NATIONAL QUALITY FORUM

+ + + + +

SURGERY STANDING COMMITTEE

+ + + + +

WEDNESDAY
FEBRUARY 13, 2019

The Standing Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, NW, Washington, D.C., at 8:30 a.m., Lee Fleisher and William Gunnar, Co-Chairs, presiding.

PRESENT:
LEE FLEISHER, MD, Co-Chair
WILLIAM GUNNAR, MD, JD, Co-Chair
ROBERT CIMA, MD, MA, Mayo Clinic*
RICHARD DUTTON, MD, MBA, United States Anesthesia Partners
TEMAYA EATMON, Patient Advocate
ELISABETH EREKSON, MD, MPH, FACOG, FACS, Dartmouth-Hitchcock
FREDERICK GROVER, MD, University of Colorado School of Medicine
JOHN HANDY, MD, American College of Chest Physicians*
MARK JARRETT, MD, MBA, North Shore-LIJ Health System*
AMY MOYER, The Alliance
CHRISTOPHER SAIGAL, MD, MPH, University of California Los Angeles
SALVATORE SCALI, MD, FACS, RPVI, University of Florida-Gainesville
ALLAN SIPERSTEIN, MD, Cleveland Clinic*
JOSHUA D. STEIN, MD, MS, University of Michigan
LARISSA TEMPLE, MD, Memorial Sloan-Kettering Cancer Center*
ADOLPH YATES, MD, University of Pittsburgh Medical Center

NQF STAFF:

ELISA MUNTHALI, Senior Vice President
MELISSA MARINELARENA, Senior Director
KATHRYN GOODWIN, Senior Project Manager
CHRISTY SKIPPER, Senior Project Manager
VAISHNAVI KOSURI, MPH, Project Analyst

ALSO PRESENT:

MARK ANTMAN, Society of Thoracic Surgeons
VINAY BADHWAR, MD, FACS, FACC, West Virginia University

GAETANO PAONE, MD, Society of Thoracic Surgeons

SEAN O'BRIEN, Duke University*

DAVID SHAHIAN, MD, Observer

*** present by teleconference**

CONTENTS

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

(8:30 a.m.)

MS. SKIPPER: All right. Good morning, everyone, and welcome to the Surgery Standing Committee's in-person meeting. Good morning to all of you in the room, those of you participating by the phone and, also, to our measure developers here.

Hi, my name is Christy Skipper, Senior Project Manager at CSAC. It's been a couple of years, since we had our last meeting.

So we're not going to keep you too long this morning. We do have 15 measures to -- I'm not going to talk too long, because we do have 15 measures to try to get through today.

So we'll do our standard introductions and disclosures of interest, briefly, briefly, review the portfolio, because you all know it, go through the overview of the evaluation process, jump into reviewing those measures.

And I want to note that, in addition to being logged in into the slide deck platform,

you also need to be logged into the Poll Everywhere link to vote later on today and then, we'll open the call for member and public comment and then, hear next steps.

And so again, my name is Christy Skipper and I'll just turn it to, I'll start with Elisa, to give, to say, introduce yourself and then, we'll go through the rest of the project team.

MS. MUNTHALI: Thank you, Christy. Good morning and welcome, everyone. My name is Elisa Munthali, I'm the Senior Vice President for Quality Measurements, and I wanted to thank you for being here, in person and on the phone.

MS. MARINELARENA: Good morning. My name is Melissa Marinelarena, I'm the Senior Director. Welcome, all of you. I know a couple of you in from other projects, but it's a pleasure meeting all of you, for the first time face-to-face those of you that are here. Thank you.

(Telephonic interference.)

MS. GOODWIN: Good morning and welcome. This is Katy Goodwin, Senior Project Manager.

MS. SKIPPER: Thanks, all. And so just a brief note, for those of you on the phone, please don't put the call on hold and I hear a little bit of feedback, so please try to mute your microphone.

And we do have a transcriber, court reporter here today, so as you speak, please say your name before you speak. And then,we'll also, for those of you on the phone, be monitoring the chat room, but you're encouraged to speak out. And, again, say your name before you speak, so that we can know who's speaking, when we go back and look at the transcript for this meeting.

Okay, so now I'll turn it over to Elisa to go through introductions and disclosures of interest.

MS. MUNTHALI: Thank you. And, before I do, I just, I think we missed one of our teammates, I wanted her to introduce herself, as

well.

MS. KOSURI: Hi, everyone, I'm Vaishnavi Kosuri, and I'll be helping out with voting today.

MS. MUNTHALI: Thank you, Vaish. So we're going to combine disclosures of interest with introductions and you, probably, remember, when you were named to the Committee, we asked you to fill out a pretty lengthy form that asked you a number of questions about your professional activities, as they're related to the Surgery Standing Committee.

And so what we're asking you today is to orally disclose anything that you provided us on that form. Before we go through the process, I just wanted to give you a couple of reminders.

You sit on this committee, as an individual, you do not represent your employer, or anyone, who may have nominated you for the Committee.

We're interested, not just in paid activities, as they're relevant to the work in

front of you, but also those that are unpaid. And this is, probably, the most important reminder, just because you disclose does not mean you have a conflict of interest. We go through this --

(Telephonic interference.)

MS. MUNTHALI: Sorry about that. We go through this process in the spirit of openness and transparency, and so we do have a number of people on the phone. We'll start with the disclosures of, disclosures of interest --

(Telephonic interference.)

MS. MUNTHALI: -- in the room, first. We'll start with your co-chairs. We'll ask them to introduce themselves. Right now, you only have one. I understand, Lee's on his way.

Introduce yourself. Let us know who you're with and let us know, if you have anything to disclose. We'll go through everyone here in the room and then, on the phone, I'll call out your name and let us know. So, Bill.

CO-CHAIR GUNNAR: William Gunnar,

Veterans Health Administration, and I have no conflict of interest. I'd also like to just say, take a moment, because -- to welcome everyone, those who are here, who, who had the intestinal and constitutional fortitude to actually travel to D.C., thank you.

And, two things. One, the Josh Stein, this is your first in-person meeting. You've been on the Committee for a while, so welcome, in person, it's nice to put a face to a name.

And I'd also like to welcome TeMaya Eatmon, who's our new patient representative for the Committee, so thank you for your commitment. So.

MS. MUNTHALI: Thank you for that. TeMaya, welcome, and your turn.

MEMBER EATMON: TeMaya Eatmon and I'm a Patient Advocate for several cancer organizations and I don't have any conflicts of interest.

MEMBER MOYER: Amy Moyer, I'm the Manager of Value Measurement for the Alliance,

we're a healthcare purchasing cooperative, a non-profit, and I have no disclosures of interest.

MEMBER YATES: Adolph Yates, Orthopedic Surgeon, from the University of Pittsburgh Medical Center, and I have no conflict of interest relevant to any of the measures that we'll be looking at today.

Although, I am involved in orthopedic measures and a whole bunch of committees and I, unfortunately, I have no commercial conflicts of interest, which is good news and bad news. And I will pass it on to Richard.

MEMBER DUTTON: Rick Dutton. I'm an anesthesiologist. I practice in Dallas. I'm a Chief Quality Officer for U.S. Anesthesia Partners, which is a large national private practice.

I have no conflicts of interest with any of the measures, today, although, like Adolph, I, I participate on lots of committees and technical expert for measure development.

MEMBER SAIGAL: Chris Saigal, a urologist from UCLA, I have no disclosures.

CO-CHAIR GUNNAR: I just wanted to point out that Chris came from L.A.

(Laughter.)

(Off the record comments.)

MEMBER SCALI: Good morning, my name is Sal Scali, I'm a vascular surgeon at the University of Florida, in Gainesville, as well as the Veterans Health Administration.

I have no financial conflicts and I do act as the Chair of the ER Quality Committee for the Society for Vascular Surgery's Vascular Quality Initiative.

MEMBER EREKSON: Hi. Elisabeth Erekson. I work at Dartmouth-Hitchcock. I'm a urogynecologist and the interim chair for the Department of OB/GYN there. As discussed, I have no conflicts of interest with any of the measures today. I do sit on a number of committees and I'm the National Advisor for the OG's Quality Improvement Network.

CO-CHAIR GUNNAR: I just want to point out that Elisabeth drove an hour-and-a-half to Manchester, through the snowy Massachusetts countryside, and then took a plane here to D.C., just pointing that out.

MEMBER STEIN: Hi, everyone, I'm Joshua Stein. I'm an ophthalmologist and health services researcher at the University of Michigan. I have no relevant conflicts of interest.

MS. MUNTHALI: Thank you very much to everyone in the room. On the phone, I'll start with Robert.

MEMBER CIMA: This is Bob Cima, in Rochester, Minnesota, and our highways were closed yesterday, so I couldn't get out. I have no conflicts.

MS. MUNTHALI: Thank you, Bob. John Handy.

MEMBER HANDY: John Handy, Thoracic Surgeon, Portland, Oregon. No conflicts of interest.

MS. MUNTHALI: Thank you. Mark Jarrett.

