEMBER HOLLANDER: Or do they come up on the next slide?

CO-CHAIR KOTTKE: Well, yes. Talk about reliability.

MEMBER HOLLANDER: Okay. So, from a reliability point of view, they did testing in one urban medical center and one EMR, which I don't think is terribly adequate. They did have 100 percent agreement for the numerator, which is whether they got the EKG and only 94 percent agreement for the denominator, which is, I guess patients in the ED or non-traumatic chest pain, patients being discharged from the ED. They did not give a kappa value for that.

And Liz could weigh in. I mean those numbers sound like they are good numbers but I think the kappa could probably be wildly

different within those numbers. So, it is hard for us to really say there is robust reliability. My sense is that this should be reliable and not that difficult.

But if it is an e-measure, it should probably be tested in more than on EMR because there is a whole bunch of EMRs that need to make this work and it probably should be tested in one medical center.

So, I don't think there is very high reliability testing here.

MS. HIBAY: Do you mind if I just put a little input into there? So, based upon the time constraints of this submission, this actually came to us last year and it is quite extensive, protracted, and onerous to do the EHR testing. We did give a wave, and I did put that in, in the preliminary analysis related to the testing. We gave them a wave on our preferences, three different EMRs at three different sites. So, we did say that for this measure it is acceptable.

Future e-measures will be required to have testing in three EMRs in three different sites. So, just to be clear if that drum keeps getting beat, that they should not be held to that level of account.

CO-CHAIR KOTTKE: I've been instructed to ask you about numerator, denominator exclusions, beta source.

MEMBER HOLLANDER: Let me pull up and read it exactly. The numerator, I believe was, and I don't see it right here and I don't have it on the notes, was basically -- oh, here it is. Patients who have a 12-lead EKG performed the denominator was all patients aged 40 and older with an ED discharge of non-traumatic pain, which seems reasonable. The exclusions are -- and this is not reasonable, is the medical reasons for not performing an EKG being documented. And I can safely see having reviewed tens of thousands of charts in my research studies. I have never seen somebody write I did not do an EKG because of.

So, there are effectively no

exclusions.

CO-CHAIR KOTTKE: So, does that make the exclusions inappropriate? If a tree doesn't fall in the forest --

MEMBER HOLLANDER: So, I think it sort of makes them a little funny but the goal is there is really, I can't think of many reasons why someone should write I didn't do an EKG. You know, what is that, all four limbs are amputated? I mean they can still do that.

So, I don't see it as being terribly relevant but I think it is not terribly well-addressed either. So, I could live with that exclusion, even though I think it is a little silly.

CO-CHAIR KOTTKE: Okay, yes, Liz.

MEMBER DELONG: So, Judd, your previous question didn't seem to get an answer. And that would have implications for the denominator. Are you satisfied that the -- I mean the developer wasn't clear on the denominator either as to whether those were

patients who were not admitted to the hospital or came through the ED and -- does this clarify it?

MEMBER HOLLANDER: So, I don't think I have a good understanding of that, as you obviously have. I think, however, the reliability of that is probably not different. So, whether they want to say it is all admitted and discharged patients or just discharged patients, I believe both of those should probably be reliable to pull out.

Obviously, before we approve a final measure, I think we need to know what we are approving with certainty and I don't know the answer to that. But I think I can be okay with the reliability.

MS. TIERNEY: This is Sam Tierney. We just want to add we have since looked at our specifications and apologize for not knowing earlier, Kim, my colleague, will speak to that issue.

MS. SMUK: Yes, so our specification does not limit the population based on their

discharge disposition and where they are going to. So, it is open to all discharge locations.

MEMBER HOLLANDER: So, let me just ask you, and maybe you don't have any insights into this, but your agreement for the denominator was only 94 percent. And based on what you just said, all they basically needed to do, I guess the non-traumatic chest pain would be where there is disagreement or lack of concordance over whether they meet the criteria. I am trying to figure out where there could be disagreement on the review in the denominator.

MS. TIERNEY: Well, for the disagreement on the denominator, what was found on the automated report versus the abstractor was different for three cases. But otherwise, for the other cases, they all were in agreement just for the denominator.

MEMBER HOLLANDER: Thank you.

CO-CHAIR KOTTKE: Okay, seeing no other action, let's vote on reliability.

MS. LUONG: Polling for high

reliability starts now; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Reliability passes with 18 percent for high, 82 percent for moderate.

CO-CHAIR KOTTKE: Validity.

MEMBER HOLLANDER: Yes, I have no issues with validity. You sort of either got an EKG or you didn't get an EKG. It is kind of a simple concept. I think it has face validity and it is an important thing to be doing for these patients.

CO-CHAIR KOTTKE: Seeing no other urge to comment, let's vote on validity.

MS. LUONG: Polling for high validity starts now; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

CO-CHAIR KOTTKE: People on the phone, be sure to comment, if you have the urge to comment.

MS. LUONG: For validity, 65 percent voted high, 35 voted for moderate and it passes.

CO-CHAIR KOTTKE: So, Judd, an ECG is

not just an ECG. Feasibility.

MEMBER HOLLANDER: So, this I actually have questions based on the comment that Sharon had. You know if we expect everybody to do an e-measure but yet we don't expect it is reasonable to ask them to test it within a year's period of time, well, then how feasible is it to really do this?

And so I wouldn't have thought of that until I heard Sharon's comment. I thought they just did it because it was easy with one EHR. But if sort of the official explanation is it is really, really hard to do this, then it is not terribly feasible.

CO-CHAIR KOTTKE: Sharon, do you want to explain?

MS. HIBAY: Yes, so, there is a difference between the measure being feasible and testing within feasible. So, testing itself requires the recruitment of both vendors and as well as practices. And you know when you put a testing model together and I say this with

experience, as an e-measure tester and an e-measure developer, it is very hard in this environment we are in, people trying to work toward meaningful use, people trying to work with the ICD-10 conversion, people working with all sorts of constraints to get people to buy on.

I can tell you when I did my recruitment for e-measure testing, well, how hard could it be? It took six months and I came begging, screaming, kicking, trying to get people to come onboard because there are so many pulls, twists, and turns that practices have coming at them. So, it really is hard.

The ability to get testers is not the same as the measure is feasible. The feasibility of testing itself, the process of getting people to test, it is a little challenging right now.

So, one of the things I also want to share with you is we were in very good discussions with the developer about the process and about testing itself. And this measure, initially, was potentially going to come to us in

phase 1, as opposed to this phase in phase 2. And so to get testing up and running, they would have had to have it completed when it was a very new requirement as of October last year and they never would have been able to get it done by December of last year, when phase 1 measures were due.

And so by the time we were able to make decisions about what was phase 1, what was phase 2, it still left them with a gap. And we recognize the onerous amount of effort that goes and expense, et cetera, that goes to testing. So, we gave them a wave on having to have three EMRs at this time.

Just to let you know, the data does come from data abstracted or reported from and EHR and it was identified. And Sam, can you correct me if I am wrong? It was identified and tested in 2010, based upon claims generated at that time for visits during that time. But the data itself was abstracted from EHRs and other electronic data.

PARTICIPANT: So, this is Joy from the AMA PCPI. At the time the testing project was conducted in 2010 on data pulled from 2009, the feasibility of the data are not required for certified EHRs and we did encounter some difficulty with capturing the measure exceptions in a structured format. But since that time, we were able to determine at the site that 93.6 percent of the data was stored in a codified format and documented in an ED EKG flow sheet. And all of the information was pulled from the EHR.

MS. HIBAY: Does that provide you with the answer? Okay.

CO-CHAIR KOTTKE: Okay, any urge to comment on feasibility? Let's vote on feasibility.

MS. LUONG: Voting starts now for feasibility: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Feasibility passes with 29 percent for high, 65 percent for moderate, 6 percent for low.

CO-CHAIR KOTTKE: Usability and use.

MEMBER HOLLANDER: I don't think I have much to say over the discussions we have already had. And this seems like this should be able to be done and be useful for the people at the tail end and be laggards and move them up.

I think in the future we should look and see whether or not there is going to be continued improvements and there will be utility down the road but for now, I think we have made it past that point. I thought it was reasonable.

CO-CHAIR KOTTKE: Any discussion?

MEMBER SPANGLER: Tom, just a quick comment about I know this happens with e-measures but I am always a little concerned when there is no public reporting of this and it is planned to be done because we never really know if it is going to happen. I mean we are hopeful that it will happen but just a caution because I think this will come up again and it has come up in the past.

I am not saying that we should rate it

low usability because I agree with Judd, I think the usability and use will probably be good. But I'm always just flagged that just because there is no public reporting on this at all but we are planning on doing this in the future.

MS. HIBAY: So, Sam, do you want to give some clarification on the current state of reporting now?

MS. TIERNEY: Sure. So, the measure is being used in PQRS. The PQRS program, as we noted in the form and as many of you are aware has publicly reported some measures but this measure is not currently one of the ones that is publicly reported. I mean we know we are moving towards that and having more of the measures in the PQRS program to be publicly reported but this one currently isn't one of them.

MS. HIBAY: Correct. And that is in its current state in claims and registry. This is an EHR submission and so just for the committee's use, you will see this, as Jason is saying, you will see this come again with other

measures that there is a plan to submit meaningful use three or whatever. And a lot of that ends up impinging upon what we do here.

CO-CHAIR KOTTKE: Okay, ready to vote on usability/use?

MS. LUONG: Polling for usability and use; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Usability and use passes with 6 percent for high, 82 percent for moderate, 6 percent for low, and 6 percent for insufficient information.

CO-CHAIR KOTTKE: So, now, overall vote.

MS. LUONG: Polling starts now for overall suitability for endorsement; 1 for yes, 2 for no.

Measure 0090 passes for overall suitability for endorsement with 88 percent yes and 12 percent for no.

CO-CHAIR KOTTKE: Related competing measures, anything Judd?

MEMBER HOLLANDER: And I don't recall whether 0289 was just for STEMI or AMI patients or for everybody. I think it was just for MI patients. So, I think they are a little bit complementary. They are not really strictly overlapping.

CO-CHAIR KOTTKE: Okay, thanks. I think that is it for -- go ahead. Sorry.

MEMBER SPANGLER: Actually, I had a quick question for Sharon. Can you remind me the process of retiring a measure? Because I am trying to remember if it is the same as this process where we are looking at maintenance of measures we decide no, which is different than, obviously, rejecting the measure.

And because I think this is an example of I think some of us probably had the feeling, should this be retired now. Should we do one more round and then see what performance is and then maybe retire it?

But I wanted to know, is that the same as this process or is that a different process?

MS. HIBAY: Yes, I'll let Karen speak to the reserve status.

MS. JOHNSON: I think you are talking about reserve status. You may have remembered that from before. So, actually, I think since the last time you met, we have clarified our reserve status policy. So, what we would have done with this measure, if you had voted that there was no opportunity for improvement, that you felt that it was totally topped out, you would have been thinking about this only for reserve status.

So, if it had passed, we would have automatically now put that into reserve status.

Since you did not say with your vote that it was topped out, then it is just regular endorsement.

CO-CHAIR KOTTKE: Okay, that is it. The next two measures are 1524 and 1525. Sana, you have to recuse yourself. And I think those are the -- aren't those yours?

MEMBER AL-KHATIB: No.

CO-CHAIR KOTTKE: Oh, those aren't yours?

MEMBER AL-KHATIB: No, those are not the ones.

CO-CHAIR KOTTKE: Oh, you are right. Sorry. I woke you up.

CO-CHAIR GEORGE: Okay, the discussants are Leslie, who is on the phone, Judd, and Joel. And we will start with the developers.

DR. HO: Thank you. My name is Michael Ho. I am a general cardiologist from Denver, representing the ACC/AHA today.

As many of you know, atrial fibrillation is one of the most common cardiac arrhythmias in the U.S. It is estimated that between 2.7 and 6.1 million American adults have this condition. And it is expected to double in the next 25 years and it accounts for significant morbidity and mortality. And the cost for treating atrial fibrillation has been estimated to range from $6 to $26 billion a year.

So, given kind of the high incidence and prevalence of this condition, we are proposing two measures. Measure 1524 is about assessment of thromboembolic risk in patients with atrial fibrillation and then 1525 is about appropriate anticoagulation therapy in patients who are candidates.

Both of these measures have strong support from evidence. They are both class 1 recommendations from guidelines. Both of these measures were developed by the ACC/AHA and the AMA PCPI and they have been both through peer review and public comment process. And they have also been tested for feasibility, validity and reliability. And these were assessed through the PINNACLE Registry, which currently has 172 practices, 3,000 providers. There is over 3.6 million patients in the registry and there is over 690,000 patients with atrial fibrillation in this registry.

And so with that, I am happy to take any questions.

CO-CHAIR GEORGE: Okay, we will go on to our discussants to talk about the evidence.

Leslie or Joel?

CO-CHAIR KOTTKE: Leslie's not on.

MEMBER HOLLANDER: No? She's off. Okay, sorry you guys have got to listen to me some more.

I actually love this measure. There is only one thing about it I don't love. I mean the developers actually made a great case that this is a really important thing. The CHADS2 score is well-validated. There is no shortage of data. It is a great way to move the field forward. There is a total lack of appropriate anticoagulation in all cases, from the evidence to support the measure.

It is really good.

CO-CHAIR GEORGE: Any comments on the evidence?

Are we ready to vote on the evidence?

MEMBER HOLLANDER: It is the only conversation we had that lasted less than like

five minutes.

MS. LUONG: Polling starts now for evidence for measure 1524: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient, and 5 for insufficient evidence with exception.

Evidence passed with 94 percent and 6 percent for moderate.

CO-CHAIR GEORGE: We'll move on to performance gap disparities.

MEMBER HOLLANDER: Oh, the performance gap is huge. I mean the median documentation in the clinical registry in 2012 was 22 percent. There is a long way to go.

CO-CHAIR GEORGE: Any discussion? All right, we will vote on the performance gap.

MS. LUONG: Polling starts now for performance gap: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Performance gap passes with 100 percent.

CO-CHAIR GEORGE: Moving on to priority.

MEMBER HOLLANDER: It is a high priority.

(Laughter.)

CO-CHAIR GEORGE: Any discussion on priority? All right, we will vote.

MS. LUONG: Polling starts now for high priority: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

High priority passes with 100 percent.

CO-CHAIR GEORGE: Okay, we will move on to the specifications and reliability.

MEMBER HOLLANDER: So, I don't remember exactly where I am supposed to say the numerator and denominator but I thought as we got into this, this would be a good spot.

So, the numerator is whether off the PINNACLE Registry they are measuring the individual elements. I think is what Linda was going to say before, somebody smarter than me and Linda, who is also smarter to me, pointed this out and I think it is actually really relevant. It is that they are recording the individual

elements of the CHADS2 score, which is heart failure, hypertension, age greater than 75 but nowhere are they actually recording whether anybody adds that up to calculate the CHADS2 score, which is the basis for the decision-making.

So when you do clinical decision rule validation, you have to actually apply the rule, not the individual elements. And I think that is actually the big flaw -- maybe the only flaw in this measure, as I see it. So as best I could tell, they are just looking at whether the doctors collected the individual information and collected it -- and documented it, but they are not looking at whether they actually ever figured out what is important in that. So is that clear to everybody, the difference I am making?

That being said, when they looked at signal-to-noise ratio, it provided the high discriminatory value that we accept for reliability being okay. And so I think it passes the reliability, but we don't actually know that

anybody is ever thinking about the CHADS2 score on the basis of what is being measured. Unless I am misinterpreting something.

CO-CHAIR GEORGE: Can you address that?

MR. CHIU: So my name is Jensen Chiu. I am from the American College of Cardiology. I think it is a good point you raise.

A lot of elements that make up the measure are on the form, the PINNACLE form. That is a slight limitation that, hypothetically speaking, someone could just check yes, as assessing from the risk factors or not.

But the key point is that a lot of those elements are easy -- already we are jumping ahead to feasibility. A lot of those elements are easy to capture, and they actually do have to capture all of the different elements of the CHADS2 score. So even though they say yes or no, you still have to -- the data quality, you still have to check if they have had all of the other elements of it that make up the thing you

actually -- the physician or whoever is doing the form, needs to actually check those elements. But you are right, it isn't exactly, you know, in terms of the equality -- it isn't exactly a one for one, the ideal to do that. That is true.

But the other thing I would just add is, the challenge is when we eventually move to like CHA2DS2-VASc and other risk scores, the challenge also is balancing the burden, as well as to some degree. Because if you have that in one form, it starts to getting a little unwieldy, because not just CHADS2, CHA2DS2-VASc. There's other HAS-BLED and others. You know, it is kind of balance.

With PINNACLE -- two things I would add. PINNACLE first off, captures both A-fib, CAD, heart failure and hypertension. So, it doesn't just focus on A-fib. And then secondly, while these measures were tested in the PINNACLE Registry, they really could be used in any modality in a physician office. So, that is the other thing I would add.

CO-CHAIR KOTTKE: So you don't actually know that the doc calculated the CHADS2 score?

CO-CHAIR GEORGE: I could see the reverse, if the measure was to document CHADS2 score there could be problems in just asking for that as well.

DR. HO: I guess you could have some sort of consensus where they didn't really know what all the components and said it was a score of three. But then they didn't necessarily have all the risk factors. So, I think it is important to have those components to be able to calculate the score.

MEMBER DELONG: This is Liz. Just a clarification. This is sort of an all or none measure. They get credit for an individual patient if they record that they saw all six. Is that what it is?

MR. CHIU: I would suppose, yes. We didn't call it an all or none composite but in essence, it is like that, yes.

CO-CHAIR GEORGE: Tom?

MEMBER JAMES: I figure this is probably the place to bring this thing up, that this is based on the PINNACLE Registry. Not all cardiologists are involved in that. So, we are talking about a smaller subset. And a large percentage of patients with atrial fibrillation are managed by primary care doctors, who certainly never have access to the PINNACLE Registry. Which means then, we are going to be dealing with -- instead of a national measure, we are dealing with a measure which is a small population of physicians. That is my concern. Whether that affects the validity, I will leave to Liz.

CO-CHAIR KOTTKE: I question that. I mean, aren't we talking about CHADS -- as endorsing CHADS2 as a measure? That doesn't mean it has to go through PINNACLE. It just says this is an appropriate measures for risk of stroke with atrial fibrillation and then we have got to figure out how to get it out into primary care.

MEMBER JAMES: Yes, that is one of the issues that we have dealt with at the AQA, was trying to see how can we take registry-based information and be able to apply it outside of the registry so we have got a much broader scope of physicians? Because I agree with the concept. It is just, are we going to be measuring everybody? Or do we have, in fact -- is that the point that we need to use the measure as a lever to change clinical practice across the country?

CO-CHAIR GEORGE: Linda and then Sana.

MEMBER BRIGGS: So, I think that we all agree that doing the CHADS score is an important step in deciding whether anticoagulation is important for patients, but collecting those six data points by themselves contributes nothing.

So, if your primary care physician fills in the check boxes but he doesn't say that this patient has a risk score of three and I need to put this patient on warfarin, what is the point?

So, I don't think this measure really gets at what we really intend. They need to understand that they need to calculate a CHADS score and what those components are. So, you really need to sees the CHADS score or the CHA2DS2-VASc score, rather than just the components in a registry.

MEMBER AL-KHATIB: I completely agree with that but I also wanted to add one comment that as we continue to use more EMR. I could easily see that being captured, you know, from EMRs. In terms of not just capturing the different data elements, but also whether a documentation of what the CHADS score is or CHA2DS2-VASc score is, in the EMR and then deciding on whether the patient needs to receive an anticoagulant or not.

So, that is why I am not as bothered by the fact that this was done in PINNACLE. I am really hopeful that with EMR, maybe many health systems and places will be able to benefit from this measure. But I do echo the concern that

Linda voiced, and I wonder if the developer can shed some light on that.

MR. CHIU: The thing I would add to both your comments. You know good points about actually having to score. You know, it was somewhat -- in terms of feasibility and burden, there was a decision not to do that, but that's something we could take back.

But one point I would add is that -- I know it is somewhat biased with ACC and AHA, a lot of the practices we do would be EHR. Unlike our inpatient sister industries, if you will, the Cath and all the other heart registries, that issue isn't as a big a issue for us because almost -- I think, PINNACLE, if I remember correctly now. When it first started we had paper charts and EHR -- a lot of it was EHR. I think all of it, almost 100 percent, 99.9 percent EHRs. So, that issue really is, through the uptake, it automatically is -- you can technically know the score by just -- because all of the elements are actually there.

MEMBER AL-KHATIB: The elements are -- take any medical record. If it is an EHR, in particular, will document a patient's sex, their age, whether they have hypertension or not. If they have had an HMP in the EHR, they would have whether they had a stroke or not. Those are like key elements to this.

But whether anybody connects the dots between those pieces needs to be captured. That is what you really want because you want people to say, I need to calculate the CHADS score for this patient and then I need to do something with that data. That is the piece that you are really trying to get to.

MR. CHIU: And if I can just add really quickly. I think that is a great point. I think one thing that we do, separated from obviously the measurements used by many groups focused in discipline, and on PINNACLE, since ACC implements it mainly in PINNACLE at this point. In the form itself, we do, it is kind a work around in that, in two ways.

One, is there is asterisk to let people know what the elements are in it. Just so that they know what it is. I mean most doctors should know it, but we just put it there. A.

And then B. When you go and calculate it, it is kind of linked to 1525, the anticoagulation measure, which I know we are not on that one. But you would hope -- we hope that the doctor would know that those elements are there, then oh, they have got to be given the meds. And if they aren't given the med, they have to give a reason why they are not. I know I am jumping to 1525, but this is kind of linked to 1525 -- both of them are kind of linked together.

CO-CHAIR GEORGE: Mladen?

MEMBER VIDOVICH: In the previous measure we looked at obtaining an EKG, but it assumed that the interpretation is included with an EKG and that actually somebody looked at an EKG. Similar to this is right, you know?

So data are collected but interpretation is not assumed, right, as you

mentioned. So, this is just to corroborate what you said. CHADS2 score data are very simple. They are everywhere. Right? I mean gender, age. I mean so it is not that hard to collect, but it is more than the sum of its parts. I mean that is the great thing about the CHADS2 score.

So, I do feel very strongly that I think there should be some evidence of documentation of interpretation because there is more to it than just collecting the data. It is actually very easy to say in any EHR they looked at the CHADS2 component. Meaning, ten years ago we did look at those components and we didn't know what to make out of it and then CHADS2 came out. Right? And now we --no more. Right?

CO-CHAIR GEORGE: Judd?

MEMBER HOLLANDER: I was going to say this sort of process recommendation to NQF, because I could imagine that the first thing you need to do if you want people to pay attention to it, is get them to collect the data elements. Right?

So let's say, hypothetically, we approve this measure today and it collects the data elements and then we have data three years from now or whenever it is up for renewal that people are doing it right or not doing it right or somewhere in the middle. And someone around the table is going to say well, they did it three years ago, we should keep it going. There should be a way for us to potentially approve this measure today with a note that stays with it that says the goal is, in three years this should be looking at do you actually calculate the CHADS2 score?

Because continuing to just collect these things, we don't think, is terribly useful. And the next committee could choose to use that recommendation or not, but it transmits thinking now. Because I think, listening to comments, there is nobody in this room that would disagree that if you just collect the data and don't use it, it is useful. But we need to a way to transmit it so we could go to the next step

because otherwise, these things just have a life of their own, staying at the low level.

CO-CHAIR GEORGE: Yes, I just feel like it is important to, if you are looking at your A-fib population, you do want to make sure that you have recorded these things.

MEMBER DELONG: I guess I was concerned about the linking with the next measure. The next measure is almost conditional on what is found at this stage, is that correct? Or is that a broader population?

DR. HO: I mean, it is the same population. I mean, you have to assess risk before you find --

MEMBER DELONG: Right.

MS. JOHNSON: If I might, let me go ahead and just address Judd's question, or comment, actually. Those kind of recommendations, we do try to make sure are in the report and next time around, especially as you -- as the Standing Committee, the idea would be that you would have that history and be able

to look at that and hopefully think about that in the future, when you look at these again.

But we don't have any -- it wouldn't be a formal condition, or anything like that, on re-endorsement next time. But it would be a recommendation from the Committee that would go with the report.

MEMBER MITCHELL: At what point in time would you consider this measure and the following measure as paired measures?

MS. JOHNSON: You could consider it today. Basically paired measures can be asked for by the developer. If they ask that they be paired, we can certainly do that. The Committee can ask that measures be paired.

So, the pairing is really more in thinking about how things are reported together. So, basically anybody can ask that something be paired, and if the Committee pretty much agrees with that, in a very informal way, that is fine.

MEMBER MITCHELL: Have you guys thought about pairing these measures?

MR. CHIU: No, we have -- just like many others, like heart failure beta blocker, ACE are paired together, we could definitely consider it. I think just the challenge has been that a lot of the people that we serve, physicians, really like to have those individual -- to see it up, to see individual components.

Because at least for the PINNACLE side, there is like a dashboard, a physician dashboard, as you know. That all the PINNACLE measures, there is like 40 or 45 of them, are basically rolled into the dashboard but all of them we don't actually send to National Quality Forum. But this one and the subsequent one we thought, last time we submitted this, were very important measures to submit forward. That is something we did consider. The difference between paired and composite, we can also explore, too. I know there are differences in that.

But we did want to in A-fib there really only are a handful of measures that

PINNACLE has. This is two of the three. So, I know NQF has moved to just having like one measure on multiple things. But it is something to take under consideration.

CO-CHAIR GEORGE: Any other comments before we vote on reliability?

MEMBER AL-KHATIB: I just have a quick question for you. It says here that this measure was endorsed back in 2012. Is that correct?

MR. CHIU: That was during the last cycle, I believe it was 2012, yes.

MEMBER AL-KHATIB: And so how is this different from the previous measure? Could you --

MR. CHIU: Different than the previous from the --

MEMBER AL-KHATIB: From the one that was endorsed in 2012.

MR. CHIU: So the only thing that is different is the specifications itself. We added the -- actually, no. The assessment, this one, I don't think anything has changed, 1524, but 1525

has changed. The only thing that has changed in 1524 is the title, CHADS2. We originally didn't put the title CHADS2.

MEMBER AL-KHATIB: Okay.

MR. CHIU: So people on the previous steering panel, steering committee, thought that this might have related to something else. But this measure really related to CHADS2, realizing that there is a new title

MEMBER AL-KHATIB: But all the specifications are the same?

MR. CHIU: All the specifications in 1524 are the same. 1525 they are different.

MEMBER HILLEGASS: So for clarification, I may have had a lapse. Can we determine how we would vote on this in the sense of, if we want them to incorporate this, not just individual factors but to actually come out with a CHADS2 score, can we turn this back to the developer while we accept this? Or if we feel it is very important to have a CHADS2 score come out of this, then we have to reject this measure?

I mean, is that on the table? Did I lose something somewhere along the line?

MS. JOHNSON: No, you basically need to think about this measure as it is in front of you. So, this measure is what it is. And that is what you are voting on.

You could certainly, and I think I have already heard a recommendation from the Committee to go further and to maybe, either make an additional measure or something that would actually look at the calculation of the CHADS2 score, and maybe even further than that. But -- so, does that answer your question, enough?

MEMBER HILLEGASS: Well, you sort of did, but so we could send this back, but we could accept this and ask them to develop another?

MS. JOHNSON: Right. But you would only send it back if you feel like it is not conforming to our criteria. So, the basic question before you in reliability is, is this a reliable measure. The recommendation to go

further or do something else would be simply a recommendation.

Now at the end of the day, if you go through all the criteria and you still don't like it, you still have the option of your thumbs up, thumbs down recommendation for endorsement. But that is kind of after you have weighed all the criteria and what you feel like --

MEMBER HILLEGASS: Okay, but your actual title says Assessment of Risk Factors, which they do, but then there is this parenthesis, CHADS2, which they don't.

Correct? Am I reading that correct?

CO-CHAIR KOTTKE: Can I ask a quick question while we are asking these?

MS. JOHNSON: Go ahead.

CO-CHAIR KOTTKE: So tell me again, the reason you did put a calculation of CHADS2 in there, is that the PINNACLE form doesn't have that on it? Is that -- did I misunderstand?

MR. CHIU: The PINNACLE form can certainly add it on there, just at the time we

didn't. We just thought, simply -- I mean, I guess it could be a simple oversight, but the thought is really to simply assess to get to the 1525 that we have actually done the medication in 1525, anticoagulation.

The other issue simply is that there is going to be multiple scores. Assume the CHA2DS2-VASc, HAS-BLED, and so do we actually build in all those scores in there. That is actually kind of a tricky --

CO-CHAIR KOTTKE: I mean, my interest is not the PINNACLE Registry. And I think, I mean probably cardiologists do not such a bad job on this, even though it doesn't look all that great. I mean the problem is most atrial fib is taken care of, probably by primary care and this is a tool. Let's try and get docs to really use CHADS2 and just not have it, somehow, in the electronic record.

Yes -- so, I mean, the answer was in 1524 that all the elements were in the PINNACLE, but not -- except for the calculation.

MR. CHIU: That is correct.

CO-CHAIR KOTTKE: Okay.

MR. CHIU: And also just to add this. The rate of documentation is very poor in this estimate. So anticoagulation, it is also very low, but this is kind of a gap in that that we need to be documenting better. And that is partly why the measure is made, realizing that it is not the best flavor of a measure as opposed to some of the others we put forward.

CO-CHAIR GEORGE: Gerard and then Sana.

MEMBER MARTIN: I was just about ready to say that it seems like 1525 is the action that you are really interested in, that you identify the high-risk patient, that you actually add up the risk elements, and come up with a score that says you need to have anticoagulation and you treat them.

But then as I look back, it isn't completely. 1524 is, probably deals with the that you are not treating some people

unnecessarily. So, you have they really are paired. So, one is you are getting at don't treat someone unnecessarily and the other one, you are trying to treat the person who needs to be treated.

And it could be solved very easily, that one thing, by either truly pairing them or by spelling out that -- in 1524, that you are adding up the elements and coming up with a risk. Which is what you want, rather than just by blind luck having an EHR that tracks those things, which is not the intent.

MEMBER AL-KHATIB: So just to go back to the comment that was made by Ellen. When I look at the validity and the reliability testing that you did, you actually did do those based on the score, not on the individual elements. Am I correct in interpreting those data?

MS. JOHNSON: Perhaps I could just make sure that, from what I am seeing, it looks like that they did what we call performance score testing. So that is, the overall percentage of

the -- this is at the clinician level, facility level -- clinician level. So, when we say performance measure score, we are not talking about the CHADS score. We are talking about the computed score for the clinician.

MEMBER AL-KHATIB: Oh, I see. Okay, got it. Thank you.

CO-CHAIR KOTTKE: So, it's actually worse than it looks. If only 22.8 percent had all of the elements, then the actual calculation was probably far worse.

MEMBER HOLLANDER: I guess the question is, because now I'm confuse. You know there is reliability testing with a good number, but if it is not reliability testing that gets to a CHADS2 score in the document, which is what the measure is sort of calling for, then should it fail that because there was no testing?

And you know and then I am stuck with so if it passes this, it might actually go through and pass every single data point and then half this room might vote against it because they

think it is not a measure that is useful, even though it is a step in the process to get there.

And so, I am a little confused about how to think about it at this point, and I think that is what I hear Ellen saying as well, is do we want to approve something that we really think can't accomplish anything but would just be a step to accomplishing it? And then there is a measure coming afterwards that says we accomplished it or we didn't.

And so what is the added value of this? And I think the, sort of the branding point, like CHADS2 score should clearly be out of the title if it is approved. I don't think we can approve something that says we are measuring the CHADS2 score when we are not. That is just a bad name for it to begin with.

So, there has to be some tweak on it, I think, to go forward from what we are hearing. But I am not really sure how to assess reliability of individual data elements if it is actually the CHAD2 score that we are trying to

get at.

So, I can either say yes, each data element is really reliable where it is documented but the CHADS2 score, there is not testing at all. So, do I vote yes or do I vote no?

CO-CHAIR GEORGE: I have one question I think you can answer relatively quickly but it says in the registry, most missing values are interpreted as no.

MR. CHIU: That just means if you are missing an element, that basically that would fail the measure because you are hoping that someone actually -- basically it is like a meta. If they didn't give it, you have to assume -- if they didn't document it, you are actually failing it.

So, it is not gaining it because it is something missing and somebody didn't put it, arguably, to be gained and then be removed from the denominator. When in fact, now, you have to have all those elements that it is missing or you are actually not treating the patient.

CO-CHAIR GEORGE: -- individual risk factor are those default no, as well?

MR. CHIU: All those risk factors need to be all those risk factors, the EHR or whatever, actually needs to follow all of those.

CO-CHAIR GEORGE: So missing for those doesn't mean no, with a default no.

MR. CHIU: Correct. The other thing I would add, if I could just add about the CHADS2 in the title. Historically, we actually didn't have that in the title but it was actually asked by the previous steering committee to add CHADS2 back in the title. But we could actually remove it. We are open to removing it again.

CO-CHAIR KOTTKE: You could put CHADS2 elements instead, but you know, Judd, this caused a question about what we were voting on in priority.