MEMBER JARRETT: Hi, Mark Jarrett. I'm Chief Quality Officer at Northwell Health. I apologize not being there in person, but I have grandchildren and infectious diseases and I have no conflicts.

MS. MUNTHALI: Thank you, very much. Barb, I'm not sure if you're on the line?

(No audible response.)

MS. MUNTHALI: Barbara Levy?

(No audible response.)

MS. MUNTHALI: Okay. Keith Olsen?

(No audible response.)

MS. MUNTHALI: Okay. And, Allan?

MEMBER SIPERSTEIN: Hi, Allan Siperstein, at the Cleveland Clinic. It's always warm and sunny here, and I have no conflicts.

MS. MUNTHALI: Thank you very much. We do, also, have Fred Grover, who just joined us in the, in the room. Can you introduce yourself? Let us know, who you're with and, if you have any

conflicts.

MEMBER GROVER: Yes, I'm Fred Grover, STS, so that's, obviously, a conflict today.

CO-CHAIR GUNNAR: I just want to point out that, Fred flew from Denver to get here and he's not voting on any of the measures today.

(Telephonic interference.)

MEMBER GROVER: I can always weigh in, occasionally.

(Off the record comments.)

(Laughter.)

CO-CHAIR GUNNAR: But I was pointing out to those, who actually -- we, we have quite a few empty chairs, here, in the room. I was just pointing that out.

(Off the record comments.)

MS. MUNTHALI: Yes.

(Laughter.)

MS. MUNTHALI: Thank you to everyone, who made the effort to participate, in person and on --

(Telephonic interference.)

MS. MUNTHALI: -- the phone. I just wanted to --

(Telephonic interference.)

MS. MUNTHALI: -- let you know, what William was saying, Fred will be recused from voting and discussing, unless there is some specific questions that the Committee has for him to address. And, I think, somebody on the phone may be trying to chime in.

MEMBER TEMPLE: Yes, so this is Larissa Temple, I'm on the Committee, as well, and I, unfortunately, I'm still in Rochester, New York, and all of our airports are snowed in.

MS. MUNTHALI: Okay. Thank you, so much. Sorry, for missing you. Okay, so before I turn the meeting over to Christy and colleagues, just wanted to remind you that, at any time, if you remember that you have a conflict, we want you to speak up.

You can do so in real-time, or you can get in touch with any one of us on the NQF staff, or your co-chairs. Likewise, if you believe that

one of your colleagues is acting in a biased manner, we want you to speak up. Thank you.

(Off the record comments.)

MS. SKIPPER: Okay, thank you, Elisa. So now, we'll just jump into a brief review of the Surgery Portfolio. As you all know, most of these measures are inpatient and the level of analysis is, mainly, at the group practice, or individual, level.

And the majority of the measures are within, fall within cardiac surgery, vascular surgery, and we have a great number of outcome measures, within a portfolio, and a handful of process structure and composite.

Two of our measures that you all will be reviewing today, were sent to the Scientific Methods Panel, they were deemed complex. And this panel consists of individuals with methodologic expertise, to help ensure a higher level of evaluation of the scientific acceptability of complex measures.

Okay, now, I'm just turning to Katy,

for roles of the Standing Committee and so on.

MS. GOODWIN: Thanks, Christy. And, most of you have been through this, quite a few times, now, but we wanted to remind you of the process and procedure for today's meeting.

As a reminder, as Standing Committee Members, you are acting, as a proxy, for the NQF Multi-Stakeholder Membership and, the main charge for today is to evaluate each measure against each measure evaluation criterion.

During the discussion, we ask that you indicate your rationale for the rating and the extent to which each criterion is met. This is to ensure that we adequately capture your discussion in our report.

You will be making endorsement recommendations to the NQF Membership and, as a reminder, NQF Members have an opportunity to submit comments on your recommendations and, also, express their support, or non-support, of the measures.

Some ground rules for today's meeting.

Please attend the meeting, at all times. If you do need to step away, please, let us know, for those of you on the phone, you can let us know using the chat function. It's just really important that we know who is present and ready to vote, because we do need quorum to vote. If we lose quorum, we will not be able to vote.

Please announce your name, prior to speaking. We ask that you remain engaged, during the discussion, and keep your comments concise and focused and, of course, base your evaluation and recommendations on the measuring evaluation criteria and guidance.

So the, the process for today's measure discussion, the measure developers, we are fortunate to have them here with us, in person, today. They will be invited to provide a brief introduction of the measures set for three to five minutes.

After the measure developer introduce their measures, Committee Members, who are the lead discussants on the measures, will then begin

the discussion of the measures, in relation to the measure evaluation criteria.

The lead discussants will be providing a summary of the pre-meeting evaluation comments that Standing Committee Members have submitted to us. They will, also, highlight, or note, if needed, the preliminary rating by NQF Staff.

As a reminder, the NQF Staff rating is intended to be used, as a guide, to facilitate your discussion. You are still free to discuss openly and, and vote, as you wish.

So we are using a new voting platform today. We no longer have those fun remote clickers that you've used in the past, so we're using a new platform called -- oh, can you go back one?

(Off-microphone comments.)

MS. GOODWIN: That's fine. It's called Poll Everywhere, and we did send a link on how you can access Poll Everywhere and you will need to be on that link, open that in a separate tab, on your computer, in order to vote.

We'll walk through that again when we get closer to our first vote, so if you don't have that Email open, or if you don't have that Email, just let us know and we'll be sure to forward you the link.

As for the discussion, we will be starting with the first criteria, which is important to measure and report. As a reminder, this is a must-pass-criteria, which means, if the, the measure does not pass, either, evidence, or gap, the discussion will stop there.

We'll then move on to Scientific Acceptability. Under that criteria are two sub-criteria reliability and validity, both, of which, are, must pass, as well. We also have a couple of composite measures, and so we'll be voting on the quality construct under Scientific Acceptability.

We'll then move on to feasibility, followed by use, which is, now, a must-pass-criteria for maintenance measures, and we do have several maintenance measures that are

under review today.

We'll then talk, discuss, and vote on usability and, finally, the overall recommendation for endorsement. Also, as a reminder, you will be voting on the measures, as they are before you, today.

If you have recommendations for the developers, they are present, they are here. We will also note that in the report, but your recommendations and, and voting is to be on the measures, as they are before you.

A reminder on achieving consensus, quorum is 66 percent of the Committee, so for this committee that is 15 voting members. We do have quorum present, on the phone and in the room, so we will be able to vote on the measures.

In order for a measure to pass, or be recommended, that requires greater than 60 percent of the yes votes of the quorum. If the vote falls in between 40 and 60 percent that is consensus not reached, or the gray zone, as we call it.

This means that this type of measure, or recommendation, would move forward to the public and NQF Member comment and the Committee will re-vote and re-discuss the measure, during the post-comment call, which is scheduled in May.

(Off-microphone comments.)

MS. GOODWIN: The measure does not pass, or is not recommended, if it is less than 40 percent yes votes of the quorum. Any questions on the process, or voting, for today's meeting?

CO-CHAIR GUNNAR: Rick.

MEMBER DUTTON: Yes, hi. Richard Dutton. Dumb question. We filled out surveys on a lot of these and put a lot of comments in, in advance, is there any way to access those? Because, I don't necessarily remember what I wrote, but I might be smarter than I am now.

MS. SKIPPER: Yes, and I can send the workbook to you all. I thought I had, but all your comments have been incorporated and they're going to be in the light pink boxes on the --

MEMBER DUTTON: In the measures?

(Simultaneous speaking.)

MS. SKIPPER: Yes.

MEMBER DUTTON: Oh, okay, good.

MS. SKIPPER: Yes.

MEMBER DUTTON: That's fine, I can find them then.

MS. SKIPPER: Okay. And so, Lee joined us. I just wanted to welcome our other co-chair, here, Lee, to our meeting this morning.

CO-CHAIR FLEISHER: Sure, Lee Fleisher. Thank you. Train just got in. No disclosures. Other than I'm on the Board, I guess. Thank you.

CO-CHAIR GUNNAR: So and, and the other, just as an overview, thanks for these -- many of us are very familiar with the process, but it's changed a little bit.

I think, expanding some understanding of the scientific methodology would be helpful. The second is, all of these measures, most of them, were, except for the composite, which we

endorsed in 2014, all of these have been, this is the second round, we, we endorsed these in 2007.

And so can you explain how the use of criteria, now, because many of these measures are part of a composite measure, but they're not independently publically reported, so can we get some explanation, or help, on clarifying how we should interpret the use portion?

MS. MARINELARENA: Sure. So use -- the criteria, itself, has not changed. For use we require that a measure be in an accountability program, within three years of endorsement and publically reported, within six years.

The difference is that, now, use is a must-pass-criterion for maintenance measures. So technically, the measure, if it doesn't meet that criteria, a maintenance measure, then it should fail and then it would not move forward to recommendation for overall endorsement.