Was it priority of the measure or priority of treating atrial fibrillation with anticoagulation? I mean, very high if we went back and said well, it is high priority to treat

atrial fibrillation with anticoagulation, but is this measure high priority?

MEMBER HOLLANDER: Well, I think it is a high priority to assess the risk and this is a way in the risk. And the way I interpreted that is, as long as it is anything in the process that gets you towards treatment, we accept process measures along the line. And I suppose it is perfectly fair to have the individual elements as the first process measure but the next process measure would be, did someone add it up and do it?

MEMBER DELONG: Could I get a couple clarifications? Number one, Mary read that the coding, at least in this PINNACLE database, says if it is not captured, it defaults to no. But you are saying that is not correct? If an item -- if a risk factor isn't captured, it defaults to it doesn't exist. Are you saying that statement is incorrect?

MR. CHIU: It defaults to missing.

MEMBER DELONG: It does default to

missing.

MR. CHIU: Right.

MEMBER DELONG: So, wherever that statement came from, that is not correct.

MEMBER HOLLANDER: When I looked at the form, and the instructions for the form, and I did this briefly. I could be totally wrong. I only saw two options, yes and no. I didn't see an option for missing.

MEMBER DELONG: And my other question is totally out of naiveté. Are you saying, Gerard, that this measure is basically a rule out measure, in terms of whether they need anticoagulation?

MEMBER MARTIN: No, what I was saying is that -- and I am not a developer, but I'm just saying as I read the two of them, that the second one, 1525, is identifying the high risk and getting them started on appropriate treatment.

This one is that you just measure the risk. And presumably, you measure the risk, you wouldn't start them on treatment. So to me, it

makes -- they are kind of asking, did you do it? They are complementary but it almost seems that what you are trying to get at is don't treat the patient that doesn't need it. Treat the patient that needs it. And how can you -- and it is a really important measure. So, how could you correct that?

MR. CHIU: Like I said, we can certainly take that idea of a paired I mean the only thing I would say, you know, kind of considering pairing it. The only thing I would say is one of the measures is being used -- I keep saying 1525, because they are both kind of related. 1525 is in PQRS QCDR, 1524 is not. And so if we do pair it, we just have to figure out logistically how we do that. That is something we can definitely take up.

It is easier probably, for us, to do a paired than a composite. A composite is a little trickier conceptually and things of that sort. I see a lot of people not agreeing, so that's -- I know sometimes NQF prefers us to do

composites but when we tried in some measures -- some that you have seen, the Action Registry, there have been a few you have seen. But I think we decided not to do a composite but we can certainly take up a paired issue, take that back up.

No, I'm sorry. So the missing actually is interpreted as no. The options are just yes or no.

MEMBER DELONG: If that is incorrect, if missing is coded as no, then that has a huge impact on the CHADS score. Right?

MR. CHIU: That is correct.

MEMBER AL-KHATIB: But in the implementation of the Registry, at least with other NCDR registries, you could, if you had a performance measure and you determined that certain data elements are so crucial to the measure that you wouldn't accept the submission of the form by the practice, unless they complete those particular data elements. So, there is a way around that.

MR. CHIU: Right.

MEMBER AL-KHATIB: Thank you.

MR. CHIU: So, it is like the data quality program -- PINNACLE's function is a little different than the inpatient setting. But the inpatient setting is a little bit more robust but there is, what they call data quality program. Where there are certain elements that are truly key elements, variables and things of that sort. I would say Cath is probably the most robust. But those basically you would have, in most Caths, a certain amount of threshold. 80 percent or something like that for certain key elements, LEVF, heart failure, things of that sort. You can miss a few times but if you are missing too many of them, you basically don't pass, basically green, if you will. So, they can't even get through getting --

MEMBER AL-KHATIB: No, but the issue right here is capturing elements with the CHADS score.

MR. CHIU: Right.

MEMBER MITCHELL: Jensen, just for clarification. There are singular elements that comprise the thromboembolic risk factors that are yes/no. And then there is a separate data element that says thromboembolic risk factors assessed yes/no. True?

MR. CHIU: Correct.

MEMBER MITCHELL: Okay. And so the question here is it talks about the number of practices missing, the variable thromboembolic risk factors assessed. There are 33 who are missing that information. Would they then be coded no?

I think this is why I am trying to link what Liz was saying to what is written in this code.

MR. CHIU: If I remember, those would probably be coded as -- I see what you are trying to say.

Yes, those would probably then be -- those would be, I guess, ones that, if they were missing a lot of data, we would not be able to

capture.

MEMBER MITCHELL: And so the following question is that, since I left before this part of the project was completed, do you have a data auditing program in place for PINNACLE?

MR. CHIU: That is ongoing. It is not as robust as Cath in an inpatient setting. So, that, unfortunately, we do not have. I don't know what we cited here but that is something that we are working on in the PINNACLE setting.

It is a little bit tricky, not to give PINNACLE an out. But in the EHR environment, it is a little bit tricky to do chart audits, if you will, and things of that sort. That's something we are actively working on. I don't have an exact time when that will be sorted out but it something that we are there is a limitation in the outpatient setting.

And that is not just applicable to this one, but unfortunately, it is applicable to many of our measures that have been endorsed or are still under review right now.

CO-CHAIR GEORGE: George.

MEMBER PHILIPPIDES: So what Liz just mentioned, stirred some thoughts. So, here if data elements were missing, the hypertension, diabetes, and it is coded as no, sort of by default. It would change the CHADS score commission but this measure is not looking at CHADS score. It is not calculating CHADS score. It is only calculating what percentage of the time you gather the data, the parts for the CHADS score.

So, that wouldn't change what -- you guys are basically saying, if you don't document it, we are calculating it as if you didn't do it. And that probably makes sense for this.

However, on the next one, where there is an assumption that we have a CHADS score that is accurate and, based on that, we are going to see how you did as far as treatment. If you use the same methodology there, then we have problems. And I had not actually contemplated that until you mentioned that for this score.

So, and that is the most I can say before my brain is about to explode, I think, on that.

So, I think on this one -- I think there are three parts to treating A-fib correctly. One, is do you gather the data? That is this element. The second is, now that you have got the parts to the car, can you put it together and come up with a risk score? That is not being done here. It is actually not being done in the next one I'll get back to that. The third part is, once the risk score is high enough, are you getting people, like you said, on proper anticoagulation?

The other one does the back end of that. It looks as if a risk score is calculated by the EMR, are they on anticoagulation? But it never assesses whether anybody stopped and said this is the CHADS risk score.

So, we have two ends of the process. What I think is missing is the middle part, the contemplative part, where someone says this is the risk. So, we can either --

MEMBER DELONG: But even --

MEMBER PHILIPPIDES: just let me finish because it is hard for me to think of this.

(Laughter.)

MEMBER PHILIPPIDES: So, we can either say this is not adequate or we can say this is the first good lead-off batter and we will go with this because there is a performance gap. People are not even documenting this stuff. Without documenting this, we can't get to the CHADS score. Those are the choices there, I think.

So, that is my feeling on this measure.

MEMBER DELONG: So, I agree with Sana that there is a workaround for the first one. But for the first one, if it is missing, it is not coded as missing. It is coded as no. So, in the overall assessment, it is coded. So it gets credit for being there when it is not.

MEMBER PHILIPPIDES: Correct. So,

missing data is coded as --

MEMBER DELONG: As no, and present.

MEMBER PHILIPPIDES: So then, this measure doesn't do anything based on the form. Right? Because all you want to know is whether it is assessed and there is no way to say it is not assessed. It is present or absent and it is assumed to be assessed. So, it would not be valid then. Right? Is that what we are saying?

MEMBER DELONG: But they are finding some that aren't there. So, I don't know how that happened.

MR. CHIU: If I could just add. I would say that 1525 is -- the denominator in 1525 is the patients in 1524 that you have assessed. A lot of people that are missing, obviously, those individuals are missing.

But 1525, that measure is measuring anticoagulation medication given. If something is blank there, it is actually failed.

MEMBER DELONG: I think where we are confused is how do you come up with a not

measured, if every one of the six is either a yes or no, and if it is missing, it gets coded as node?

MR. CHIU: So the assessment score -- if I could just answer really quickly. If the assessment score is a yes, and thromboembolic risk factors is yes, that is the denominator of the anticoagulation.

If somebody decides not to give a med, any of the oral anticoagulants, dabigatran, warfarin, all those others. If that is blank, that is actually performance fails. Because in measures --

MEMBER DELONG: Now, you are talking about 1525.

MR. CHIU: That is what I am saying. So, that one -- I think you alluded to, but that second measure is different than 1524. It actually is. 1525 is different than 1524. If you are leaving that blank, that is actually performance failed. Because the point of the second one is giving the medication.

MEMBER DELONG: I am talking about the first one.

MEMBER HOLLANDER: So, I am going to read right from the PINNACLE form and we are going to, I think, solve this problem, which is the unfortunate answer.

So diabetes, in an example, the coding instructions are: Indicate if the patient has a history of diabetes, regardless of duration or disease or a need for anti-diabetic agents, and the selection choices are only no or yes. Which means you can't say they didn't document it, which means that this measure can't document whether information is missing.

(Simultaneous speaking.)

MEMBER HOLLANDER: But nowhere in the instructions does it say, you can leave it blank. So, if I am a coder -- I am looking at the Coder Data Dictionary right now. And that is what the Coder Data Dictionary says. It doesn't say I have a choice to leave it missing. So, some people may be doing that but that is not in the

directions in how to code it.

CO-CHAIR GEORGE: I think some of this conversation pertains to how it is implemented in a system, which we don't necessarily have to be so concerned about. We can just provide that information so that maybe clarification can be updated.

MEMBER DELONG: But the report they presented is based on that database also.

MR. CHIU: The testing, that is correct.

CO-CHAIR KOTTKE: And this may have huge implications of not having or not recorded. When we were doing this for a history of myocardial infarction -- this was a hell of a long time ago when I was a fellow, but case fatality rate for yes, prior history, was 12 percent. The case fatality rate for no was 12 percent. The case fatality rate for not recorded was 18 percent.

You know, you just can't record it as not no.

MEMBER HOLLANDER: I now don't know what I am interpreting, which is actually back to at least the reliability.

CO-CHAIR GEORGE: So yes, we are still talking about reliability.

MEMBER HOLLANDER: So, I really just have no idea how to interpret a no answer at this point, which makes it really hard to have some confidence in the reliability of the data.

I mean I suppose if there was something up-front that said, if it is not yes or no, leave it blank. That would be great, but that is not provided in the data dictionary, or at least the portion of it that we have access to.

CO-CHAIR GEORGE: And I agree but we are not telling them how to implement PINNACLE.

MEMBER HOLLANDER: No, but --

CO-CHAIR GEORGE: We are telling them this could be a potential issue, depending on how you have it coded and implemented in your system.

MEMBER HOLLANDER: Well, but it is

our job to say whether we believe the measurements and I am saying that if I don't see direction to say leave it missing, if it not addressed. I don't have any way that I could say this is reliable or valid because the instructions I see force you into a yes or no answer and both a yes and a no answer means they addressed it.

So, I don't see a way to document it is not addressed.

MEMBER MITCHELL: So if you look, in that same document that you were referring to, there is a measure called thromboembolic missed factors assessed. And the actual definition of this is indicate if the patient's thromboembolic risk factors for AF or AF flutter were assessed and documented in the chart. And then the answers are yes with, in parentheses, all risk factors were assessed. Two, no. Medical reason, two, no. Patient reason, four, no. System reason.

So, the question is, this data

element, in and of itself, just about answers his question here about whether or not thromboembolic risk factors were assessed because it specifically asks the question yes/no.

What we are talking about right now is each individual measure, whether or not age -- all the different, six different, risk factors were individually documented. There is a failsafe in quotation marks, in the data set itself, that specifically asks the question about all six risk factors.

This is separate, a completely separate data element. It is not derived -- it is an absolute data element that you have to fill out. At least that is how I built it.

CO-CHAIR KOTTKE: So, I now found that but there is no instructions on -- it says all risk factors. It doesn't say all of CHADS2 risk factors. It is whatever somebody thinks is there. We don't have reliability testing. We don't have reliability testing as far as I know for that, as a matter of fact, I couldn't even

find that measure on the case report form but I find it in the data dictionary.

CO-CHAIR GEORGE: So, we do have the signal-to-noise.

MR. CHIU: We have that -- at least in the PINNACLE data collection form, on the bottom, we do note that the six factors have to be the ones you are using. Obviously, we agree the limitation is, the score is not, we don't actually have a score at this point.

MEMBER BRIGGS: So, I just have one comment on that. This measure was approved in 2012 and you have been collecting data on that. And we are talking about this measure being a process or a baby step. If this was already approved in 2012 and we still aren't calculating a CHADS score, isn't the next thing that we should be doing actually the CHADS score? So, that is what we should be doing now, rather than saying let's continue this measure the way it is.

MS. HIBAY: Do you mind if I provide a little bit of clarity for the missing data? I

think it is actually quite easy, just to remember, we are going to focus on the measure here that is in front of us as we continue to talk.

So S22, missing data. This is what the developer provides: If data required to determine if an individual patient should be included in a specific performance measure based upon defined criteria is missing, those cases would be ineligible for inclusion in the denominator and, therefore, the case would be deleted. So, you don't have all the data elements related to an encounter or the patient population. Right? Okay.

If data required to determine if the denominator-eligible patient -- so now you know your population, qualifies for the numerator. The numerator here is do you have six risk things documented.

So again, if denominator-eligible patients qualify for the numerator, or has a valid exclusion or exception -- if it is missing,

these cases would represent a quality failure. That is an answer to your question. they not thrown out of it.

MEMBER DELONG: Sorry. I think Sana made a great point. There is a workaround here. But what you are saying is it has to show up in the file as missing and that is what we are having a problem with, because it doesn't show up as missing in the coded file. It defaults to zero if it -- if it is not filled in, it defaults to zero.

But as Sana said, there is a workaround. They can fix that. And as Kristi said, if we accept that all of the CHADS measures, all six of them, are incorporated in that summary score then that fixes it. But missing doesn't show up and that is the problem.

MS. JOHNSON: So, I wonder. I'm confused now, unfortunately. I think I hear that there are six data elements that go into a CHADS score and there is boxes that you would check or not, depending on your patient. And then there

is a seventh box that says did you or did you not assess it. So, my question for the developer is are you basing this measure on that seventh box?

MR. CHIU: 1524, yes, we are basing it on that.

MS. JOHNSON: So, on that one box that is yes or no -- or actually it is not yes or no. There is a few other things. Okay.

MR. CHIU: It is basing it on that and the six -- all those seven, but again, it doesn't actually have a score itself, the exact score. It is basing all the seven -- so all seven, if something is missing, there is still that issue as you are talking about, but it is going on that yes/no, for environmental risk factors, and all those elements need to be there.

So, the stroke, TIA, diabetes, and all those other things.

MS. HIBAY: So just a clarifying question. So if any of those seven boxes, six being the risk factors, the other one being the yes, they are all there. If any of those are

blank, it is a quality failure.

MR. CHIU: Although we will check in our notes, I am pretty certain that if something is blank it is a failure because the thought is that they need to be documenting the measure -- I mean documenting the risk factors. The point is so low, people aren't documenting this measure. So, if they are not documenting, you can assume that they are not doing their job.

MS. JOHNSON: Right, Nursing 101. If it is not documented, it is not done. I'm sure Medicine 101 as well.

MEMBER HILLEGASS: Okay, so I'm confused because on 2(b)(7) under Missing Data, it states here the developer notes that in the PINNACLE registry, most missing values are interpreted as no.

So, I hear what you said, Sharon, but that contradicts what you are saying, I think.

CO-CHAIR GEORGE: Well again, I think it depends on how that is implemented. They say most, I'm not sure what most means, but you can

implement it as a checkbox where one has to be checked. If they are both unchecked, it is missing. Or you can implement it as a radio button, where it has a default and you have to change it. So, it depends how they implemented that in the coding that -- how they built their system.

MEMBER HILLEGASS: But the developer is reporting this.

MR. CHIU: Which section are you referring to?

MEMBER HILLEGASS: Under 2(b)(7), page 6, on 1524 under Missing Data. Page 6, 2(b)(7). It is called Missing Data.

MR. CHIU: Yes, I actually think that is inaccurate, the description, the 2.2(b)(7), because there is another section in S -- that section you guys pulled up earlier, S-33. Sharon, you pulled up something in S.

MS. HIBAY: I did, it was S-22, but if you go to page 39 on the full, it says in PINNACLE, missing values are interpreted as no.

So, you are correct that there is conflicting information here.