So we have, we have seen, in the past, some measures that are beyond six years and not publically reported. There have been reasons for

it and the Committee can decide, if that's an acceptable reason, some of them have changed measure developers and they're in the process of getting it, it publically reported. So the, the Standing Committee has said that the measure developer is making good faith effort to get it publically reported and in an accountability program. So that's something to consider.

CO-CHAIR GUNNAR: Yes.

MS. MARINELARENA: But, we did get, include, you know, initial endorsement, re-endorsement, and they are accepted to be, again, an accountability program and publically reported are loosely defined, but when we get to that part, we can, we can have that discussion.

CO-CHAIR GUNNAR: So, so help me just a follow-on question, help me understand. So if a, if an individual measure is part of a composite, but not independently has been publically reported, or, or fails in, in and of itself, to meet the use, how do we, that, does it, how do they maintain it, within the

composite?

MS. MARINELARENA: A measure does not -- part of a composite does not need to be NQF-endorsed.

(Off-microphone comments.)

MS. MARINELARENA: A part of a composite, a measure that is part of a composite does not have to be NQF-endorsed.

(Off-microphone comments.)

MS. MARINELARENA: Yes.

CO-CHAIR GUNNAR: So we'll go with A.J., then, then, Reg.

MEMBER YATES: Well just, just, just as a point of order, the vast majority of the measures we're talking about today are part of composites, but they're not publically reported.

MS. MARINELARENA: Correct.

MEMBER YATES: We could make a vote, right now, as to whether that we're going to accept that definition of being reported and go home five minutes from now, for about -- because that should be the first question, if it hasn't

been publically reported.

And we're going through this exercise to endorse something that we're going to turn around at the very end and say, well we can't endorse it, because it hasn't been used in that, with that definition.

I would argue that it is used for internal quality metrics and, and it is reported back to the surgeons and they find out, where they are in life, in terms of, their abilities and their skill sets and how they're taking care of patients.

But, I, I read this and, and this is troublesome, because I don't understand why these measures are up for re-endorsement, at all, if they haven't been publically reported?

MS. SKIPPER: Elisa.

MS. MUNTHALI: So that that's a conversation we want the Committee to have and, to Melissa's point, it is up to the Committee's discretion to talk about, you know, the degrees to which, it meets accountability and public

reporting.

So we are not -- we don't define what that is, but because you're experts in surgical care and you bring different perspectives, I believe, you'll be able to land in a place -- now we, we don't only ask that the measures be publically reported, we also ask for a plan, if they're not.

So that could be a way forward, if the developer can articulate a plan to get these measures in an accountability program that's acceptable to us and would meet the criteria, as well.

MEMBER JARRETT: This is Mark Jarrett, on the phone. I, I, kind of, concur with the last speaker, you know, I'm all for public reporting and transparency and accountability, don't get me wrong.

However, I think having valid and reliable and scientifically-proven measures that are used internally, across the country, at local sites, is also critical, because the reality is,

we measure things, not so we can publically report them, we measure them so we can, actually, get better at what we do.

So I think that, although, accountability is critical, I, personally, kind of, tend, in my world, to believe that, you know, not everything has to be publically reported, again it can be very valuable.

(Telephonic interference.)

CO-CHAIR FLEISHER: So Elisa, I'm going to push in that, these rules are defined by the CSAC, correct?

MS. MUNTHALI: They're defined by NQF and upheld and, by the CSAC, so we've require accountability and public reporting --

CO-CHAIR FLEISHER: Right.

MS. MUNTHALI: -- not defined --

CO-CHAIR FLEISHER: Not defined. So being that I'm now taking a constitutional law class --

MS. MUNTHALI: Oh gosh, okay.

CO-CHAIR FLEISHER: -- with the Dean

at our law school, it's not in the Constitution that, how it's defined, as public reporting, so we could push back and say, based upon the fact that they're part of composites, that we feel it's important and CSAC can, either, uphold, or, or vote against us, in that way, is that an acceptable --

MS. MUNTHALI: That is an acceptable --

CO-CHAIR FLEISHER: -- decisionable from this group? So we, we could, actually, inform the higher courts, so to speak, that we feel that it meets a criteria, which I'm not saying we should, or shouldn't, but that is an acceptable, from your standpoint, approach to that?

MS. MUNTHALI: Yes, because what Melissa said, I just want to underscore that. In the past, components were required to be NQF-endorsed.

CO-CHAIR FLEISHER: Right.

MS. MUNTHALI: A few years ago, we

changed that, so the components have a composite do not need to be NQF-endorsed, so that could --

CO-CHAIR FLEISHER: And, in particularly, since the sites gets the individual scores back, the surgeons, we could help inform that STS has maybe has approached us from a higher perspective and, and therefore, we don't want to de-endorse something that's of such quality and such importance. So I'll -- I think that's --

CO-CHAIR GUNNAR: But, but as a point of order, normally, is the way this would work, based on NQF rules, if we get to use, I mean, we will say, it does not pass, because it doesn't meet the criteria, NQF criteria for use, but then you're supposed to stop. So what happens here?

MS. MARINELARENA: So you would have the conversation and then you would vote, if you, you know, if it passes, we need to capture a rationale, or justification, from the Committee, as to why you think, no it does meet.

And again, public reporting for us

doesn't mean that it has to be on, you know, hospitalgov.com that, that's not publically reporting.

It's, again, a loose definition. It means that you were -- the public can access it, in some way, we just need to be able to capture a justification for you passing it.

But it does, they have to meet all the other criteria. They have, there has to be evidence. There has to be a gap and they have to be reliable and valid, and then we can talk about use. Yes.

MEMBER STEIN: I have a separate, or a different question. I'm relatively new to the Committee. If, if we have suggestions of ways a developer can enhance their measures, if we pass, if we vote to pass a measure, as it is, is there any incentive for them to listen to our suggestions, or do we have to not pass the measure, so they can listen to our suggestions?

MS. MUNTHALI: So I think any suggestions and recommendations we will be taking

down, as part of the commentary, in the reports, and, you know, we, the developer will have that input and, hopefully, they will use it to, in their improvement, as they conduct maintenance. So don't feel -- don't withhold from giving recommendations that's part of the process.

MEMBER STEIN: But, but they're --

MS. MUNTHALI: But it doesn't go --

MEMBER STEIN: -- but their incentive to act on our recommendations is less, if we've been, if we've already endorsed the measure, correct?

(Simultaneous speaking.)

MS. MUNTHALI: So we wouldn't have like a condition recommendation, so we don't do that, but it could be part of the commentary, as you're discussing, you know, perhaps, look at this, or look at that, but not as a condition of endorsement.

CO-CHAIR GUNNAR: Rick.

MEMBER DUTTON: Yes I, I appreciate Lee's effort to find language that will make

everybody happy, but I really think we should tackle this one more head on.

The vast majority of cardiac surgery hospitals and cardiac surgeons in the country participate in SDS and we know that these measures are used, internally, by all of those people, to improve the quality of care.

Now, it seems to me, the, the spirit of what we're trying to achieve here, and I would consider that acceptable use for all of these, but if that's not going to be case, then I certainly agree with A.J., we should go home.

CO-CHAIR GUNNAR: A.J.

MEMBER YATES: Yes, I was just going to -- A.J. Yates. I was just going to follow up on the fact that, if NQF has made it no longer necessary for composite measure to have its components endorsed, it runs the risk of outside groups taking their composite measures and breaking them down, but then saying that those individual parts carry the weight of endorsement, because they were part of something that was

endorsed.

And so as blowback to the committee that decided on all this, I think, is reasonable that we, again, give a circuit court opinion to satisfy --

(Off-microphone comments.)

MEMBER YATES: -- so to satisfy Counselor Fleisher's recommendation, or Judge Fleisher. And -- but I would argue that -- I would agree 100 percent with Rick.

And I think that the, the value of the composites is made stronger by making sure that those parts that could be broken out, someday, are endorsed and ready to be broken out, but it doesn't do, it, it doesn't lower its value, in terms of, its use.

And I would, I would suggest that we make, if possible, there would be extra schedulary (sic), but maybe we should vote on that concept, so that we don't discuss this and argue it on every single measure, but we can, certainly, open it up for consensus, but maybe,

we refer to this, as Fleisher Rule A, or, or Ruling A, and that might make things go smoother.

CO-CHAIR GUNNAR: So I did pass my constitutional law class and --

MEMBER YATES: Good.

CO-CHAIR GUNNAR: And so -- and, you know, the rules are established and are, and are well-recognized up-front, when you present your measure for endorsement on the, an initial approval.

And the rules are now that, if it's not -- that particular measure is not publically reported within a period of time, then it fails the use. It no longer passes the -- that's the rule.

It doesn't require that that individual measure be, which is part of a composite measure, be endorsed individually to be a member, or a part of the composite.

The composite's passed, I mean, fundamentally, it -- the issue is, in this -- what we're, what we're faced with, for the first

time, as a group, is this, the, the fact that an individual measure was not publically reported in the period of time that was clear at, at initiation.