So, you have the S-22, which says they are a quality failure if they are blank. And you have in 2(b)(7)(1), which is on page 39, it says in PINNACLE missing values are interpreted as no for most variables.

MR. CHIU: So, I think that we will have to take this testing back just to see why there is incongruency here. Just when we submitted this last year -- this was submitted a year ago, or I actually didn't submit this document. So, we will have to take this back in terms of the testing.

But the intent of the measure is that you have assessment and all those variables need to be present. If they are missing, that is actually performance failure because you need to be actually having those things noted, realizing we don't have the score. That is a limitation.

CO-CHAIR GEORGE: Are there any new issues to bring up in regards to reliability?

MEMBER HOLLANDER: It is sort of a new issue. I don't know what I would be voting on, and I am being totally serious. I actually think that sort of I throw it out there for a proposal is whether we should defer this one, too, from this point forward.

If there is no missing data, the whole measure can't go through. It is totally invalid. It is fatally flawed. And I don't know whether that is the case. And so, I can't vote because nobody could tell me how this stuff is coded and I see that there is a good noise to signal -- signal-to-noise ratio but now I don't know of what.

And so if there is a great signal-to-noise ratio of meaningless stuff, it doesn't help me.

CO-CHAIR GEORGE: So, I think we do need to vote on what we have before us today. I think the developer had one last comment.

DR. HO: I mean I guess the question here is about reliability of the measure

regardless of where it is implemented. So, I guess, isn't the question whether this can be assessed in any healthcare system, not just whether it is in PINNACLE or not? Is that correct?

CO-CHAIR GEORGE: In a reliable basis that you are assessing.

DR. HO: Right.

MEMBER VIDOVICH: If I understand you, Mary, it doesn't have to be done in PINNACLE. It can be done at Department of Veterans Affairs, whatever, anything. Right?

CO-CHAIR GEORGE: Well, if there is no new issues to raise, I think we will go ahead and vote on reliability.

MS. LUONG: Polling for reliability starts now for measure 1524; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

For reliability testing, it did not pass; 17 percent voted moderate, 28 voted low, and 56 voted for insufficient.

CO-CHAIR GEORGE: Okay, we'll move on

to validity.

MS. JOHNSON: No, actually since it did not pass reliability and this is a must-pass sub-criterion, we are going to stop discussion right now.

I think just since the developers are going to be here tomorrow, I think it might behoove us to let them look at their specs again, think about it. And if they can explain things a little bit better tomorrow, we can see if that might help.

I think part of it, at least in my mind, is I am not quite sure how this measure is being calculated. I still don't quite know that. I think that might be part of the concern. It is not the testing results. It is the precision of the specs and understanding how the specs work. That is the part of reliability that people are hanging up on.

MR. CHIU: Also, the reason we would like to back is there are multiple areas in the form where I see there is inconsistencies. So,

2(b)(7) you brought up and then earlier in S, page 6 there are some inconsistencies. So, we will take a look at that to see why there is inconsistency in how it is tested.

MS. HIBAY: Does that sound like a reasonable plan to the committee? Okay. Especially since the earlier voting was so stellar.

CO-CHAIR KOTTKE: Okay, 1525. Do we want to move forward or take a break is the question. We are 15 minutes past the break. We could take a 15-minute break right now and then come back. What's that?

(Simultaneous speaking.)

CO-CHAIR KOTTKE: Okay, the consensus is do 25. We will hear from -- is Joel on the phone? No. So, it is George and Mladen. The developers here can give us a quick -- which they have probably already done.

DR. HO: So, on 1525 is appropriate anticoagulation in patients at moderate to high risk for thromboembolic events and those with

atrial fibrillation. So, I mean we have alluded to this measure in the earlier discussion. And as you can see from the data that was provided, there is quite a bit of variability in terms of patients being on anticoagulation in those four moderate to high risk.

CO-CHAIR KOTTKE: Who is discussing this? Were you making motions there, George?

MEMBER PHILIPPIDES: I will start. So, this is, as we discussed, sort of the next step in sort of the A-fib treatment algorithm. This is looking at people with a documented, by electronic medical record a documented high CHADS score of one or above, people who should be on anticoagulation in the world of nonvalvular AF.

As far as evidence, as heard before, there were a lot of clinical trials citing the importance of anticoagulating high-risk A-fib patients. They relied heavily on the ACCF and AHA guidelines and ACCP guidelines. And I think the evidence is strong for this concept.

We then looked at, again, PINNACLE

Registry data from 2012 and I think one other year to look for opportunities for improvement and performance gaps. And that was cited, they estimated that roughly 59 percent -- pardon me?

CO-CHAIR KOTTKE: We're going to vote on evidence first.

MEMBER PHILIPPIDES: There was strong evidence.

CO-CHAIR KOTTKE: Okay. Anybody want to pile on? Okay, let's vote on evidence.

MS. LUONG: Evidence starts now for Measure 1525; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient, and 5 for insufficient evidence with exception.

Evidence passes with 89 percent for high, 11 percent for moderate.

CO-CHAIR KOTTKE: It is so impressive it gets music.

(Laughter.)

CO-CHAIR KOTTKE: Okay, George, opportunity for improvement and disparities.

MEMBER PHILIPPIDES: So, I will go

quickly. So, again, they cited PINNACLE Registry data from 2012. The mean performance was about 59 percent of these high-risk patients were appropriately anticoagulated but there is a wide-range of zero to like 99 percent with an inter-cohort score of about 22 percent. So, there was a gap and there was a range.

Interesting, there also was some data, not a lot but some data suggested that there were disparities of care in different groups. Men were treated more avidly than women. Non-African Americans less likely to be treated and patients with non-private insurance. So, there was a suggestion there of some disparities that we should all sort of know about.

CO-CHAIR KOTTKE: Are we ready to vote? It sounds like it.

MS. LUONG: Polling starts now for performance gap: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

One hundred percent for performance gap.

CO-CHAIR KOTTKE: Priority?

MEMBER PHILIPPIDES: High prevalence, high morbidity, high mortality, high cost, high priority. MS. LUONG: Polling starts now for high priority: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

One hundred percent for high priority.

CO-CHAIR KOTTKE: Scientific acceptability, numerator/denominator exclusions, beta source issues/concerns.

MEMBER PHILIPPIDES: So, I am not going to over numerator/denominator again. They are pretty straightforward I think. It is basically numerator people on Coumadin. Denominator, those over 18 with nonvalvular AF or A-flutter. That is pretty straightforward.

I think the exclusions warrant a little bit of discussion. I would like to hear what Mladen has to say about this. The medical exclusions, I think, were fine. Transient or reversible causes of pneumonia, surgery, pregnancy, I think that is right. There is no

rush to treat with anticoagulation, we agree.

On the appropriate medical exclusions, bleeding, allergy, absolutely. But I always get concerned when I see patient reasons in sort of italics that include economic, social, religious, noncompliance, and patient refusal and especially patient refusal as being an appropriate outcome always bothers me because one could make an argument, I think we have made this before, in the area of lipid management, that a system or a physician who takes time to educate a patient and goes over the risks and benefits might have less patient refusal.

And put another way, one way to game the system and not anticoagulate those who should be is to write the patient refused, when it really could be that there wasn't as much time and effort put into educating that patient.

They do give some data on this. I don't know if now is the time to get into that. Sixty percent of the physicians two years in a row in the Registry had no exclusions but of

those physicians who had exclusions, there is a wide range. And there were some physicians who used exclusions a lot and, in those physicians, I think it was 87 percent of them were patient exclusions, not medical exclusions.

So, again, it doesn't seem like there are many physicians who are using the exclusions but of those that do, most of those, the vast majority are on the patient side and again, that always worries me because I think that is sort of hard to document what is really going on there. So, that could be a source of dirtying up the data.

Mladen, any thoughts on that?

MEMBER VIDOVICH: That is always a problem. We run into this a lot with STEMI, with patient inclusion and why was I late with door to balloon time. And you always try to come up with some things to game the system. I don't know how to eliminate that. Maybe people have a good idea. It has to be there somehow because people do refuse anticoagulation frequently. They don't

have money or they don't have transportation. You know how it is in practice.

So, I believe it, if it were up to me.

MEMBER AL-KHATIB: I want to second that as well. As a practicing electrophysiologist, I see a lot of patients with atrial fibrillation. We have these discussions about the benefits and risks of anticoagulants. And I can't tell how many times, after going through all of those and trying to convince the patient that should be on one, their answer to me is no, they we refuse. And they consistently refuse. We absolutely have to have that there. We don't have control over it.

MEMBER VIDOVICH: It is more common than not. You know how it is.

MR. CHIU: Can I add just really quickly just to their points, really fast? That is a good point, I think that he is bringing up. Some of our sets in terms of patient refusal, actually, that would be considered a performance not met, like cardiac rehab and like referrals

and things of that sort. Those are recently updated, those sets.

But like for medication we might be just a little bit more leery. I see your point about it being other the flip is like over-medication. And somebody simply doesn't want it, it is a med, there are some unintended consequences. So, that is why currently as it stands it is '08 and 2011 when A-fib gets updated, that issue will be discussed further about patient refusal. We still consider that an exception to the thing realizing you are going to have a few people that are arguably gaming the system. I can argue almost any measure can be gamed to some degree.

But that is something that will be taken up when we visit CHA2DS2-VASc and all the other things for the next A-fib update.

CO-CHAIR KOTTKE: Okay. Yes. Oh, sorry, Tom.

MEMBER JAMES: This kind of issue was discussed at one of the AQA meetings. And the

consensus from that particular meeting, ACC was involved, was that patient refusal is appropriate, should be included as an exclusion but it should be monitored and tracked by a doctor for public reporting purposes.

MR. CHIU: We do track the exception rates of this. Hence, we have broken them into the buckets but those are tracked. Or at least we recommend they are tracked if people use our measure and other ones.

CO-CHAIR KOTTKE: Linda?

MEMBER BRIGGS: I just wanted to speak to the denominator statement because I think it is not as clear as we would like it to be because it talks about one or more high-risk factors or more than one moderate risk factors.

For CHADS2, there are no like high or moderate. It is in CHA2DS2-VASc, there are those additional things that we added. The high ones are the stroke, TIA, other thromboembolism. So, I am just saying that there is a lack of clarity there that maybe you want to clarify the

denominator statement a little better, maybe as a particular part number of risk factors or what.

Right now it says high or moderate. And I know what CHADS is and I had to go look that up because I had no idea where I would find that. And it is really when you look at CHA2DS2-VASc versus CHADS.

MR. CHIU: What page are you referring to? The reason I bring that up is one of our sections is --

MEMBER BRIGGS: Oh, I'm sorry. Actually, it is on page two, where it has the numerator and denominator statement and it is actually in the beginning of the specification as well.

MR. CHIU: Okay. So, it must have been cut off. So, it is an error, an oversight in our writing this but the denominator actually is probably not even here but we actually do specify it in our published document what moderate is and what high-risk is delineated. So, the high-risk being the prior stroke and the

moderate being the other.

MEMBER BRIGGS: It is delineated elsewhere. Okay.

MR. CHIU: So, I apologize it wasn't written here.

(Simultaneous speaking.)

MR. CHIU: Yes, 2006 guidelines.

MEMBER BRIGGS: Okay.

MR. CHIU: We actually cited directly to the guidelines and write it down. It is like a table that I guess I think we maybe tried to embed and we weren't able to get the table in there. So, oversight on our part.

MEMBER PHILIPPIDES: So, they performed reliability testing based on the PINNACLE data and the ICD-9 codes. I think it was a signal to noise ratio at the level of performance measure and it had a very high score of like 0.99.

So, it seems like for reliability testing, it was reliable. And should we vote on reliability?

CO-CHAIR KOTTKE: Yes, let's vote on reliability.

MS. LUONG: Voting for reliability starts now; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

So for, reliability we have 47 for high; 47 for moderate; and six percent for low. So, it passes.

CO-CHAIR KOTTKE: Validity and threats to?

MEMBER PHILIPPIDES: So, in my opinion, the specifications outline align with the evidence. I could not find any formal validity testing. They basically leaned on face validity. They cited the fact that they polled certain ACC and AHA committee members. And there was a high number of those guys and gals who felt that this was valid and that it basically outlined good quality.

We talked about the threats to validity in our discussion about the exceptions and exclusions. So, I think we sort of went over

that and feel okay with that. And as I mentioned before, the data support the idea that there were meaningful differences in regards to performance across a pretty large registry. So, overall, I had not major problems with validity, as outlined.

CO-CHAIR KOTTKE: Anybody have the urge? Seeing none, let's vote on validity.

MS. LUONG: Voting for validity starts now; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Validity passes with 18 percent for high and 82 percent for moderate.

CO-CHAIR KOTTKE: Feasibility.

MEMBER PHILIPPIDES: So, this again was a review of mostly electronic medical records. I believe that is correct, guys.

And they mentioned that the people of value in the Registry in the review didn't have quote, unquote, any major problems or ask to changing anything. I had a hard time if there problems in getting specific pieces of the data.

I might not have looked in the right place. I looked in the appendix as well.

So, I couldn't find exactly what percentage of the time they missed certain data elements, so just the broad statement that there was no major issues in extracting the data. And similarly there was no mention of sort of cost of extraction or time of extraction that I could find.

And I don't know if the developers want to comment on that.

MR. CHIU: I'll comment on the cost first. We don't really have a hard and fast rule in terms of the cost. And I will say the PINNACLE, unlike all the other registries, one advantage it has is that it is actually free to physician offices, so there is no cost associated. Obviously, the real cost is the nurse or somebody actually pulling the data. That is the real cost there. We didn't actually quantify the time it would take for this specific measure.

The thought, obviously it is biased on our part, we thought that this measure was pretty much easy to pull because a lot of this stuff you already pretty much have, the diagnoses and things of that sort. It should be easy to pull, like Cerner and Epic and all those others. So, that is kind of what we think in terms of the cost.

And then your first point, in terms of the missing data we will probably have to get back to you. That relates to the earlier question about assessment on the risk factors. So we can get back on specific numbers. I don't think there is a lot of missing data but we can certainly check, at least on the few key elements needed for this measure and get to your point, I think, Dr. DeLong about what we do with the missing data because there is a little incongruence in the form.

MEMBER PHILIPPIDES: I think in regard to the missing data, in this metric that makes sense that if somebody doesn't comment on whether

or not they anticoagulated or not, then that is a no. This one it fits with the classic nursing or doctor rule, that if you didn't document it, you didn't do it.

So, my suspicion is that 2(b)(7) pertains to this metric and not to the last one on an individual element level. I am just suspicious that that is what you are going to find.

MR. CHIU: Yes, we are going to confirm it. This one, we are pretty certain that if you are not document, you are failing. So, the point is you need to be giving the anticoagulant.

MEMBER PHILIPPIDES: So, those are my comments in regards to feasibility.

CO-CHAIR KOTTKE: Seeing no movement, let's vote on feasibility.

MS. LUONG: Polling for feasibility starts now; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Feasibility passes with 29 percent for

high and 71 percent for moderate.

CO-CHAIR KOTTKE: Usability and use.

MEMBER PHILIPPIDES: So in regards to usability, at present, this measure is not being publicly reported. It is being used in PINNACLE, I guess in some ACC practice improvement pathways. It is sort of percolating on that level.

There is mention that this might be picked up in CMS in the future and it is sort of a more robust for maybe part of the PQRS complex of metrics. But that is sort of where it is. It is a little bit early on, I guess, in its mention.

MR. CHIU: When we wrote this, this was actually -- just to give you a little history, this was actually written I think in December 2013. And it was supposed to be reviewed last time and it wasn't. So, some of this are the factors of it being old.

So, hence, this is actually, one thing we can say, I think we wrote this as we weren't

sure if it was. Now, we are certain it is in PQRS. This measure, 1524 is a piece of this.

MEMBER PHILIPPIDES: Okay, so I should have read the more PQRS. It's in there. Thank you.

And overall, I have no major issues with usability.

CO-CHAIR KOTTKE: Unintended consequences?

MEMBER PHILIPPIDES: I don't have any.

CO-CHAIR KOTTKE: Let's vote on usability and use.

MS. LUONG: Polling starts now for usability and use: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient information.

Usability and use criteria passes with 41 percent for high and 59 percent for moderate.

CO-CHAIR KOTTKE: We will vote on the overall.

MS. LUONG: For overall suitability for endorsement, polling starts now; 1 for yes and 2 for no.

Measure 1525 passes for overall suitability for endorsement, with 100 percent.

CO-CHAIR KOTTKE: Any issues of competing measures, George?

MEMBER PHILIPPIDES: There are two competing metrics out there but both of those have to do with patients who have already suffered a stroke ad they are both, I believe in-patient to traditional care medicine out-patient based. So, it really is a slightly different population. This is primary prevention trying to stop the first row. So, I don't think there is a major issue.