CO-CHAIR FLEISHER: But I do want to say and, and, sorry, is that, my recollection of the discussion was, this was not a case that, really, was thought about that, from the CSAC perspective, or from when this was developed and, like Elisa said, in that it's used, but only as part of a composite.

I mean, it really was trying to move the field forward to say, let's not endorse measures that never go into public reporting. Nobody thought that this very narrow use case of a measure that's part of a composite that's public reporting that's why I'm parsing it to say pushing back --

CO-CHAIR GUNNAR: Arguing the Fourteenth Amendment --

CO-CHAIR FLEISHER: Right, but --

CO-CHAIR GUNNAR: -- all right.

CO-CHAIR FLEISHER: -- but, pushing back --

CO-CHAIR GUNNAR: Also --

CO-CHAIR FLEISHER: -- on --

CO-CHAIR GUNNAR: -- he's also being a regionalist and then, he's saying that the initial intent of the founders was --

CO-CHAIR FLEISHER: Absolutely.

CO-CHAIR GUNNAR: -- this was not their original intent.

(Simultaneous speaking.)

CO-CHAIR FLEISHER: This was not the original intent that I remember, but Elisa --

MS. MUNTHALI: That is correct. I think it was a broader assessment of the entire portfolio and I don't think, you know, they were looking at the components of composites and, whether or not the, the actual composite was in, you know, publically reported in an accountability program and then, whether or not those components were. That did not come up in discussion.

But, this is the first time that we're having this discussion and this is why this dialog is important and how you look at public reporting will be reflected and should be reflected in your vote and go along with your recommendation.

CO-CHAIR GUNNAR: So to be clear, there's two paths here. The first path is we kick the can upstairs. We endorse -- we just go ahead and override the NQF rules on use, pass it on that basis, conceptually, send it upstairs and let the powers that be manage this individually. Alternatively, we follow the rules, we don't pass those individual measures that, that aren't publically reported, allow that the, the developers to respond to that.

It will then be, come back to this committee, at which point, a decision would be made, collectively, about those and a recommendation from the Committee for decision.

So one way, Path 1, is to send it forward and eliminate the opinion of the

Committee in front of CSAC. The second path, actually, fundamentally, it brings it back to us, in my opinion, and it allows us to weigh in, as a committee. Does that make sense?

(No audible response.)

CO-CHAIR GUNNAR: Am I -- am I tracking and are you tracking?

MEMBER YATES: There's also a third path, which is to go through the process on each, endorse them to the point of usability and then, put them in limbo, to go back to Path 1.

CO-CHAIR GUNNAR: That's -- yes. That's Option 2. That, that just is, you know, we can call it 2A and 2B, but it's, it's fundamentally the same, we get to weigh in, as a Committee, on that.

MEMBER YATES: But, if we are sending up endorsed, other than the fact they're not used, if we send it up as endorsed, except for that fact, we do the higher committee service by saying that the measure looks pretty good and would be endorsed otherwise if it weren't for

use.

CO-CHAIR GUNNAR: I hate being overruled and we sent it up endorsed, with the potential for being overruled.

MEMBER YATES: No, I'm not saying it would be endorsed, it would be endorsed, but -- well, on hold, for their opinion.

MS. MARINELARENA: No, it would be endorsed.

MEMBER YATES: You can't do that?

MS. MARINELARENA: So just to confuse you a little bit more, assuming the measure meets all of the criteria, you passed it on use. You say, you know, we think that it does meet the use criteria. You give it an overall recommendation for endorsement, then you have to look at the measures for related and competing.

We did pull them together. They're not competing, but they're related, because now you have these components, the individual measures that are part of the composites.

So we will ask you, to go to, through

the exercise and the questions are, do you need these additional measures, in addition to the composites?

Set of questions, you can say yes, or no, but we don't ever have the discussion about related and competing measures, until the measures have been endorsed, because that's a separate conversation.

If you say yes, we capture that, we need, we just need to have your discussion in the rationale, for having multiple measures and the measure goes forward.

CO-CHAIR FLEISHER: So just to be clear that, for the use case, and, Amy, I don't know, if you would, intermittently -- if we say that we believe that, being publically reported part of a composite measure is good enough to meet that use case, we can do that in this committee?

That's our definition of use, so we're -- and CSAC, in our report, we can say, this is our definition of use, why we passed it, they can

overrule us for our definition of use.

But it's our definition, and we make it very clear that if it's not part of a composite and it's just out there, we would have turned it down, our use, we've changed the use criteria slightly, or we interpreted the use criteria within our committee's purview?

MS. MUNTHALI: Right, and just to be clear, to the extent that the CSAC is overseeing our entire portfolio, we don't want them -- we want these cases -- there is oversight, so we don't want them to frequently overturn the expert committees, this is you, on surgical care.

And so if you were the experts, you understand the implementation of these measures, you have more knowledge of that, and you have justification, it would be challenging for the CSAC even to do that.

So we want to make sure the process goes forward. In case it doesn't, they are there, as an oversight body, to ensure that, you know, everything was considered and the process

and the criteria were applied appropriately.

MS. GOODWIN: This is Katy. We also have Robert Cima, on the phone, who'd like to make a comment.

MEMBER CIMA: Yes, I agree with what has been said, about the composite, but I still go back to Dr. Dutton's point is that, what, how do we define public reporting? I don't think we have the leeway of saying what public reporting is.

I mean, A, AHRQ has a very good statement about it and it says, it needs to be public reporting to be effective, reports need to provide consumers with transparent, timely information they can trust to help them inform conversations with their providers and payers, to help guide their healthcare decisions.

So if this information isn't available -- I mean, I understand internal quality control, but if it's not available to the public, in a way they understand and can access, in a timely and a transparent manner, then I'm not sure we're doing

our job.

MS. MARINELARENA: This is Melissa from NQF. So Christy is pulling up our definition of transparency, or public reporting. So here we say, transparency is the extent to which performance results about identifiable, accountable entities are disclosed and available outside of the organizations, or practices, whose performance is measured the capability to verify the performance results adds, adds, substantially, to transparency.

CO-CHAIR GUNNAR: I'd say that's pretty clear. So with that preamble, should we jump in with the first evaluation, then we'll just see how it goes?

MS. MARINELARENA: Let's do it.

CO-CHAIR GUNNAR: The first one is 0114. I don't have my sheet with me, or do you have it -- it's --

MS. MARINELARENA: I'm sorry, it's 2561.

CO-CHAIR GUNNAR: Oh.

MS. MARINELARENA: Here.

CO-CHAIR GUNNAR: Sorry, I didn't have it. So 2561, STS Aortic Valve Replacement Composite Score. And do we have -- Cliff is not, it's Cliff Cove (phonetic), Barbara Levy, and Robert Cima, but I don't think Cliff and Barbara are participating.

But Robert is, right? Robert, or Dr. Cima, are you prepared to present this all solo?

MEMBER CIMA: Yes, I can do this. This is an STS measure. We've gone through many, many of these --

CO-CHAIR GUNNAR: And my -- call to order -- we need to bring up STS Representatives and --

MEMBER CIMA: Okay.

CO-CHAIR GUNNAR: And they'll introduce the measure. My bad. Yes, this one is -- it won't address the issue until we get --

(Off-microphone comments.)

CO-CHAIR GUNNAR: So welcome. Whose -- oh, we have everyone coming up. I don't know

-- I can't move down.

MR. ANTMAN: Okay.

CO-CHAIR GUNNAR: So, welcome. If you'll introduce yourselves, for the record? Is there anything else we need from -- as far as, the introduction is concerned?

MS. MARINELARENA: They can, probably -- if you, you can, probably, introduce, both, 2561 and 2563, since they're similar.

MR. ANTMAN: Okay.

MS. MARINELARENA: Okay?

MR. ANTMAN: May I speak to the measures, in general, to begin with?

MS. MARINELARENA: Sure.

MR. ANTMAN: Okay. So good morning, everyone. I'm Mark Antman. I'm Senior Manager for Quality Metrics and Initiatives, at the STS. I'm joined by Dr. Gae Paone and Dr. Vinay Badhwar, also, representing the STS.

Dr. Dave Shahian is, also, attending today. He's acting, as an observer. I think, we also have individuals, on the phone, Dr. Sean

O'Brien, Dr. Maria Grau-Sepulveda, may be on the phone, to speak to some technical aspects of the measures.

I want to thank the Committee and, and NQF Staff, for allowing us to have 15 measures reviewed today. I know that's beyond your usual limit of measures, per meeting, so we appreciate that.

These measures are calculated reported back to participating surgical groups and publically reported, in our perspective, based on clinical data submitted to the STS Adult Cardiac Surgery Database.

The database was launched in 1989 and contains more than 6.5 million cardiac surgery procedure records and represents the participation of over 90 percent -- probably well-over 90 percent -- of the surgical groups performing cardiac surgeon in the U.S., and the proportion of those surgical groups that participate in voluntary public reporting is now up to just about 70 percent.