CO-CHAIR KOTTKE: Than you. It is break time. Fifteen minutes.

(Whereupon, the above-entitled matter went off the record at 3:21 p.m. and resumed at 3:33 p.m.)

CO-CHAIR GEORGE: This is Measure 2461. Our developers are here. I just will let everyone know that Sana has recused herself from this measure. And the discussants are Carol,

Joseph and Tom James. So, developers, if you will introduce the measure.

DR. KUSUMOTO: Great. Thank you very much for this opportunity. My name is Fred Kusumoto. I am from the Mayo Clinic. I am an electrophysiologist here representing Heart Rhythm Society. So, thank you again. I appreciate it and do want to acknowledge that Sana is one of the principal developers for this and actually recording some of her work here.

So, this first data measure, 2461, really looks and gets at this sort of first in-person evaluation of someone with a new implantable device. Remember that a lot of people are implanted with devices, 200,000 to 300,000 people and we are really talking about new devices. And the reason we want to focus on new devices is once these devices get implanted, it is important to realize they then shift their environment of care to an outpatient situation where the monitoring then and the care of that

patient then changes over time because a device is something that can be prescribable or changed. In other words, this isn't an inert thing. It is not like getting a hip or a prosthesis of some kind. This really is a new tool that then can be used and taken care of.

So, a couple things. So, let's take a look at the evidence base, briefly, before you guys speak about it. Remember that the in-person evaluation has been around since the 1990s. It has been endorsed not only by Hearth Rhythm Society but also by the Canadian Society of Cardiology. The reason for this is because this first visit is critical first of all for coordinating care but also because the great majority of complications occur during this first month anywhere from two to five to six percent of lead issues where in fact lead problems get recognized.

This first visit is also very critical and unfortunately, there is a big gap for care. Sana has done wonderful work looking at the

Medicare sample where there has been 40,000 patients with devices in place where, in fact, only 40 percent of those patients actually received the appropriate guideline-directed follow-up.

In addition, there is significant consequences to not having this appropriate follow-up for the same group in a group of patients with ICDs. So, these are devices then to defibrillate patients when they have sudden cardiac death or ventricular arrhythmias. In a group of 70,000, in fact, again, 40 percent of patients did not get appropriate follow-up and those patients, when you looked at one-year, two-year, three-year mortality had a seven percent increase in mortality.

This is not just in isolation. If you look in the Canadian database, again, 10,000 patients from Ontario, in fact who had ICDs, they have much better follow-up, 86 percent of those patients, in fact, got appropriate follow-up within this time period. But in fact for those

patients who did not get follow-up, there is a 30 percent increase in mortality in that group.

So, critical, a gap, a lot of variability and very impactful.

CO-CHAIR GEORGE: Thank you. And who is -- Carol?

MEMBER ALLRED: Thank you. Now, can you hear me? Great. Thank you for that wonderful introduction. You really set the stage.

Okay, can you hear me now? All right, very good.

I think that was a great introduction and set the stage well for the measure. I was particularly struck by the first follow-up visit being so important because of the complications and because of the education of the patient and their family. It is a wonderful time for them to finally get their act together and ask questions. This is a personal experience. I do have an ICD implant and between the time of implant and that first visit, lots and lots of input from people

with things that some were true, some were not true, but everyone had an opinion about what was going on and what I needed to know. So, it was good.

I would agree 100 percent with the numerator statement here. I would disagree with you on the denominator statement in that I think the person with a repeat procedure still needs that follow-up visit because the chances for infection, the chances for lead movement, the chances for any number of things, and sometimes just the reprogramming. I had a different device put in. I had extra leads put in. I needed that same follow-up that everybody else did. So, I would include those people.

CO-CHAIR GEORGE: Can you speak to the evidence right now?

MEMBER ALLRED: The evidence really doesn't address the follow-up person with a repeat procedure. It is one of the exclusions in there.

The evidence itself is one study with

a fairly small number. I thought could have been a little bit better evidence but it was very good and I thought what was there was accurate.

DR. KUSUMOTO: Great, thank you. And let's bring up, when we talk about the denominator statement because your point is incredibly well taken with regard to thinking about this problem, about this difficulty because those patients who have devices in place are an incredibly important group. The reason, just as a quick aside, just to acknowledge this important point, is that it suddenly makes our group very variable. And so that it makes it somewhat more difficult.

MEMBER ALLRED: Absolutely.

CO-CHAIR GEORGE: Tom, do you have a comment on the evidence?

MEMBER JAMES: Yes, and there is nothing like meeting with the developer to start changing some of my thoughts but much of it has to do with definitions. Because when I started looking at the evidence and started thinking in

terms of, from the title, in-person evaluation and then looking at the EKOS study -- you didn't mention that one.

DR. KUSUMOTO: I didn't. I was going to let you do it.

MEMBER JAMES: Okay, this is where we had that discussion. But the part where the EKOS study looked at ambulatory measurement from remote monitoring, as Fred was pointing out, this was subsequent to an in-person evaluation.

I had gone to the Heart Rhythm Society and the European version to look at what evidence they had. Again, so much of this was definitional, I was convinced from speaking that if I changed the definitions to something that a general internist like me could understand, then I think we have got good solid evidence behind this one. So, I would be supportive on the evidence level.

CO-CHAIR GEORGE: Tom?

CO-CHAIR KOTTKE: So, you have just confused me. What is the problem you have with

the definition?

MEMBER JAMES: The definition, it says in-person and yet it looked like from some of the studies, it was done by remote monitoring. But the point was that that study started after the in-person visit. And secondarily, it is the definition of who is an appropriately trained clinician. Because I thought well, heck, I see people who have had a pacemaker in my office. I can't do anything about it but I know they are there.

So, if we have the clarification on the training level, who is the appropriate one, and then finally what is missing from the evidence what actually happens during that meeting, that in-person evaluation. And that is better defined in this document than it is in the measure.

And I think just adding those as definitional elements will satisfy this.

CO-CHAIR GEORGE: Any other comments on the evidence? Ellen?

MEMBER HILLEGASS: The only comment I had was I was concerned about the two- to twelve-week window. I thought twelve weeks was too long but I know you base your statement on the evidence and the evidence did say two to twelve weeks. But I still believe that is too long.

DR. KUSUMOTO: I can't help but agree with you. So, our practice at the Mayo Clinic is ten days, seven to ten days. And so, there is no question I absolutely agree with you.

Having said that, the evidence which then goes with the consensus statement is two to twelve weeks. And because of that, that is where our evidence is, that actually Sana's study has looked at, as have others, the Ontario database and also Hess et al. when they looked at the NCDR database and really called that out.

So, when you start to look at evidence with large numbers greater than 100,000 patients, it really, sadly, is two to twelve weeks, even with our personal issues associated with that.

MEMBER HILLEGASS: Yes, I wish there was evidence to show a shorter amount of time because I really think you need that.

MEMBER ALLRED: I agree with that, too, because if you are going to have an infection, you need to catch it sooner, rather than later.

DR. KUSUMOTO: And that is the one thing I wanted to emphasize in this measure. It is really a care coordination measure. It really is making sure that we transition and put responsibility on the implanting physician that hey, if you are going to implant it, you are responsible for then making sure that this patient sees someone at some period of time.

MEMBER ALLRED: I agree.

CO-CHAIR GEORGE: If there are no other comments, we will vote on the evidence.

MS. LUONG: Polling for evidence testing starts now. I'm sorry, polling for evidence starts now: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient, and 5 for

insufficient evidence with exception. And this is for measure 2461.

For this measure, 38 percent voted high and 63 voted for moderate.

CO-CHAIR GEORGE: Okay, we will move on to performance gap and disparities.

MEMBER ALLRED: Okay, performance gap. There is definitely a performance gap. I think only 42 percent of patients actually received that first follow-up visit during that period. So, that is an opportunity to improve.

There is some disparities information that indicates that white Anglo-Saxon people have a higher incidence of having that follow-up visit than minorities do. So, that is an important area. We need to address that disparities gap.

MEMBER JAMES: I would corroborate that from our own data within a Medicaid plan.

CO-CHAIR GEORGE: Any other comments or discussion? If not, we will vote.

MS. LUONG: Polling for performance gap starts now; 1 for high, 2 for moderate, 3

for low, and 4 for insufficient.

For performance gap, it passes with 81 percent for high and 19 percent for moderate.

CO-CHAIR GEORGE: And priority.

MEMBER ALLRED: It is a high-priority item. It is a growing area because there are like 55 percent more people with implants in the last ten years. So, currently, there are 2.9 million people in the United States with implants.

I think it is an expensive area and, obviously, patients without good care aren't going to do well.

CO-CHAIR GEORGE: Comments?

MEMBER DELONG: I have a question out of total ignorance. Are these procedures all sort of equivalent in their need for follow-up and their risks?

DR. KUSUMOTO: So, I can answer that with regards to the sort of clinical need. You know certainly we have sort of one-lead, two-lead, and three-lead models of devices. In

fact, this first follow-up, although you are going to recognize more lead issues, obviously if you have more, the more leads you have, the more likely that you are going to run into issues. As Ms. Allred pointed out, the big issue here is this change in care environment. So, I would make the argument that even though there is some complexity associated with it, you are more likely than to find problems let's say with defibrillators compared to pacemakers because those are designed, as Janey said to be pacemaker pluses, right, the plus is the defibrillation portion.

The big issue that comes here that the importance of this measure really is making sure that the patient is touched personally and some of these questions can be answered.

MEMBER JAMES: Two responses. First, as a clinician, the ability to follow-up and ensure that the questions are answered and that patients understand, helps to generate that compliance. We talked about that with drugs

previously. The same thing applies with devices to be used appropriately.

Secondarily is a health plan. This is considered an expensive type of investment in a patient. you want to make sure, as a health plan, as a payer, that you are protecting that investment. Doesn't that sound terrible?

MEMBER ALLRED: No, it is not terrible. I think it is great. Absolutely.

CO-CHAIR GEORGE: Any other comments?

MEMBER PHILIPPIDES: A very brief comment. Is there anyone here who places these? She just left. That's too bad.

I think sometimes the payment for these procedures are bundled. Is that correct? Such that, if you touch a patient within a certain amount of time after you place this, there is no additional income coming in to the system. So, this is another one of the sort of many metrics that is sort of going upstream against sort of the RVU and payment tide and trying to change behavior in sort of a very

difficult time.

So, my suspicion is sometimes when people go outside of these ranges before they arrange follow-up, that is maybe one of the first times they can actually start to get reimbursed again for the work. I know that sounds somewhat cynical but I think that is the reality.

So, I like this measure because it starts to push back against some of that stuff.

CO-CHAIR GEORGE: If I could just add that as a surgeon I am horrified that we even have to have a measure like this.

MEMBER ALLRED: Absolutely.

CO-CHAIR GEORGE: All right, let's go ahead and -

MEMBER HOLLANDER: Actually, I'm going to raise it. It is not that relevant to the voting but it will be in two or three years. I'm just going to throw it out there.

(Laughter.)

MEMBER HOLLANDER: It is the definition of a visit. And so my job now is I am

running a huge telehealth program for Jefferson and everybody is looking into doing visits by telehealth.

And so I think as measures like this roll out, this may be a perfect thing, particularly if you have a patient traveling 50 miles or 75 miles to get to a referral center and they have it planted, that you could actually do this by video conference.

So, I just put it out as we start to think about it in this and other measures, defining a visit is going to become important in the near future and we may want to do that.

MEMBER ALLRED: I would have to disagree in that I don't think you could do the first initial visit by teleconference because it is the hands-on. It is the looking at the incision. It is all of the different things going into it plus the personal interaction between you and the person who has put it in and you get your questions answered.

MEMBER HOLLANDER: My point is still

we should define the visit.

MEMBER ALLRED: Yes, you're right.

MEMBER HOLLANDER: It may or may not be appropriate for every condition but we should at least know what counts and what doesn't count as we go forward.

MEMBER ALLRED: Right.

CO-CHAIR GEORGE: All right, let's go ahead and vote on priority.

MS. LUONG: Polling for high priority starts now; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

High priority passes with 69 percent voting for high and 31 percent voting for moderate.

CO-CHAIR GEORGE: Okay, we will move on to scientific acceptability, the specifications and reliability testing.

MEMBER ALLRED: Okay, I am going to throw it to one of my colleagues.

MEMBER CLEVELAND: I'm happy to take over, at least for the reliability.

So, I think the strength of this, again, is emphasized. So, the clinician-level analysis I think that is a very -- for the specifications, I think that is the place where this needs to be.

The specifications will include codes. I guess, it is going to obviously transition in the administrative from ICD to ICD-10 codes. And so it is well thought of that way.

There is a fairly sophisticated algorithm that is detailed but in looking through that algorithm, I think that is a reliability, at least for the specifications present.

The validity testing, in terms of -- I don't know if we want to talk about the specs only or talk about the reliability testing, too, as well.

Okay, I don't know. Tom, do you have another thoughts?

MEMBER JAMES: No, you are summing that up right. I think the attribution methodology is going to be the key thing because

if it is billed out as a group, who is going to be responsible?

CO-CHAIR KOTTKE: We don't have trouble with attribution at Health Partners. We have got the algorithms and the algorithms are there. It is not a big --

MEMBER ALLRED: And I think the measure actually states that the person who implants the device has primary responsibility.

CO-CHAIR GEORGE: Liz?

MEMBER DELONG: Okay, that was my question. Is it the interventionalist or the PCP? I thought this was a transition, partly a transition measure, in which case you would expect the PCP to pick up.

DR. KUSUMOTO: So, my apologies for confusing you with regards to the definition. It is a transition measure in the sense that it is transitioning from a hospital situation to an outpatient situation. But it is the hospital -- so, this is important because it squarely puts the responsibility of this transition and the

correct hand-off into one unequivocal place. That is what is critical. The implanting physician.

CO-CHAIR GEORGE: Gerard?

MEMBER MARTIN: So, I have been waiting for my first time to do this today and that is, representing children. I know this is described as being for adult patients. This is absolutely one measure that there is no reason why it should be just adults because of the -- and this applies to children as well. And I don't know how you all deal with that. It is something that pediatric centers could do. I am sure that we actually do it. I hope. And I don't know why we are keeping this just to the adult age.

DR. KUSUMOTO: It is an incredibly important point. So, we will talk about this on the next section.

With regards to making sure that the test was done appropriately et cetera, Medicare fee-for-service with critical, just with regard

to the billing piece, at least for this, for our testing.

CO-CHAIR GEORGE: Any other comments on reliability, specifications? If not, we will go ahead and vote.

MS. LUONG: Polling for reliability starts now: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Reliability passes with 50 percent high and 50 percent for moderate.

CO-CHAIR GEORGE: All right, we will move on to validity.

MEMBER CLEVELAND: So again, validity for this measure was conducted at the level of the data element and, again, using Medicare fee-for-service claims. Essentially, the data was compared from claims to date in the patient chart and then computing sensitivity specificity, positive predicted value, negative predicted value. Those were all in the 95 to 100 percent range. So, that implies, again, at least supports, I think, indirectly the validity of

this measure. So, I, personally, did not see any problems with validity.

CO-CHAIR GEORGE: Any questions on validity? All right, we will vote on validity.

MS. LUONG: Polling for validity starts now: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Validity passes with 75 percent voting high and 25 percent voting moderate.

CO-CHAIR GEORGE: Moving on to feasibility.

MEMBER CLEVELAND: The data is --

MEMBER JAMES: It would be interesting.

CO-CHAIR GEORGE: Any other comments on feasibility? Okay, we will take a vote on that.

MS. LUONG: Polling for feasibility starts now: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Feasibility passes with 31 percent voting high and 69 percent voting for moderate.

CO-CHAIR GEORGE: Usability?

MEMBER CLEVELAND: So, for this measure, the proposed result is okay, but I think --

MEMBER JAMES: The only parts that I had any concerns about was thinking in terms of what public reporting would be. For me, as a primary care physician, how would I go about judging a cardiologist based upon having information on this?

So, I think for a cardiologist, it is great. I'm not sure for the rest of us what we are going to do with that information.

MEMBER CLEVELAND: Send him to somebody else.

CO-CHAIR GEORGE: Kristi?

MEMBER MITCHELL: My question to the developers: have you thought about the implementation of a measure like this in an ACO or some other closed system, such that it will address the issue about the data and be able to follow Mrs. Jones across time and setting?

PARTICIPANT: So, Kristi, the proposed rule that came out the other day, electrophysiologists, so if they were the ones using this measure, are actually not included in the one ACO piece; they are actually part of the exclusivity, they are excluded from exclusivity. So, obviously, that is in the proposed rulemaking. It is a three-day old proposed rule. So, we did -- that is something we looked at right away. So, to answer your question, the implications at this point if the proposed rule looks like the -- if the final rule looks like the proposed rule that physicians, electrophysiologists will continue to have the flexibility to practice in multiple settings.

And then if the patient's primary care provider is in a particular ACO, then that will allow them to ensure that the patient is going back to their hub.

CO-CHAIR GEORGE: Any other discussion on usability? All right, we will vote.

MS. LUONG: Polling for usability and

use starts now: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient information.

For use and usability, both criteria passes with 31 percent for high and 69 percent for moderate.

CO-CHAIR GEORGE: All right, so we will vote on the overall suitability for endorsement.

MS. LUONG: Polling starts now for overall suitability for endorsement: 1 for yes and 2 for no, please.

For Measure 2461, it passes with 100 percent yes for overall suitability for endorsement.

CO-CHAIR GEORGE: And are there any competing measures? We will move on to Measure 2474.

CO-CHAIR KOTTKE: Okay, cardiac tamponade and/or pericardiocentesis following atrial fibrillation ablation.

DR. KUSUMOTO: Okay, thank you very much again for the opportunity.

So, 2474 really looks at a procedural complication, pericardial tamponade during a procedure that electrophysiologists do. You heard from Michael from ACC earlier, and you guys all know AF is everywhere. Right? I mean it is in the water. I see it in my clinic all the time. And really a very, very difficult problem.

We had spoken about the stroke issue earlier with regards to the anticoagulation. The second issue is symptoms.

So, a fair majority of patients are asymptomatic with atrial fibrillation, but a large number of them actually have significant reductions in quality of life with this. We have medications and they can be used for this. Unfortunately, the medications at the end of the year failed half the time, if not more, and they are associated with significant risks.

For example, in the affirm trial, those patients who are on our strongest medicine, amiodarone, actually had a higher risk of being hospitalized in the ICU because of pulmonary

complications and other issues.

So, our alternatives for treatment are very poor. And for this reason, over the last ten years, it is really remarkable, I did my first atrial fibrillation ablation back in 1996 that in fact this has now emerged as one of the principle procedures that is done for the management of atrial fibrillation.

It is now being done more and more frequently, and the big issue with this is complications. As I tell all of my patients, this is an elective procedure. And what we need to do is to make sure that we avoid risk -- risk of stroke, risk of urgent surgery, risk of et cetera, et cetera, et cetera.

When you look at major risks of pericardial tamponade, where the heart is inadvertently perforated in the heart wall and, in fact, you then have fluid going around the heart is actually a signal important event. This is really, this event should not happen. You really think about this as a serious adverse

event.

I look back on our experience at Mayo Clinic, again, because we have systems in place because of this issue, we actually have not had any of these events over the last ten years at our place. And this is not trying to spout, et cetera. This is really because we took great effort to take a look at this, per se.

Now, my colleagues at the Mayo Clinic Rochester, however, have had some events. Again, not disparaging, again, different patient populations, et cetera. And it is anywhere from one to two, to three percent of patients. And the clinical outcomes with these problems are key.

So, if the Mayo Clinic Rochester experience about another 15 percent of those patients had pericardial tamponade actually went on to surgery. Those patients who have pericardial tamponade have longer hospitalizations. In a 100,000 patient database, looking at the nationwide hospital inpatient

sample, in fact the hospitalization was seven days for those patients who had pericardial tamponade versus those patients who did not, which was about a day to day and a half.

In addition, those patients who have had pericardial tamponade, even if treated, if you look at Chinese data, in fact 25 percent of those patients actually developed what we call postcardiotomy syndrome, where in fact they get chest pain and so forth. So, this is something that lives with them.

So, this really is an event that should not happen. And for this reason, we feel that this is a very important thing to measure, report, and hopefully minimize.

CO-CHAIR KOTTKE: Great. Thanks. Joe, Jason? Who?

MEMBER SPANGLER: Let me turn my speaker on. Sorry about that.

This is a negative or a complication measure. So, the lower the number, the better the quality, just so we keep that in mind. We

had a few of those earlier to discuss.

I had some issues with the evidence with this, mainly because I didn't feel like there was any QQC provided. It seemed like it was only expert opinion, even though it was based on a guideline, but the expert opinion didn't seem very specific to me for the actual measure that we were looking at.

So I -- going through the algorithm that we have, I actually thought the evidence was not low but actually insufficient; I didn't think there was evidence that went directly with the measure that we were looking at. So, that was kind of my major issue.

MS. JOHNSON: So, let me interrupt here and just make sure everybody understands our criteria for evidence for outcome measures.

So, for outcome measures, we do not necessarily ask for the QQC. So, what we want the developers to provide for us is a rationale for saying that at least something that they can do can help that outcome. So, that is what you

need.

MEMBER SPANGLER: Okay, so I apologize for that. Sorry about that.

MS. JOHNSON: No, that's fine.

MEMBER SPANGLER: So, then I would say there is a rationale. I would agree with that, but I still feel like that the evidence -- there is a rationale, but it still feels to me that the evidence that they are providing is not directly related to the actual measure that we are looking at.

CO-CHAIR KOTTKE: Other comments? Joe, do you want to talk about why it is bad to poke a hole in the heart?

MEMBER CLEVELAND: Thank you, Chairman. I think the only other thing I think that confounds this a little bit -- I mean I think this is something where I can understand the rationale and as a cardiac surgeon I agree this is a bad problem when it occurs. And at our place, fortunately, I think it has been over five years since we have had to take anybody to the

operating room for this. So it is infrequent but when it occurs it is disastrous.

So, the only question I have, and this is in our measure worksheet, is the question about other structures or processes of care. And I guess I would ask the developers, are there standards for periprocedural management of anticoagulation? Because I assume all these people have -- well, maybe they are not all on anticoagulation as we learned, but a vast majority of them are. So, are there standards for, in terms of target INRs before one proceeds with this and things like that? So, I could see if one lab is willing to do a catheter-based ablation with someone with an INR 23 and everybody else has to be less than two. How do we control for that? Because that could affect the outcome.

DR. KUSUMOTO: So, a great point with that. So, the problem with that is that there is going to be variability with regards to then you have introduced variability into that, which

makes it somewhat difficult.

For example, for patients with persistent atrial fibrillation, when they leave the lab, they are going to be in sinus rhythm. So, in fact, because this is their highest period for having stroke, in fact we demand now actually before we used to do low molecular weight heparin, actually we now do those procedures with patients on full anticoagulation. And actually our lab has looked at the, they call them novel, but the new oral anticoagulation agents. In fact, it is actually fairly safe. We have not seen, albeit in small numbers, any change sort of in our outcomes with regards to pericardial effusion and tamponade. And that is why we specifically chose tamponade as opposed to just the development of effusion.

I mean, this is something that is very definable. This is someone who now has a second procedure with a needle in their chest or open-heart surgery, et cetera, which really then becomes a signal event.

CO-CHAIR KOTTKE: Thanks. Henry Ting online. Welcome Henry. You had a question?

MEMBER TING: Yes, not a question. Just a comment. I thought the evidence was a pass because this is an outcome measure.

CO-CHAIR KOTTKE: Okay, thank you. Gerard, are you voting? No. Judd is, though.

MEMBER HOLLANDER: Yes, so I guess you are, effectively, trying to make this a never event. And so then my question concerns unintended consequences.

So, I have no knowledge in this area; I don't know how much of this is operator error and how much of this is patient-related factors. But my question is: are they going to be patients who might have received an ablation before who now won't because the operator considers them too high risk?

DR. KUSUMOTO: So an absolutely great point. But again, and there are some patient characteristics that are associated with the development of pericardial effusion that we will

obviously go through if this passes this hurdle. Having said that, the great majority of this, albeit small, number is really relative to the physician himself or herself, in fact, in how they are handling those catheters inside the atrium. I mean it really is where gentleness is really critical, no matter how much you are ablating or not ablating, et cetera. The gentler the ablation is, the better the catheter handling, et cetera. So, this really is a quality piece.

CO-CHAIR KOTTKE: Great. George?

MEMBER PHILIPPIDES: To follow-up on the comments you both made: would a practitioner, after this goes into effect, shy away from doing this procedure on patients who require anticoagulation because, by definition, they have a much higher risk of tamponade and effusion?

DR. KUSUMOTO: So again, an interesting point. And I think this data is evolving dramatically and quite quickly and couldn't be included into the application. In

fact, I would say that 95 percent of our patients that we ablations on actually now, over the last several years, because our mainly persistent atrial fibrillation are on anticoagulation.

So, while I think there is certainly no question that if you are on anticoagulation are more likely than to bleed and then have tamponade. But then what is going to happen, I will give just a personal anecdote here from a case from a few days ago, actually. A woman who actually came to us for an ablation for atrial fibrillation went and did this transseptally. You have to go from the right atrium to the left atrium to sort of get your catheters into place. Well, as part of our sort of zero tolerance for this event, we have an intracardiac echo in place during this procedure. What this is is an ultrasound so we can monitor the effusion.

Well, she had a slight effusion to begin with, which is common in patients on anticoagulation. We did this and there was this question: Was there a slight increase in the

effusion?

So, because of these sorts of systems in place, we then stopped the procedure. And then, because of our zero tolerance, we actually then woke the patient asking are you having any chest pain. In fact, she was having a little bit of chest pain. So, we in fact then stopped the procedure. And then what?

Well, I am not sure whether or not she would have, if we would have then anticoagulated her further as we would often do as we do for the rest of the procedure, would she have then gone on to develop tamponade. But the point is that even in these patients who have higher risk for tamponade and for developing pericardial effusion, there are ways to monitor this, to make this still a never event.

CO-CHAIR KOTTKE: Let's see. Jason and then Liz.

MEMBER SPANGLER: So, I'm going to go back on something I earlier said because I am relooking at this. Obviously, I think I looked

at it not as an outcome. So, the evidence is, I agree with Henry, not as important.

The question I have is when we are looking at a health outcome measure and the processes and structures of care, you have mentioned some of the things that can improve. I guess my question is, you know you talk about high-quality A-fib ablation and some of the things you have done. Are those the standards, the benchmarks?

So, what I am trying to figure out is if we have this type of measure and people don't score high on this measure, do they know what to do? Do they know how to improve upon this measure? So is it because they are not performing at the standard of care or is this because this is a complication that happens? I mean I guess my question is: Can this be a never event? That is, basically, my question.

DR. KUSUMOTO: My personal thought is yes or certainly pretty darn close to it. I mean, who knows? But I do think that yes, this

can be made a never event with good systems of care, and good training and teaching, and all of those kinds of important things.

MEMBER SPANGLER: And it doesn't matter whether it was out there?

DR. KUSUMOTO: Yes, I think that they are, in fact. You know again, that is why I think this sort of measure is so important. That is why I feel so passionately about that and you hear this in my voice. I think this is something that can be done.

We do it in our small system. What do we do? Well, we have people when they want to learn AF ablation, they come to our place as visiting scholars for a week to week and a half. We go through our systems with them. In fact, not just the procedural piece of this is how I move my hands, et cetera. Well, these are the things that we do. Here is the checklist that I go through. Once I have done the transseptal, I am looking at the ST segments. I am doing this. I am doing this. I am doing this, et cetera.

Here are things that can happen.

So, we do this sort of in our own sort of place in a sort of -- in an organized fashion but, nonetheless, not nationally. If you had a measure like this, there would be an emphasis, I think, of moving this ball forward to make this a never event.

CO-CHAIR KOTTKE: Liz and then George.

MEMBER DELONG: So, I am a little bit confused about perverse incentives. Is there not an ongoing trial of ablation versus is it medical therapy?

DR. KUSUMOTO: Yes.

MEMBER DELONG: So, there is an alternative and the jury is actually out as to which is better, unless there are pre-specified patients who are destine for ablation. Is that the case?

DR. KUSUMOTO: So a great point. So, let me just, again, I don't want to take the committee's time too much. There have been trials that have looked at ablation versus

medication, both in nuance in atrial fibrillation and in patients who have actually failed a prior drug therapy.

Let's take the prior drug therapy first because those were the ones that were done first. So, these were patients who had failed one drug; they randomized to another drug, which, obviously, doesn't work, and then the ablation does better. It does better in the sense of less atrial fibrillation and improved quality of life, those kinds of indices. The study that you are referring to, CABANA, is looking at hard outcomes with regards to stroke and hospitalizations.

MEMBER DELONG: So, I guess the perverse outcomes are there are alternative therapies, potentially, and to avoid doing an ablation on a patient would mean that you might be switching the patient to something that is equally effective, but you do not capture the revenue from doing that. And I don't know which one weighs out.

DR. KUSUMOTO: But you would capture

-- again, these are patients often, at least at our place, they are drug refracted. I mean, they really don't have a ton of options. I mean this is really, it is sort of the last stop. And I do think that if you make the hurdle a little higher, with regards to we are going to measure this and we are going to report it and we are going to show it off to the world that, in fact, I think that you would then put the onus on the provider doing this procedure to in fact do a high quality job.

CO-CHAIR KOTTKE: Okay, are we ready to vote? Oh, two people. Tom and then Joe.

MEMBER JAMES: Mine's a question. And speaking to the relative importance, what I found is one article from Sue showing 2.9 percent incidence of tamponade with this. The paper talks in terms of two million people with atrial fibrillation. I am trying to get down to what is the incidence of all of this to get a relative importance?

DR. KUSUMOTO: Yes, so that is good.

We will talk about it, too, hopefully again if we pass the validity and reliability piece. We in fact took the large database of 600,000 patients, a million sort of events, et cetera, and kind of looked at sort of the incidence.

And if you kind of run through those numbers, you are looking at a couple thousand sort of patients, which if, in fact, you then put this on the, what do I want to say, the people who do the majority of these procedures who are, in fact, those people who do less than 20 procedures, you would then get rid of, if you made them zero, 90 percent of the events.

Now, I am not saying that it would potentially be zero but, nonetheless, there is clearly -- if we again talk about gap and potentials for benefit from this measure.

CO-CHAIR KOTTKE: So, this really begs the -- you know I think the right denominator is the patient with the ablation, not the patient with atrial fibrillation. And this has to do with 0715, too. It is the denominator is the

patient in the cath lab.

Joe and then Gerard.

MEMBER CLEVELAND: Yes, my only other comment, I guess, to address at least Judd and maybe Liz is would people not want to do this or would people be denied this procedure? As a cardiac surgeon, there is, actually, we do open maze procedures still that actually if you are seeing this growing effusion, I am going to stop the anticoagulant and maybe this is a person who should go for a radio-frequency maze.

So, there is another alternative procedure that can be done. Obviously, it is more invasive and differs from drug therapy, but it is not as though these people will not be treated or at least not have that option to.

CO-CHAIR KOTTKE: Gerard?

MEMBER MARTIN: Then they will really have chest pain.

(Laughter.)

MEMBER MARTIN: Sorry, I thought we would at least laugh a little bit this afternoon.

With the tamponade, how much of it is due to the transseptal versus the actual ablation?

DR. KUSUMOTO: So, a little of both. And it actually, I think, varies depending on the type of ablation that in fact you do and the experience of the operators.

I actually believe that when you look at the low number of operators, is actually from the transseptal rather than the ablation itself.

MEMBER MARTIN: I mean it does raise the question of whether the title of this should be cardiac tamponade during transseptal procedures in adults.

DR. KUSUMOTO: So, we looked at that a little bit. So, there is some data to suggest that in fact there is a higher, and again, from experienced centers.

So, we do ablations on the right side, where you don't have to do the transseptal there. The pericardial tamponade raised very, very low, vanishingly low. If you add the transseptal it

adds an additional sort of half a percent or a percent. If you then do the atrial fibrillation ablation, that is when in fact you do have more.

The problem is so again when we get to the measure itself, trying to tease out those patients with regards to just doing the AF ablation only.

CO-CHAIR KOTTKE: Okay, anybody else have the urge to -- Tom is here. Okay, shall we vote?

MS. LUONG: Polling starts now for evidence help outcomes: 1 for yes and 2 for no.

CO-CHAIR KOTTKE: While people are voting, it reminds me of a cartoon we had outside of our cath lab, where the fellow is standing there and the catheter is coming out the patient's mouth and the preceptor says, I told you not to push.

(Laughter.)

MS. LUONG: For evidence help outcomes, 94 percent voted yes and 6 percent voted no.

CO-CHAIR KOTTKE: Okay, opportunity for improvement in disparities.

MEMBER SPANGLER: So, you know the performance gap is interesting because if you want it to be a zero event, then there is a performance gap. But if you look at the complication rate, and I don't have experience, obviously, with these patients at all. I mean it seems pretty low; I mean the numbers they give, 1.2 to 2.4. So, in that respect, there may not be one but if were truly think there is a zero event, then there is a gap that there can be improvement upon.