The 15 measures under discussion today include seven that are component measures of the STS CABG composite. That CABG composite in 2010 became the first, the STS's first voluntarily publically reported composite measure, and the CABG composite will actually be, be, be submitted to NQF for your review in the fall of this year.

Six, of these seven CABG composite measures, excuse me, component measures, are outcome measures. One, the anti-lipid treatment measure is a process measure. It's the only process measure of all 15, under review today.

Of the eight measures that are not for isolated CABG, six are operative mortality measures for valve, or valve plus CABG operations. And each of these, as has already been discussed by the Committee, is part of one of the STS valve, or valve plus CABG composite measures, all of which are voluntarily publically reported.

And two of those valve, or valve plus CABG composites, specifically the AVR and the AVR

plus CABG composite, are the remaining two of our 15 and those are the first measures that are scheduled for review today.

So that's my overview. I think given the preceding discussion related to public reporting, Dr. Badhwar and Dr. Paone and I will be happy to speak to what we see, as the, how each individual component measure is public, publically reported, as each of them come up for discussion.

So with respect to the first two measures under review, 2561 and 2563, these are the AVR composite score and the AVR plus CABG composite score.

Each of these have a morbidity and a mortality domain. Each of those domains, as well as the composite score, are the data for each of those measures, for each of those, each of those domains and the composite score as a whole, is reported back to all participating surgical groups for their review and for quality improvement purposes.

Each of these composites as a whole is, as I have said, publically reported, and there is a publically reported domain score for, both, the mortality domain and the morbidity domain.

So in fact, those, those mortality component measures that are under review today for these two composites, they are, in fact, publically reported separately, on the STS Public Reporting site.

There is a separate mortality score and a separate mortality star rating, as well as a separate morbidity score and a morbidity star rating, for each of those domains. I hope that helps. We'll be happy to answer any questions.

CO-CHAIR FLEISHER: Just basically if I could ask, as you discuss your own vision of how it's publically reported, if we have questions, are you open to developing any plan for additional public reporting?

If that's an option, that would be useful for discussion today, because that is part

of the conditions for approval, correct? That, if there's a plan that would satisfy the intent, that that would also be helpful and that would allow us to go forward?

MR. ANTMAN: A plan for the individual components be public, publically reported separately?

(Simultaneous speaking.)

CO-CHAIR FLEISHER: Yes.

MR. ANTMAN: I'll defer to my surgeon leaders.

DR. PAONE: I'm not sure that I can state that there's a plan specifically to publically report each of these measures. The composite was developed specifically because we felt it was a better measure of what a patient needed to know and would want to know, about what their risks for the surgery were going to be.

I think it's pretty complicated for a patient to look at this variety of outcomes and try to determine whether that's an important piece of information for them to have.

You know, our concept is what the patient wants to know is, am I going to have this operation and go home, without any major complications, and to their mind, I don't know that they can understand or recognize the difference between whether I've had a stroke, or renal failure, requiring a dialysis, or a sternal, a deep sternal wound infection.

So conceptually, this is why the composite was developed, in addition to the fact that, relatively speaking, the lower incident outcomes would have less validity over a longer term and the composite score is obviously a better way to present that data.

I can't say that we would not consider, and I would defer to others higher up in the society quite frankly, as to whether or not we would consider publically reporting these in the future, but again, I think the composite score was developed specifically to address the issues that I've mentioned.

CO-CHAIR GUNNAR: And for me, it just

is a, and, and, maybe, take this back -- I mean, is it important to you that in the individual morbidity measures of the composite are NQF-endorsed independent of the composite, when they actually form the basis of the composite?

And as you say, Gae, it would be -- does it really make, from a driving quality sense, having it in both places, so to speak, does that -- does it serve a mission, beyond what the composite does? I guess that's my fundamental question.

DR. PAONE: Yes, I would suggest, frankly, that having to have these individual outcome measures separately reported, for instance on our website, would, would do nothing more than just make it more busy and, and difficult to interpret.

Again, from the patient's standpoint, they don't really care if they get one of them, what's important to them is that they get none of them.

On a more practical level, though, I

think, first of all, and I can't tell you how many, but there are actually centers that will individually report this information on their own personal websites.

And I don't know how many that is, I know when I was at Henry Ford, we did. You could go to certain quality pages on our website and you would find many of these outcomes individually reported as the percent of stroke, percent of infection, those sorts of things.

And so I think you're correct. I think this would be, sort of, data, you know, overload, if we needed to do all this and put it all on the website.

The other part of it though is we do report them, individually, to the centers. And the centers do look at them individually and act on them when they get them.

And I think having these measures being continuing to be NQF-endorsed, furthers the significance of these measures and makes people take notice of them and pay more attention to

them, I suppose, than they might otherwise.

Although frankly I'd like to think that people pay attention to these, whether or not they're endorsed, but I do think it adds a validity to them that is beneficial.

DR. BADHWAR: Hi, I'm Vinay Badhwar, from West Virginia University. To just add, minimally I think one of the most important aspects of these NQF-endorsed measures, the subcomponents of the composite measure is to follow along with the Fleisher, Gunnar, Yates doctrine that we just listened to, which was very erudite and you're right on the money in terms of how you've interpreted it, I think it is an excellent conversation, before this began, is that the, any, or none, aspect of that, I think, is the most important, as, as Gae just -- Dr. Paone just outlined that, it's even more stringent if they have anything.

And so to go to the core mission of what we're all here for, in terms of that transparency of any type of morbidity is that,

what the family member or their, the patient, would be able to walk out of the hospital free of any bad outcome.

So having those subcomponents NQF-endorsed are, of course, of great value and, as Dr. Paone outlined, they are individually shared in the harvestry court by line item of those measures, with participant sites separately. So the question -- the short answer to that question is yes.

MEMBER CIMA: So just to -- a question to the developer. This is Bob Cima. This is one of my questions I had is, if you're reporting, if you're pulling out the -- I mean, the composite, as is, is one score.

But then I just heard that you're pulling out the separate components and reporting them separately, as two groups, mortality and morbidity, but really we endorse just the composite.

So I was wondering about the -- is it appropriate to report them separately? Shouldn't

they then be separate measures that have been endorsed separately, because you're basically picking and -- you're pulling stuff out of them and saying them separately.

I understand the rationale behind it, it makes sense, but it wasn't necessarily what we endorsed. We endorsed a single measure, but you're pulling it all apart and making it two measures and saying they're both NQF-endorsed.

I understand it's a nit-picky thing, but, really, the composite measure is the only thing that's endorsed, and it should be presented unadulterated.

DR. PAONE: I think I understand the question that you're asking and it's, again, it's an -- the purpose of adding the composite to the mortality was to sort of broaden its reliability and give it more weight.

And because there's a difference between a patient who survives with a major complication, and the patient who survives without one, we thought that that certainly

strengthens the benefit and value of the measure.

I'm not sure what the advantage I -- if I understood correctly, just to have an AVR, a separate AVR composite on morbidity and one for mortality, again, I think would be separating two things that don't need to be separated and should --

MEMBER CIMA: No, I appreciate that and I understand that, but what's -- but then you're doing it by reporting it separately. But what I'm saying is, we didn't endorse it to be reportedly separate, reported separately, but you just said, you are reporting it separately.

So then, if you are reporting it separately, then they should be evaluated and endorsed separately. You don't have the -- because of the methodology is designed, a single composite score was developed.

You should report the composite score, but from what I heard, you separate them out and report them as mortality with a star rating, and then morbidity as a star rating.

(Off-microphone comments.)

DR. PAONE: Go ahead.

DR. BADHWAR: So first of all, this is Vinay Badhwar. I think your question's very important and the process through which you are given this introspection of the importance and the scrutinizability and the use of each of those measures, we totally respect.

I think the source of your question is, how best to report them? For clarity, we are reporting these back to, or the STS is reporting these back to the participant site, not publically, each of these NQF measures there, by allowing the participant sites to create quality improvement, based on these endorsed measures.

The value, however, if we went into publically reporting each individual measure, it would most definitely get unruly, in terms of the public being able to actually interpret, for example, isolated AVR and renal failure, as opposed to isolated AVR and deep internal wound infection, hence the reason for the composite.

But to Dr. Fleisher's earlier question, if there's suggestions, of course, we are here to listen.

CO-CHAIR FLEISHER: So I actually would ask, and I'm on your site now, you've said two things that I find interesting.

One is you're saying the public doesn't understand it, but your sites have it. But if I were to go into your site and know, as a patient, that I can ask, I care about stroke and I can ask my physician, who has the data available that an individual with a stroke, you can find out how well they do in relation to the rest of the country, and that's available.

That to me gives me the next piece of information. I wouldn't know that from your site currently, am I correct?

In other words, you could be more transparent with the public with what's available to the individual sites.

But currently you've made a decision that the public --

CO-CHAIR GUNNAR: Yes, let me ask it in a different way, if I wouldn't mind, I'm just going to turn the question. In your star rating on the composite, would it be possible, to have, to be within confidence limits for mortality and four out of the five morbidity components, and yet have an increase in stroke that is buried in the computation to the star system, do you see where I'm going with this?

So, you know, it's one thing to say, you know, fundamentally, it's you have to be in line in all measures, all components of this composite to have a two-star rating, right? Or better?

One variance would get me into the one-star, right? If that's the case, then fundamentally you are reporting those substantively to the public, right? You're giving them the information they need. If not, then that's helpful to the Committee, as we look to evaluate use.

DR. SHAHIAN: Yes, I just -- well,

first of all, I wanted to clarify something that came up on the phone comment. We do, in fact, present an overall single composite score that encompasses both morbidity and mortality. That's there, so we're doing that, but we also go one step further and provide you with scores and star ratings for the individual components.

And I think Dr. Fleisher brings up a very important point. And we've actually been doing this, but we should make it more explicit.

If you look at things that I've written, you will see that I always say, ask your surgeon and ask your hospital about their results and they should have the report available, and if they won't discuss those things with you, then I'd think about another program.

I've said that publically. I've put it in publications, and I think what we need to do is just put that more explicitly on our Website and say for additional information ask your program for the drill-down information. Thank you.

MEMBER CIMA: Just to the comment you made, that goes to my question. Yes, you have reported the individual composite measure, element.

But, in doing so you, sort of, give the impression that those are all NQF-endorsed, because you pulled them out of the composite measure. That's my concern. Is it sort of, you know, not guilt by association, but, you know, endorsement by association?

I mean, we've improved the composite measure; that's what should be displayed. You're taking it a step further, but is it implying that those are -- that's not how the methodology was evaluated, it wasn't how we approved it in the past, and now you're reporting it separately, as giving the impression that it could be NQF-endorsed. That's the only thing I was asking, only --

DR. SHAHIAN: It certainly was not our intention to do that. We present the ADR, or ADR CABG, whatever composite score is, we present it

and we just happened to take it one step further. There's no implication that the individual components of the composite are NQF-endorsed; it's the overall score that's endorsed.

CO-CHAIR GUNNAR: But once again, the answer is, now that you have, it may actually be helpful to the Committee, if you're giving a star rating to the component, to the morbidity component.

The question I have, specifically, is, can I have, can I be above the confidence limits in any one particular morbidity, you know, in renal failure, in stroke, the component of that morbidity score, can can I be, can I have a high rate of morbidity of a single part of that and still be a two-star?

Or, if I'm elevated beyond expected in any one of those that I would actually be taken to a one-star rating in morbidity? Does that make sense? Am I making myself clear?

DR. SHAHIAN: Yes. The morbidity domain of the composite is treated in the

composite as a dichotomous measure. You either have -- you either are a winner or a loser in that particular domain, and it doesn't matter what the rate is, if you have renal failure, you fail; it's any or none for an individual patient.

So for an individual patient, it's just like death, an individual patient either, lives, or dies, and an individual patient, either, has one of those complications, or doesn't have that complication and it is treated in the calculation as a dichotomous event.

Now, is it possible that somebody could have, somebody could have a poor score on the overall morbidity domain and still have a two-star rating, it is and it's a very complex algorithm?

And Sean O'Brien, our Statistician, may want to speak to that, but the mortality domain is rated about four times the importance of the morbidity domain, just so you know that. So the mortality is graded much more importantly.

CO-CHAIR GUNNAR: But you publically

report on morbidity as an isolated -- you give a star rating that you split them apart, as Dr. Cima states, and that's what's on the website, and I appreciate that.

That's actually -- and it's helpful to the Committee to have the understanding that it's the entirety of those five morbidity domains that make up the score.

And if I'm reading back what you just said, it is possible to be really good in renal failure, but have a higher stroke rate than expected, and still be a two-star because the statistical treatment of that entire morbidity domain, all five, allows that to occur?

DR. SHAHIAN: That's exactly right, yes.

(Simultaneous speaking.)

CO-CHAIR GUNNAR: Okay. I mean, that's, that's just the way that it is --

DR. SHAHIAN: I believe that's theoretically possible.

(Simultaneous speaking.)

CO-CHAIR GUNNAR: Yes that, that's just the way it is, and it's helpful, very helpful background for the Committee. Who do we have -- oh, Amy, yes.

DR. PAONE: If I could just add that, on the website, the components of the composite score, the individual scores for those are not noted, but the components themselves are noted.

And to a point that was made earlier, that does potentially or at least theoretically provide an opportunity for the patient to then go to the surgeon and ask how their results are specific to each of those components, and the STS report to that institution provides that information for them to be able to provide to the patient.

MEMBER MOYER: So I've actually used this measure to help a friend get care for a family member in the past couple of months. It was great that the information was available. I hate not being able to answer people's questions about things like that.

The way it's reported on the STS site didn't actually provide meaningful differentiation for us in terms of the star score. We actually went to the kind of data that's on U.S. News and World Report. That was a little more helpful for them in terms of being able to evaluate.

And, I think it's, you know, it's great to say talk to your surgeon about this. My understanding of the reports that surgeons and facilities are allowed to share at least with us as a purchaser was not the star ratings, it's kind of a raw data and, for us, at least, it required a little bit of, you know, statistical analysis, to really understand and interpret that, that I think would be hard for the individual patient.

And, also, would question, okay, so I asked my surgeon, hey, look, they're like, I have 89 percent, well what the heck does that mean? Is that good? Is that bad? Is there someone else in your system who is better? Do you know

that?

In this case, they ended up going to a different hospital, in the same system and were really happy with it. But if you have to go to each individual surgeon to get that information or each individual facility, that's not really -- I don't think that's a reasonable expectation for a patient in terms of -- instead of just having it easily available in one spot.

CO-CHAIR GUNNAR: Yes, just before it, and this is a sort of clarifying point, just so people -- it's a question I have, which, which is, is that, the star ratings is the publically reported data is available for 67 percent of participants, and that can either be a group practice or a facility.

And Amy's point, an individual could actually be practicing at multiple sites, or could be listed once with a group practice. It's a little -- but 67 percent are publically reporting or are signed up for publically reporting.

Just to be clear, the data that underpins that star system is not just for those 67; it's for the 100 percent, correct? That, just a point of clarification.

DR. PAONE: So there's a few questions there to just address.

CO-CHAIR GUNNAR: Yes.

DR. PAONE: So, to address your comments, Dr. Gunnar, you're absolutely right in terms of how it's interpreted.

So the data's generated for all surgeons. However, this is a voluntary public reporting enterprise. So it's now 69.9 percent of all surgeons actually publically report for the adult cardiac component, more so for congenital and otherwise.

But so you're right that that star rating is available on the Website and through other side measures, but essentially on the STS Website for those that volunteer report, and they sign consent forms to participate.

Now, as it pertains to Amy Moyer's

question, about -- I think you're asking about the availability of subcomponents, were you asking -- just for point of clarity, were you asking for a star rating of that individual participant hospital overall, or were you asking for a more specific piece of information such as a morbidity domain?

MEMBER MOYER: I was asking or clarifying that if a patient were to ask either a facility or a surgeon group, or surgeon, for information, as we've been saying, that group isn't going to then give to that patient or that purchaser.

So purchaser, here's our star report and we can share that with you, is my understanding that's, kind of, proprietary information, but they can share the raw or the risk adjusted registry data, which is a little harder to interpret.

DR. PAONE: Oh. So I clarify to say that any participant site, again, a participant site is a hospital, which quite frankly now in

the current era is the vast, vast, vast majority of participants.

But as Dr. Gunnar had mentioned that, in the past, when there are large surgical groups that signed up, as a participant, that was actually defined as a participant.

So that, as you know, very well, with the employee models that exists now, the vast majority are hospitals, and so if that hospital is signed up for publically reporting, then that data is completely available, both on the website of the STS, it's searchable, both on the morbidity and mortality domains, so that answer is totally publically reported.

However, if they're not signed up for public reporting, then that data is not readily publically accessible, and yes you do have to ask them for their numbers. Does that answer your question?

DR. SHAHIAN: But -- could I make one point of clarification? We do not restrict what a hospital may share. They can share,

voluntarily, can share with a peer or a patient anything that they get and it's about a 200-page report.

It's all available. There's a few quirky little things about STS overall results that for technical reasons we ask them not to share, but in terms of their own results and benchmark results, they can share anything they want. It's not proprietary.

(Off-microphone comments.)

CO-CHAIR GUNNAR: Oh, sorry. A.J.

MEMBER YATES: But just point of clarification, it's site-specific information to this point in terms of public reporting, but not surgeon-specific.

And the second thing is, is the data that the hospital has access to that can be shared, if they choose to share it, does it have the breakdown by surgeon at this point that they can share?

DR. PAONE: So, as Dr. Badhwar mentioned a moment ago, the overwhelming majority

of the over 1,000 programs now, would represent institutions, or surgical programs, or hospitals, and not any individuals, or not even anymore any individual surgical practices, which was more common in the past.

And, as far as what a hospital can share, it can share anything it wants, it's their data. It's once they receive it, they can put anything they want on their websites, they can share it publically in any form they want.