There is no, I didn't see any data around disparities.

CO-CHAIR KOTTKE: Okay, Henry, you have a comment?

MEMBER TING: Yes, I do. I don't think this can be a never event. I sort of disagree with the presenter because these are things that happen at large centers that do a lot of cases. And the Mayo Clinic Rochester was

referenced as one of the best and highest volume centers that do these procedures. And some of us do the transseptal, but some of those are also due to the aggressiveness of the ablation. And so when you are doing patients who are coming back with recurrent atrial fibrillation who have already had an ablation, this can -- if you are not willing to do the procedure and be successful and try very, very hard, then you will never have tamponade or pericardial effusion. But if you are actually trying to cure the disease and trying very hard, then you can cause a pericardial effusion.

So, I don't think this is a never event. I think if you are going to try to do these in complex patients who have recurrent atrial fibrillation after several ablations, this is something where you can have this event.

I want to make a few other comments before actually I need to go to another meeting. It is probably out of place, but the other thing I wanted to say is it is now said as and/or. It

can't be and/or. It says and/or cardiac tamponade or pericardiocentesis. It is either and or it is or. I don't know how to interpret and/or in a numerator. So, that has to be fixed.

And then the other thing I would point out is that this misses all patients who have a pericardial effusion caused by the procedure, which may be moderate to large size, but did not require pericardiocentesis or cause tamponade. Not all pericardial effusions cause tamponade.

So, when you do these procedures, you wind up with a moderate to large effusion and you don't have tamponade, this is completely missed in the numerator.

And I do think the distribution of evidence, opportunity for improvement is quite low because if you are talking about 1.2 to 2.4 percent, and that is a range, that is not a big area for improvement.

DR. KUSUMOTO: I'll agree with Henry in the sense that that is why it is chosen as we can pick on pericardial tamponade and that

physiology because pericardial effusions are present. And in fact, as I mentioned, you can see this at baseline in some of our patients. So I think that that is a point.

I will differ a little bit with regards to the never event, as I mentioned earlier. Maybe it could never be a never event but boy, it sure would be nice to have it be a never event and be, obviously, as close to zero as possible, even with these complex ablations that we are doing. I mean that really implies that maybe the way that we are doing it is not as good. So, I would make that argument sort of with that.

I do want to note that there is one study that has come out since this application went in that in fact did show some disparities where women were more likely to have pericardial effusion than men. I don't know if that is just because of atrial thickness compared to men. There are a whole host of issues one can think about but, nonetheless, that data is out there

now.

CO-CHAIR KOTTKE: Anybody have any thoughts on Henry's objection to and/or?

I mean they are basically saying if it is just an or, if you have both, you are out. But why would you be out if you had both pericardial tamponade and a pericardiocentesis? I mean, pericardiocentesis is the treatment of tamponade, and I mean I think the and/or is fine myself.

MEMBER TING: Is it correct English? I think it is just or, then. If you have either one, isn't it or?

CO-CHAIR KOTTKE: But then if you have both, you are out.

MEMBER TING: If it is or, you have both, you are in. It is or.

CO-CHAIR KOTTKE: Oh, okay.

MEMBER TING: From a medical perspective, I don't think you are supposed to use and/or.

CO-CHAIR KOTTKE: Oh, we just got an

official inspect language. It is "or."

Okay, Liz.

MEMBER DELONG: So, the conversation seems to imply the need for risk adjustment. But clearly, the event rate might not support risk adjustment. But it is an outcome. I mean it is a bad outcome.

It would seem we are still aiming to drive it down. I just wonder about ranking sites according to this outcome if they haven't been credited with the risk that they are taking on.

MEMBER TING: I would agree with that comment, Liz, because if you only adjust it for age and gender, which was done for this outcome, clearly this is not the only thing that correlates with the outcome.

CO-CHAIR KOTTKE: Other comments? Are we ready to vote?

MS. LUONG: Polling starts now for performance gap: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Performance gap has 6 percent high, 47

percent moderate, 35 percent low, and 12 percent insufficient.

MS. JOHNSON: So this is in our gray zone. We do continue.

CO-CHAIR KOTTKE: Okay, priority.

MEMBER SPANGLER: I would go back on an earlier comment I made within the National Quality Strategy just because it mentions cardiovascular, whether everything here is a high priority. So I would question whether this is a high priority or not.

Going back to something Tom said, it seems to be low prevalence overall. I agree severity can be high. There was no data presented about severity. You did mention the hospital stay, but that wasn't in the application. So that is new information and there is no cost data.

I think the priority is questionable. It is definitely severe when it happens, but everything else around priority, I would say would be low.

DR. KUSUMOTO: I would apologize. The hospitalization data was in the nationwide sample, which was in the thing, but again, buried way in the references, et cetera. So my apologies to the group on that.

CO-CHAIR KOTTKE: Other comments? Joe.

MEMBER CLEVELAND: I agree in theory that, while infrequent, I would still say for an elective procedure this is a calamitous type outcome. So, it just depends on how you view it.

I mean if there is a surgeon running up to the EP lab to open somebody's chest, it is not a good day for you or the patient.

CO-CHAIR KOTTKE: My feeling is that a patient has the right to be safe. And I think this makes it -- the right denominator is the patient going to the cath lab. It is not all patients with AF because it is an elective procedure.

And I disagree somewhat with Henry. If risk of tamponade is that high, maybe you call

Harzell Schaff and have him do the procedure as an open; it is an open chest procedure instead of messing around with catheters.

MEMBER TING: That is a patient preference issue. So, if the patient chooses that --

CO-CHAIR KOTTKE: Yes, sometimes you stop, though.

Okay, any other -- yes, Judd and then --

MEMBER HOLLANDER: So, I am trying to put my patient hat on. This is something I would love to see publicly reported. If my family members or I need to go for an ablation, I want to be able to pick the doctor that has done a lot of procedures and has a low complication rate.

So, I think with A-fib growing, to me that makes this a high priority. I want to know the answer to that.

MEMBER TING: Judd, what needs to along with this is success rates from atrial fibrillation. You know sort of like how many

patients have successful ablations with this operator, as well as recurrence rates at one year and two years.

CO-CHAIR KOTTKE: And tamponades.

MEMBER VIDOVICH: I just wanted to say it is somewhat of a crude measure like perforation yes or no, but it is a good quality measure. And again, it may differentiate from good operators or from centers. And I think it is not like a publicly reported, I don't know, CABG outcomes in New York, you know if people actually went to a different state. I think that this actually will improve overall quality and move through it. It is a rare event, but it is a worthwhile one.

CO-CHAIR KOTTKE: George and then Joe.

MEMBER PHILIPPIDES: Henry, I think I missed what you mentioned a moment ago. Are you saying that looking at this metric without also knowing how many successful tough ablations somebody did is incomplete and might not give a fair picture of a person's quality?

MEMBER TING: Absolutely. So, if this is equated to angioplasty and stenting, you would want to know successful procedure without complications. An equivalent complication would be coronary artery perforation requiring coronary artery bypass surgery as an emergency part of the procedure.

You know you would want sort of a family of measures that includes -- if you are truly the patient, you want to know what are the changes that this operator would be successful, their experience, as well as their complications; this is one of the complications. This is not the only one. And it is probably not the most common complication from atrial fibrillation ablation either.

MEMBER PHILIPPIDES: Yes, because I do think sometimes patients get sent on -- a physician in the community might send on the tougher cases into Mecca to be done. And so by definition, they will do the easier cases but send the tougher cases on. And I do think there

has to be some understanding of the different patient populations. Surgeons see this all the time in second opinions for bypass.

MEMBER TING: And there is no risk adjustment with this measure right now, just age and gender.

CO-CHAIR KOTTKE: The point I would make is what Henry and, I guess, Liz have brought up previously, is that is where risk adjustment would help with that type of thing.

MEMBER PHILIPPIDES: And that wasn't done, except what Henry mentioned already.

CO-CHAIR KOTTKE: Gerard, are you voting? I can't see -- oh, that's George. He's hiding.

Okay, we are ready to vote, it looks like.

MS. LUONG: Polling starts now for high priority: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

For high priority, 13 voted high, 56 voted moderate, 25 voted low, and 6 voted

insufficient. It passes.

CO-CHAIR KOTTKE: Okay, scientific acceptability and reliability.

MEMBER SPANGLER: Do you want me to keep going? Do you want to take over? What do you want to do? Or Henry, do you want to give it to Henry?

MEMBER CLEVELAND: I'm out of my league here. I was going to actually bunt this or whatever. I will be honest with you. I don't understand the crosswalk that is done. So, if you do, you can help me.

MEMBER SPANGLER: So, I mean I think the measure specifications are clearly defined, as is the data source. It seems to be implemented.

There has been discussion about the denominator, but I didn't have issues with the denominator. I don't know if other people wanted to speak to that. And if we want to talk about that, we can now. I can go on to testing, reliability testing if you want or if we want to

talk about that, we can.

CO-CHAIR KOTTKE: Sure. Are we going to talk about reliability testing?

MEMBER SPANGLER: Testing?

CO-CHAIR KOTTKE: Yes.

MEMBER SPANGLER: All right. So they did reliability testing through the beta-binomial model, measuring signal to noise ratio. We talked a little bit about signal and noise earlier. And there is also demonstrated, I thought, high reliability.

So, I didn't have any issues with the reliability.

CO-CHAIR KOTTKE: Okay, any -- let's vote, seeing no dissension.

Who had questions? Jason, did you have questions about the denominator?

MEMBER SPANGLER: No, I didn't. I thought somebody else did but I didn't.

MS. LUONG: Polling starts for reliability: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

For reliability, 38 percent voted high, 56 percent for moderate, 6 percent for low.

CO-CHAIR KOTTKE: Validity.

MEMBER SPANGLER: It looked like empiric and face validity and face validity was done in measure development. They did some testing and the results showed ablation was the most common procedure and one with the complications being measured. So it seemed to correlate. The validity seemed to correlate with what they were looking to measure. So I thought the validity was fine.

CO-CHAIR KOTTKE: Did you have the urge to --

MEMBER PHILIPPIDES: I have a quick question.

So we are proceeding as though that says cardiac tamponade or pericardiocentesis. Is that correct? Okay.

CO-CHAIR KOTTKE: Joe?

MEMBER CLEVELAND: I was just going ask in terms of the three-year rolling average,

the rationale for that. Is it just because of the infrequent nature of this complication?

DR. KUSUMOTO: Correct.

MEMBER SPANGLER: Sorry, I didn't know this was part of this. Again, we multiply have talked of this but I think it is important that there isn't really risk adjustment done as much as should be in this. I think that is an important point to keep in mind.

CO-CHAIR KOTTKE: Tom?

MEMBER JAMES: Yes, a question. Do you think there is sufficient data to create a differentiation among the measured entities? It is going to take enough volume in any one place and will there be sufficient differentiation with this measure than can help create some improvement or is this not a comparative measure but one in which, against each other, that measure against the absolute.

DR. KUSUMOTO: Well, again, I think it is the absolute that you really want zero. And I am going to, again, argue the other side of the

coin with Henry and agree with my surgical colleague here. You know we talk about outcomes in terms of yes, a successful ablation. Well, that is kind of like well, you know, the patient survived whatever that line is. I mean you really want to avoid complications. And this, again, when you look at all of the evidence from these large claims databases, et cetera, pericardial tamponade is the most common serious complication that occurs. I mean you can make arguments about stroke and other things having bigger sort of issues but when you look at absolute numbers, it is pericardial tamponade.

CO-CHAIR KOTTKE: Judd.

MEMBER HOLLANDER: Getting at Tom's point, maybe I am a little redundant. But if we are comparing across providers and that is sort of our hopes for this, how many cases does the typical operator do a year and is there any statistical difference between them?

DR. KUSUMOTO: Yes, that is a great point. So, in the paper that was included in

your piece looking at the nationwide inpatient sample, in fact there was a dividing point at about 25. So, if you did less than 25, in fact, you had a higher complication rate with regards to pericardial tamponade compared to those patients who did greater than 25, which was borne out actually in our data, too.

What is important to know that is if you take those 25 and 90 percent of sort of the complications, mainly because they have the higher volume, are in fact in that group of patients, even given a three-year rolling average. So, I think the shortcut for these physicians, I think.

You either get into it, you do it, you do it well and you get publicly reported or you don't.

MEMBER SPANGLER: What are those numbers, comparing the greater than 25 or less than 25? Or are they --

DR. KUSUMOTO: Oh, so pardon me. So, it is looking at the ranges are in single

percents. So, it is going to be when you go greater than 25, it is about one to one and a half percent. And then when you go less than 25, then it goes to two and a half to three. That is correct for that data. That's right. Correct.

MEMBER HOLLANDER: So, that wasn't exactly my question but I think that is interesting to inform volumes better.

I am saying if I am trying to choose between three providers, do they each do 100 cases a year, so if one is two percent and one is one percent, it is a one-patient difference? Or are you doing 500 patients a year, where maybe a two or three percentage is statistically relevant?

DR. KUSUMOTO: Yes, you know this question is going to be how does this go. Because I do think that the way that this procedure should go, as a sort of a personal editorial, is to sort of large places that do a lot of them can be very good, particularly if they are going to be very complicated and have

higher risk.

CO-CHAIR KOTTKE: Looks like we are ready to vote.

MS. LUONG: Polling starts now for validity: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

For validity, it passes with 6 percent voting high, 63 percent voting moderate, 25 percent voting low, and 6 percent voting insufficient.

CO-CHAIR KOTTKE: Feasibility.

MEMBER SPANGLER: So, the data is collected through administrative claims and there is, it looks like, readily available electronic form for this. There didn't seem to be any identified areas of concern. So, I thought the feasibility is pretty good.

CO-CHAIR KOTTKE: Seeing no motion, let's vote.

MS. LUONG: Polling starts now for feasibility: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Feasibility passes with 56 percent voting high and 44 percent voting moderate.

MEMBER DELONG: So, again, a naive question and I'm sorry I didn't ask this. If the patient does experience one of these complications, is it unlikely that they would have already been discharged and reporting to a different hospital or something? I just don't --

DR. KUSUMOTO: I'm sorry. So, again, so the complication happens. You are going to be in the hospital for a week because of the complication.

MEMBER DELONG: But the complication is evident immediately. So, they haven't gone somewhere else.

DR. KUSUMOTO: Oh, yes, absolutely. No, no, this is procedural.

MEMBER DELONG: Okay, that is all I wanted to know. I appreciate it.

CO-CHAIR KOTTKE: Usability and use.

MEMBER SPANGLER: So it is not currently publicly reported but it is being

considered in the quality programs for CMS for 2015.

So, again, I think the comment I made earlier, just a little red flag, we don't know because it is not being publicly reported but I think this probably has moderate to high usability, depending on implementation but I don't think there will be any problems.

PARTICIPANT: And since we submitted the application and actually this measure is in PQRS 2015.

MEMBER SPANGLER: So, it was included in the final rule, I guess.

CO-CHAIR KOTTKE: Joe?

MEMBER CLEVELAND: I'm just going to with a comment made earlier today about public reporting. And I fully submit that I think for this measure to really have full impact, it is going to have to be publicly reported. So, I completely endorse it.

CO-CHAIR KOTTKE: Let's vote.

MS. LUONG: Polling starts now for usability and use: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Usability and use passes with 44 percent voting high and 56 percent voting moderate.

CO-CHAIR KOTTKE: So, overall?

MS. LUONG: Polling starts now for overall suitability for endorsement: 1 for yes, and 2 for no.

For overall suitability for endorsement, 81 percent voted yes and 19 percent voted no.

CO-CHAIR KOTTKE: And I assume there are not competing measures.

MEMBER SPANGLER: No.

CO-CHAIR KOTTKE: Okay, thank you very much.

If somebody can find Sana. Oh, she left? Oh.

CO-CHAIR GEORGE: All right, we will

go on to the least measure of the day, 0715.

CO-CHAIR KOTTKE: Unless somebody rebels, we are going to finish up today, the last measure.

CO-CHAIR GEORGE: Any rebelling?

CO-CHAIR KOTTKE: Seeing no rebelling, we will go.

MEMBER SPANGLER: Do we need to let the -- if people are expecting public comment at this time, do we need to let them know or that is just assumed? Only because the schedule says 4:45. So, I am sure they are assuming that we are not finished, that they have to wait but I don't know if people are waiting right now for that.

MS. ISIJOLA: Operator, can you open the line for public and member commenting?

OPERATOR: And at this time, if you have a public comment, would you please press *1? Again, for public comment, please press *1.

And there are no public comments at this time.

MS. ISIJOLA: Okay, thank you.

CO-CHAIR GEORGE: All right, so Gerard, Liz, and Tom Kottke are the discussants. We will hear from the major developer.

DR. BERGERSEN: Good evening. Hi. My name is Lisa Bergersen. I am an interventional pediatric cardiologist from Boston Children's Hospital. I am the measure developer sponsored by Boston Children's Hospital to be here today.

First, I will just start by defining the metric for you, which is the standard adverse event ratio in patients less than 18 years of age for the outcome adverse events.

The numerator for this metric is the occurrence of adverse events at an institution divided by the risk-adjusted expected rate of adverse events due to the case mix complexity at the institution.

The denominator is derived using CHARM methodology, which is a risk-adjustment methodology based on three procedural characteristics and patient characteristics; one

procedural characteristic, two patients.

The procedural characteristic is the procedure type risk group. I will give you a little bit of background on that. Patient characteristics being age and the presence of hemodynamic variables.

We have shifted gears considerably from adult measures. As some background, congenital heart disease affects one in a hundred children. And our field of pediatric interventional cardiology is rather young.

Over the past couple of decades, tools and equipment and procedures have evolved considerably, such that where it was primarily diagnostic a couple of decades ago, we are doing more and more interventional procedures which either complement or replace some of the surgical techniques used in congenital heart disease. We don't do just one type of procedure. There is a multitude of different types of procedures.

About a decade ago, reports of adverse events came from single institutional

experiences. And institutions would contend that you couldn't compare their adverse events to others because their case mix complexity might be different. Thus, there was a need to develop a way to allow for equitable comparisons among institutions.

Adverse events were agreed to be an important outcome from these procedures and, therefore, risk adjustment methodology was necessary. None existed. Therefore, in 2006, we put together a small group of institutions which were put together to have a geographic distribution, have some variation in case volume, with the hope that that data set could support a generalizable risk adjustment methodology for the field.

It was limited to eight institutions to allow for reliable data capture and assurances that there wouldn't be biases in attribution and classification.

This data set was collected between 2007 and 2010 and from this dataset, we were able

to develop procedure type risk groups. Ultimately, four categories of the multitude of different procedures that we performed categorized into four groups of similar risk; half of them, being in category one, with decreasing frequency with increasing risk.

This was, by far, the most important factor in adjusting for the risk of important adverse events in our population, which range from two to eight percent across the institutions.

That risk adjustment methodology, the model, all of the factors were published and the methods are completely translucent within the literature. To date, there hasn't been a publication in a separate dataset that I can cite for you as a validation dataset for the methodology.

Therefore, in the materials that you received, we have provided a preliminary analyses or an analyses on another multi-center database. So, the C3P0, after achieving our initial aims,

our goal was go beyond benchmarking and risk adjustment methodology and explore quality improvement initiatives.

So, we invited additional sites and expanded our participation to 15 sites with the goal to reduce radiation exposure. While that is the primary goal of our current project, we still report the standard adverse event ratio to the institutions for their internal quality improvement processes, both by actually institution, as well as provider, which we also do internally.

This metric, although it is put forth to you as an institution-based metric, internally we use it for performance reporting, for providers to the Board of Registration and Medicine at Boston Children's Hospital, as well as the Department of Public Health. The measure put forth to you is at the institutional level.

Oh, so the data set that I showed you has not been published. We put it in a validation data set from January to September.

Our intention is to audit -- actually not our intention -- we will be auditing 2014 data in 2015 what you have presented before you.

I think that is about what I wanted to say about some of the background on the measure. Thank you for your time and I will answer more questions as we go along.

CO-CHAIR GEORGE: Thank you and we will go on to the discussion of the evidence.

CO-CHAIR KOTTKE: Measure 0715, standardized adverse event ratio for children less than 18 years of age undergoing cardiac cath. It is the ratio observed to expected clinical important adverse events, risk adjusted using the CHARM method. It is facility level. It is an outcome measure. And I would say the evidence is high.

CO-CHAIR GEORGE: Any discussion on the evidence? Tom.

MEMBER JAMES: First, as a pediatrician, I am glad to see this. But secondarily, why the exclusion of those

facilities with less than 50? We just had the discussion about volume-related improvements.

DR. BERGERSEN: Yes, so the derivation data set, as well as any testing data sets thus far have been in centers that are freestanding pediatric heart centers. So, it hasn't been centered in centers that perform a small volume of cases, as this point.

CO-CHAIR GEORGE: Any other comments on the evidence?

If not, we will go to a vote.

MS. LUONG: Evidence of outcomes polling opens now; 1 for yes, and 2 for no for Measure 0715.

One hundred percent voted yes for evidence of outcomes.

CO-CHAIR KOTTKE: Okay, opportunity for improvement. Observed adverse event rates from eight pediatric hospitals used in testing are included in Section 2(b)(5.2). Rates from these facilities range from 1.71 percent to 7.86 percent. The developer cited several studies

that likely include rates of complications following cardiac cath in children. So, there is no information on disparities.

And some of the data referenced from 58 centers were just reported at AHA, nearly 20,000 procedures, adverse events of 1.9 percent.

So, it seems to me there is opportunity for improvement. And the denominator, I think the denominator is kids going into the cath lab, not the total, it is irrelevant that it is one percent of kids have congenital anomalies. Kids deserve to be as safe as possible going into the cath lab.

CO-CHAIR GEORGE: Liz?

MEMBER DELONG: So, to further demonstrate my ignorance with respect to complications but the list of complications that you are accumulating is pretty long and it wasn't clear to me that they were all really significant. And there was no tabulation as to when you did collect these. What was the tabulation? What were the complications? And

did they fall more in the moderate range or did they fall in the severe range? I mean I have no clue.

DR. BERGERSEN: So, in some of our other publications, we have elaborated. The overall rates of what we are calling clinically important adverse events is four percent. And of those, about one percent are going to be these life-threatening high-severity events. And the other two to three percent are going to be the other major adverse events.

One of the unique characteristics of the registry is that we have assigned a severity level, which has been adopted by the international pediatric congenital cardiac code for severity of adverse events. And according to EKOS, there is definitions for severity levels. So, clinically important events would come under level 3, 4, or 5 events and the less important events were not included in the measure.

MEMBER DELONG: I guess my question would be suppose you have got a site that has a

two percent mortality rate versus a site that has at three percent rate but nobody dies. They have got a three percent rate of some of these relatively less severe complications. They are going to look worse.

DR. BERGERSEN: That is a good point. Fortunately, within cardiac catheterization for congenital heart disease mortality is exceedingly rare, which is one of the reasons we couldn't use that as an outcome measure for the field.

MEMBER DELONG: Sometimes we use composite measures because we don't have enough events. But once we use composite measures, it seems that they all have to be serious enough so that we can compare the event rates.

CO-CHAIR GEORGE: Any other comments? If not, we will vote.

MS. LUONG: Polling starts now for performance gap: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

For performance gap, it passes with 13 percent voting high, and 80 percent voting

moderate, and 7 percent voting insufficient.

CO-CHAIR KOTTKE: Priority. The developers state that congenital heart disease is the leading cause of morbidity/mortality affecting one percent of infants and that cardiac cath has become a common quote interventional procedure with therapeutic goals complementing surgical strategies and, at times, eliminating the need for surgery.

Stated like all the others, it is a National Quality Strategy priority area. Again, I would say that if you are going to do pediatric caths, you ought to be good at it and you ought to have low complication rates. And I think the appropriate denominator is the child going into the cath lab. Really the one percent doesn't have anything to do with that.

And so, I think it is high priority.

And I think also the fact that it is internal rather than external, so that pediatric cardiologists who are doing this or are saying we need standards for our own practice raises the

priority.

MEMBER MARTIN: I just wanted to, again, second your comment about high priority in congenital heart disease. We have both surgical-based interventions and catheter-based interventions. There are now registries tracking both the surgical-based procedures and the catheter-based procedures.

This registry, the C3PO and the CHARM methodology has been critical in forming another registry that is up and running that is available now in over 90 sites in the United States. And it is this methodology that is informing that registry, so that we can start to take this to a national level. So, this has been a critical piece of information for us and it something that we would expect, with broader use, we will see even greater variation in the outcome results, whether it is with the adverse events.

So, this is a high priority.

CO-CHAIR GEORGE: Liz? Any other comments on priority?

MEMBER TING: Yes, this is Henry. I raised my hand on the web links. I don't know if it was seen. Just a quick question. Is the denominator all children who go to the cath lab for diagnostic and therapeutic procedure or are those two separated between a diagnostic catheterization versus a therapeutic catheterization?

DR. BERGERSEN: That is correct, all patients who go, including those with diagnostic and interventional caths are included.

For diagnostic catheterization procedures, in addition to the adjustment for age within the model for CHARM, diagnostic catheterizations were stratified by age group with a higher rate of adverse events observed and expected among younger infants, as compared to the older children.

So, they are in three different groups, based on age groups.

MEMBER TING: So, do you stratify on the therapeutic procedures as well? Because I

could imagine those could be very, very different and much higher rates of complications, depending on what the device is or specific therapies are done.

DR. BERGERSEN: So, the CHARM risk adjustment methodology includes age as a patient factor that is adjusted for in the final model, after accounting for an independent of the procedure type risk group. So, yes.

CO-CHAIR KOTTKE: In fact, the type of procedure is in the model. She said yes, Henry.

CO-CHAIR GEORGE: Any other comments on priority? All right, we will vote.

MS. LUONG: Polling opens for high priority; 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

High priority passes with 63 percent voting high and 38 percent voting for moderate.

CO-CHAIR KOTTKE: Scientific acceptability, numerator/denominator exclusions. Numerator number of diagnostic and interventional cardiac caths for children under 18 years of age

resulting in clinically important adverse events performed by an institution performing at least 50 cases per year in pediatric patients under 18 years of age. The denominator is the number of diagnostic and interventional cardiac cath cases for children less than 18 years of age performed by institutions performing at least 50 cases per year in that pediatric population.

Exclusions, primary electrophysiology cases, ablation cases, pericardiocentesis only, thoracentesis only. The data source is electronic clinical data, electronic registry, paper, medical records.

I didn't have any concerns there. It looks like Liz.

MEMBER DELONG: I just wonder about standardization of data elements because if you are using different kinds of sources like the EHR and a clinical registry, et cetera, is that harmonized?

DR. BERGERSEN: That is a great point. And there has been a lot of effort over the past

decade to develop common nomenclature in the field.

The procedure types, as defined in the procedure type risk groups, although you may use different nomenclature within your own reporting systems, there is clear one to one mapping. So, while the C3PO Registry uses a nomenclature for procedure type risk groups, when we attempted to map that to data elements within the IMPACT Registry, which uses the IPCC nomenclature, we were able to do that in a reliable fashion.

MEMBER DELONG: Yes, I guess I am more concerned about the outcomes, the complications. I mean one of them is monitoring. Is that well-defined?

DR. BERGERSEN: That is a good point. Within the registry, which I presented you the data that has been tested, as I stated earlier, the adverse events are further classified, according to severity.

So to give you an example which would illustrate your point, an arrhythmia, just saying

an arrhythmia would be an event that would be included. An arrhythmia that is self-terminating would not meet the definitions and would be of a low severity. However, one that required medication to terminate would be in a severity level 3. And if you required cardio version, it would be a severity level 4.

We have tried to be as clear as possible with our definitions.

MEMBER DELONG: So, my question is really would those all be coded the same way across different databases? Would you pick up the arrhythmia that required medication similarly?

DR. BERGERSEN: Thus far, it has been an abstraction from medical records and there has been limited testing in other data sets. So, I can't answer your question sufficiently.

CO-CHAIR KOTTKE: Okay. Are there some questions about data elements? Are they are defined? Developers compared information recorded in the database with medical record.

This is also considered validity and passes reliability by NQF requirements.

Data element testing was done using data abstractor from the EHRs and paper records entered into the database registry. A sample of 3,359 pediatric patients from 11 pediatric hospitals with a total of 784 cases were examined. No information about the types of facilities, where they were, size, et cetera, or patient included in testing.

So, it is unclear whether the testing sample represents the variety of entities whose performance will be assessed by this measure.

The results of the data element validity testing indicate that 85 percent of the 149 adverse events, including the medical record, were captured in the registry. All interventions performed were recorded correctly.

The developer states that all major adverse events were appropriately captured but that two events related to sedation and airway management and late identification of a growing

fistula requiring surgical repair were not recorded.

So, I think reliability is moderate.

CO-CHAIR GEORGE: Other comments on reliability? All right, we will vote on the reliability.

MS. LUONG: Polling opens now for reliability: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Reliability passes with 13 percent voting high, 68 percent voting moderate, 13 percent voting low, and 6 percent voting insufficient.

CO-CHAIR KOTTKE: Validity, the specifications do align with the evidence. The validity was tested both at the data element level and the measure level.

Let's see. The risk adjustment model appeared to be appropriate to me but I am not a statistician. Specifications, this measure was risk adjusted using a logistic regression model with three risk factors. The calculated score is

the ratio of observed expected rates of clinically important adverse events occurring during or following cardiac cath.

The developers do not provide information on how the risk model is developed. The C statistic reported for the risk adjustment model is 0.72. This model discrimination statistic represents the proportion of all possible pairs with different observed outcomes for which the model correctly predicts a higher probability of observations with the event outcomes than the probability for non-events.

I don't think I want to read all this. Anyway, when applied to the impact data set, the AHA abstract reported a C state of 0.70 as well.

So, I think it is valid.

CO-CHAIR GEORGE: Any threats?

CO-CHAIR KOTTKE: No particular threats.

CO-CHAIR GEORGE: Any comments on validity? All right, we will vote on validity.

MS. LUONG: The poll opens for

validity voting: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Validity passes with 25 percent voting high, 69 voting moderate, and 6 percent voting insufficient.

CO-CHAIR KOTTKE: Feasibility. It appears to have -- from the databases in the hospitals, I would say that it is feasible.

CO-CHAIR GEORGE: Any comments on feasibility? All right, we will vote.

MS. LUONG: Voting starts for feasibility: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient.

Feasibility passes with 38 percent voting high, 56 percent voting moderate, and 6 percent voting insufficient.

CO-CHAIR KOTTKE: Usability and use. The measure is currently used in the congenital cardiac catheterization project on outcomes quality improvement, C3PO QI, program for internal quality improvement. Public reporting is planned.

Data on improvement over time using this measure is not provided, although the developer states that progress is tracked for the participating institutions and reports are available on demand.

According to the developers, the vulnerability of the measure is potential lack of reporting adverse events. However, they note previous audit results have found a 92 percent event capture rate among high severity clinically important adverse events.

MEMBER DELONG: So, I have a question and a comment. The question is I am really confused because there have been a number of these categories that have had an insufficient, except that there were no comments. So, I feel sort of ignorant in terms of why there was an insufficient, if there were no comments. I mean I would like to hear why they were insufficient.

But my other comment is there is no time frame given. So, I don't know how you can evaluate and compare, if it is not with a

consistent time frame.

DR. BERGERSEN: I'll address the time frame. In order to have a sufficient number of events, at least in institutions performing between 300 and 600 cases, we need a time frame of about a year. So, in implementation we report both locally and within the registry, rolling four quarter averages.

So, each quarter are rolling four quarter average of the past four quarters. It wasn't in the materials provided to you. It's a good question.

CO-CHAIR GEORGE: Any other comments on usability? All right, we will vote.

MS. LUONG: Polling starts now for usability and use: 1 for high, 2 for moderate, 3 for low, and 4 for insufficient information.

Usability and use passes with 31 percent voting high, 63 percent voting moderate, and 6 percent voting low.

CO-CHAIR GEORGE: So, any final comments before we vote on overall measure

approval? If not, we will vote.

MS. LUONG: Polling starts now for overall suitability for endorsement for Measure 0715: 1 for yes, and 2 for no.

For overall suitability for endorsement of Measure 0715, 94 percent voted yes and 6 percent voted no.

And that concludes the voting for today.

CO-CHAIR GEORGE: I assume there is no competing measures.

MS. ISIJOLA: Operator, can you open up the lines once more for member and public commenting?

OPERATOR: Yes, ma'am. At this time, to make a public comment, please press * then the number 1. At this time, there are no comments.

MS. ISIJOLA: Thank you.

Okay, well today ends today's meeting. We do have reservations for you at Mio, some of you are familiar with that, at 6:30. We can try

to push that back so you can refresh yourselves and be a little more comfortable.

We will reconvene tomorrow at 9:00 a.m. for the meeting and at 8:30 for the breakfast. But please let us know if you have any questions.

(Whereupon, the above-entitled matter went off the record at 5:22 p.m.)

H

I

J

W

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Cardiovascular Measure Endorsement Project 2014 Standing Committee

Before: NQF

Date: 12-04-2014

Place: Washington, D.C.

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

Court Reporter