The only restriction would be that they're not allowed to say, we're better. They can say, we're two stars or three stars, but they can't say we're better than so and so because of our data.

(Simultaneous speaking.)

MEMBER YATES: At the level of a consumer, though, asking the hospital for data, can they ask for, is it available through the database, by surgeon?

DR. PAONE: The surgeon-specific data is not reported by STS to the institution. Within

their database they would certainly have an ability to breakdown that data in any form they wanted, and certainly included in that would be surgeon-specific outcomes for any number of reports that they wish to present, but that is not provided by STS.

MEMBER YATES: Right, it would be out of their own individual --

DR. PAONE: That would be out of their own individual database.

(Simultaneous speaking.)

MEMBER YATES: Right, but it would not have gone through the parsing and the risk adjustment that the STS does before they report it back to the institution?

DR. PAONE: Not in a composite fashion, no.

MEMBER YATES: Or as an individual measure, does it come back surgeon-specific to the institution, having gone through the risk adjustment and parsing of the STS?

DR. PAONE: Not to the same extent,

no. But they can get O/E ratios for individual surgeons on the individual outcomes, if they choose to.

MEMBER YATES: Within their own database.

DR. PAONE: Within their own database.

MEMBER YATES: Right. So in that regard, it's an institution-specific result and, you know -- I'm just making sure, I'm not sure that I'm doing, I'm, I'm making this point, because I'm well-aware that it's site and not surgeon, but I'm saying this for the public record for a new member that's a consumer representative.

DR. BADHWAR: I think you ask a very important question, A.J., but to echo Dr. Paone's answer, the O/E ratios do provide in lieu of that a risk-adjusted aspect per surgeon. These sites can have that information, but it's operation-specific.

So as you know, a surgeon does an aortic valve of CABG, a mitral valve, a double

valve, so the overall, you know, holistic experience of that surgeon is not currently they don't -- they can't just generate that; they can break it by operation though.

So you're right. They do have risk-adjusted data by surgeon, but it has to be pulled by each of those sites.

CO-CHAIR FLEISHER: So we, we do have -- TeMaya, do you want to -- TeMaya and Melissa and Elisabeth.

MEMBER EATMON: Okay. I was just, once I've seen the website as a patient advocate and a patient actually, I'm on the website and I'm seeing that it has the composite score and then it had the mortality and the morbidity score.

I know that you spoke that from a patient perspective, we may not want all of that information, but as a young patient, I want to find all of the information out there that's possible.

And so I'm not really comprehending

the mortality versus the morbidity when I scroll down to the bottom and it says that the overall composite score represents the two domain scores in a single number.

So across the board, just looking at it from a patient, I see that the composite score's a two, the mortality score's a two, the morbidity score's a two, per se.

So, for me, I want to understand what would be my risk for the mortality, what would be the risk for the morbidity, what are certain things that I would actually see that would distinguish between morbidity and mortality?

And, granted, you don't want to be there in the first place as a patient. I don't want to be -- I literally don't want to be there with you in the first place, so saying that I can walk out with one or the other is irrelevant to me because I'm there in the first place.

So at that particular point, I want to be armed with as much as I possibly can, and I want to go to a place to find the information

versus WebMD or blogs, or going into other type of places.

So granted, yes, it may seem like it's a lot of information or overload, or we wouldn't understand, I promise you that when you get a diagnosis, you are researching, you are looking up stuff, you are learning stuff and asking your doctors -- your specialists questions that they never thought they had seen since med school probably, but you are really sitting there researching.

And so I, as a patient, I just am trying to understand how that cannot be part of the overall composite, because I'm looking at it that, this mortality, this morbidity, it equals to that composite score, and so I want to understand more of what each of those actually are.

DR. BADHWAR: Well, thank you for that impassioned and very important question. So for clarity, two stars is as expected in terms of definitions, in terms of how we -- this is, I'm

talking at the 30,000-foot level first.

Three stars is better than expected. One star is lower than expected. So two stars is actually a very good rating in terms of the Bayesian analysis of how we do this statistically. That's point one.

Point two is the subcomponents of the overall rating are available. And for example, in coronary bypass, it's the overall composite score, it's the absence of operative mortality, absence of major morbidity, and then some process measure uses.

And then that, together, statistically, gets combined with other measures to form the composite. So you do have availability. So say, for example, one program has three stars in mortality and two stars in morbidity, then you can, by interpretation, note that the absence of any or none, so absence of any morbidity is lower than the overall mortality.

I know that it can be confusing to a

patient to do that definitions, but that information or the inference of that information is available on the website.

Now, to Dr. Gunnar's earlier point, is the availability, on the website of, what's my stroke risk, what is my renal failure risk? That's not there because of the composite, so these are composites of composites, if that makes sense, at a high level.

DR. PAONE: Yes, you know, as someone who has spent many, many years being interested in database and quality outcomes and the like, I feel for your, you know, desire to have more information and yet difficulty for us in trying to provide that to you in a way that makes sense to you.

One of the things in the composite, in the database results, for instance, is when you look at the two-star rating for operative mortality, there's a score below it.

And so I'm looking at one now, and it says 97.7 is an absence of operative mortality,

so that means that the mortality adjusted was 2.3 percent for that, and so you could theoretically scroll through those hospitals that you're looking at or considering, and compare those numbers, and similarly for an absence of major morbidity.

But an additional point that I make with the patients is that reality and the reason the composites, particularly the morbidity composite, is an all or none phenomena.

What we often don't talk about is that mortality for an individual patient is an all or none phenomena, and so whether you -- and patients will often ask for a number.

And then any of the surgeons that are at, around this table and on the phone will know this, they'll ask, you know, what is my risk for this?

And if you tell them 2 percent, or 4 percent, or 10 percent -- sort of the conversation that I have learned to have over the years, is I add after that number is quoted, to

the best of my ability, but that for you is meaningless, because that's a number derived from looking at many, many patients statistically. You will have either a zero percent mortality or a 100 percent mortality.

And so these are things that are intended to, in a somewhat simplistic way, help you. The other thing that I would say, and it's not really intended to throw the onus back on the patient, because we are responsible for trying to provide as much information, as we can.

But to the point Dr. Shahian made earlier, this enables you to go with some information to talk to your surgeon, or to go through your program and see who you want to have do your surgery, with the ability to ask them some questions.

And I would say to you if a surgeon will not give you that information, or seems annoyed by your questions or by the fact that you have spent a great deal of time trying to find this information, then I would suggest that you

go find another surgeon.

And in a way that's where this can also provide assistance to the patient, is it's a beginning, it's not the end of the process. I don't, I don't know if that makes sense, but I hope that it makes a little bit -- I mean --

CO-CHAIR FLEISHER: So Melissa and then Elisabeth.

MS. MARINELARENA: So I think we can pretty soon probably just start going through the, these are great issues, start going through the criteria for the measures, but the conversation that you were having right before TeMaya is actually very important, because we do pose questions for all the measures, about the level of analysis for the measures.

They are checked off as specified at the facility and clinician. And we do talk to STS about this, and if these were the two composite measures were evaluated by the Methods Panel, so they are, there's some language in the preliminary analysis, because the specifications

need to match the testing, and the testing is not at the facility level. They are clinician-level-testing, so we do ask the -- we're asking the Committee to discuss that and to give us some insight on if you think that the testing actually does include facility levels -- normally we see testing that you can compare facility-to-facility, we have clinician-level-testing on there. And they did provide an explanation to us, but we'll have that discussion when we get to testing. I'll hand it over to Elisabeth, and then I think we should get started if we're okay with that?

CO-CHAIR FLEISHER: Yes.

MS. MARINELARENA: Okay.

CO-CHAIR GUNNAR: Okay, we'll move forward. Back to any other comments from STS, before we proceed?

(No response.)

CO-CHAIR GUNNAR: Very good. Dr. Cima, 2561.

MEMBER CIMA: This is Bob, yes, I was

on mute. So this is the composite measure for Aortic Valve Replacement 2561, STS's measure steward as we know, and it is composite. I mean, do you want to go through this like we normally would? We just went through most of this.

It's a composite of mortality, absence of mortality, risk adjusted, and major morbidity. The major morbidities are the ones associated and reported on multiple other STS measures, CBA, surgical re-exploration, sternal, deep wound, sternal wound infection, and excuse me, post-operative renal failure. The numerators they enter were very straightforward. It is the more -- the one we always talk about, for mortality, is that it's, within 30 days, or within that index hospitalization, even if it goes beyond 30 days. That is always, has been, in the past, a question of concern, so I, I think they all, they do a very good job of giving us the broadest picture, possible, and trying to capture the most patients that might be appropriate for that reportable component.

The major morbidities are basically anyone that has an isolated AVR that doesn't experience the five specified major morbidity endpoints, and as was pointed out by the developers, it's dichotomous. It's not a little renal failure. They have a strict definition, and you either have it or you don't.

So that is the measure they have done, as we know, a very detailed statistical analysis and have a number -- a little army of statisticians that help them, and so we have never had any questions about the validity of those statements.

The star rating is the composite. We've discussed that. I did make a comment on my worksheet because we had always talked about it being voluntary. We always know that there's 90-plus percent of cardiac programs that are in STS, but originally it was you had to be voluntary.

I was wondering and, I guess, they confirmed already the statement that it is still

voluntary, it's not required reporting, which gets to one of the other discussions we had earlier and will probably come back to.

One question I did have for the developers about that, was the -- they didn't -- they used to, in previous ones I reviewed, they gave the distribution, what percentage of patients, of institutions fall in the one star versus three-star.

And I just, I would just appreciate, if the developers, at the end of this, or before we get to the voting, could describe that? I remember --

CO-CHAIR FLEISHER: Just for clarification, NQF, we don't review the star ratings themselves, so that's not a relevant point for endorsement. That is post-endorsement, how it's displayed, correct?

So while it's of note, it's not part of the endorsement --

CO-CHAIR GUNNAR: It's not a determining factor.

MEMBER CIMA: Okay. I just was wondering about the, the, the usability of it, in the sense of if you only have one percent of institutions, or one star and one percent of three star, does that -- is that really a usability?

So, you know, we look at disparities and gaps and we look at that, but I would think how we're reporting it also because we're talking about publically reporting, is it useful, but if it's not within the purview, but that was my thought.

So preliminary analysis, do you want to get to evidence?

MS. MARINELARENA: Yes.

MEMBER CIMA: Okay.

MS. MARINELARENA: Yes, we're discussing evidence and you can talk about the star rating in both validity under meaningful differences and usability as well.

MEMBER CIMA: Oh that was what I was getting at.

MS. MARINELARENA: Yes.

MEMBER CIMA: So evidence. I think they have demonstrated in the evidence the volumes are important, and there still remains a performance gap which is in the order of a reasonable amount of considering the volume and the degree of severity of the complications, there is a reasonable gap that would indicate that, it should -- it is important to measure.

Other than that, the quality of the data is good. The methodology is well-established and has been refined over decades. And so I did have the comment there about, you know, going back to the whole issue of, if we're really going to pull these apart, then they should be separate endorsed measures, but I defer to the rest of the Committee on that, but I felt it was preliminary composite rationale was high.

CO-CHAIR GUNNAR: So just to get us, because this will get us the first voting. Everybody's on the -- make sure everybody's on

the website, right, on the link?

MS. SKIPPER: Yes. So you all should be logged into the Poll Everywhere link, sent via email, yesterday. Should the vote be coming up?

MEMBER SAIGAL: I don't see anything on the screen yet, is it --

CO-CHAIR GUNNAR: Yes, it's just the blue screen right now, right? Or are we in -- do we have a poll just --

MS. SKIPPER: It should be a blue screen.

CO-CHAIR GUNNAR: Just this?

(Simultaneous speaking.)

MEMBER DUTTON: I've got a blue screen.

MS. SKIPPER: I haven't activated anything yet.

MEMBER DUTTON: Can, I ask a quick question, while we're waiting to activate this?

For the developers, are any of these measures today, but, this one included, have you advanced any of them to MIPS --- to CMS for MIPS

reporting, or do you use them in your QCDR, or so STS to MIPS, for reporting?

DR. BADHWAR: The point of order of MIPS is for individuals that are --

(Off-microphone comments.)

MEMBER DUTTON: Yes, it's not, it's not hospital reporting, it would be physician reporting, but that, but those are publically reported measures, which is why I asked.

DR. BADHWAR: We have not, as a society, done that yet.

(Off-microphone comments.)

MEMBER DUTTON: Any of them? Because that, that is publically reporting, right? I mean that's headed for public reporting, even if it hasn't happened yet?

MR. ANTMAN: So we do have a number of STS measures that are part of the MIPS Program. I don't have a list of those specifically; we can certainly access them.

I don't think any of the 15 -- the -- I don't know if the 15 measures that we're

looking at today specifically are, but we can certainly look that up.

And that's actually on the website, I think, under the -- if you search under QCDR, I think it'll actually list the measures. I actually think a few of them may be. It's a good point, though.

MS. KOSURI: We do have a Member who stepped away, so we'll do a test run just to make sure that we have quorum before we proceed to the actual vote, if that's okay? So I have activated the poll, as you can see. And if you can click the link and input your vote.

CO-CHAIR GUNNAR: So everybody's bar is working? Everybody -- anyone not --

MS. KOSURI: So as far as, you know, we only have 12.

MS. SKIPPER: And if you're not by a computer, there's also a Poll Everywhere App that you could download on your phone to vote. And if anybody's having trouble that's on the phone, you can definitely chat us and we can resend the link

to you.

MEMBER JARRETT: This is Mark on the phone. What we should be seeing is that it says response recorded, right, underneath it, under the test?

MS. KOSURI: That should -- that should be it. That's the indication.

MEMBER JARRETT: Okay. Thank you.

MS. KOSURI: Yes.

DR. PAONE: This -- while you're testing that, if I can just follow-up the question about MIPS and NCDR? The website does list, it has a page where it does list 11 cardiac measures and one, two, three, four, five, six thoracic measures, which can be reported for MIPS.

It would be surgeon-specific and it's requested by surgeons, so it's --

(Off-microphone comments.)

DR. PAONE: And QCDR.

(Off-microphone comments.)

DR. PAONE: If it's requested,

correct.

DR. BADHWAR: These are the QCDR?

DR. PAONE: QCDRs.

DR. BADHWAR: QCDRs, yes.

DR. BADHWAR: Yes. Yes.

MEMBER DUTTON: Yes that seems pretty obvious, public reporting news of these measures.

DR. PAONE: Exactly, yes.

DR. BADHWAR: And the question, on the phone, about the breakdowns, sort of, post-approval question, about percentages, so the Bayesian analysis in general for all these measures, it's essentially 80 percent are as expected, and there are 10 percent that are higher than expected, 10 percent lower than expected.

And it varies by measure, particularly those that are less frequent operations, it's a little bit more of a tighter bell curve, but it's roughly that. It can be 70 percent is as expected and a little higher, three-star a little higher, one star, but it's not as, who's on the phone,

but it's not 98 percent three-star and one percent one star, it's, it's still -- it's a bell curve of Bayesian statistics.

MEMBER CIMA: Great, thank you.

DR. BADHWAR: You're welcome.

(Off-microphone comments.)

CO-CHAIR GUNNAR: Hey, there, if we're coming back live, we should use our mics and, if we're not --

(Off-microphone comments.)

CO-CHAIR GUNNAR: Okay, so we don't have a quorum. So unless somebody has not voted online, the way it works is we will vote here and they will, but they will not display the vote results, and we'll just spend the day voting and then online with the Survey Monkey, the additional people will vote, and we will know the results at a later date, correct?

MS. SKIPPER: Right.

CO-CHAIR GUNNAR: Yes.

(Laughter.)

CO-CHAIR GUNNAR: Except we're still

playing.

MS. SKIPPER: All right, we're going to just take a quick break.

(Simultaneous speaking.)

CO-CHAIR FLEISHER: Can we do a -- so who's connected on the -- do we have people, on the phone, that can just, sort of, raise their hand, do we know?

MS. SKIPPER: We missed Dr. Handy, he had to step away.

PARTICIPANT: Is he going to come back?

CO-CHAIR GUNNAR: So we're at 14, we're still at 12.

CO-CHAIR FLEISHER: No we're at 13.

MS. SKIPPER: We've gotten to 13.

CO-CHAIR GUNNAR: But, I mean, so --

CO-CHAIR FLEISHER: So why don't we take a five-minute break, while we figure this out?

(Whereupon, the above-entitled matter went off the record at 10:08 a.m. and resumed at

10:19 a.m.)

MS. SKIPPER: So we are going to be voting on evidence for Measure 2561. And voting is now open.

(Pause.)

MS. KOSURI: So I think we have all the votes. So for Measure 2561, STS Aortic Valve Replacement Composite Score, for the evidence portion we have passed this measure for the evidence part with 14 votes passing. So that is 100 percent. Thank you.

MS. SKIPPER: And we can now move on to the discussion of gap. If there is not anything -- yes, discussion of gap.

MEMBER CIMA: As I mentioned earlier, there is, you know, a small band of performance that sort of -- you know, we are doing well across the country but there still is some gaps. And given the severity of this, both the scientific team and -- and my view of this, if I wanted to know what -- if I were having cardiac surgery, I think there is data to support

continued measurement of this because of the implications of it. So --

CO-CHAIR FLEISHER: Thank you. I think particularly we have a lot of measures and they are very similar and we have been here before -- that unless there is a significant concern -- perhaps in the discussion, if you feel there is gap, unless anybody else feels we need much discussion, we can just go forward and say it's -- there is a gap. Elizabeth?

MEMBER EREKSON: I just have a question about penetrance because that has been a question that we have raised about the STS database before and I think -- I believe in the measure that I reviewed there was a bump in participants from this kind of 900 into the 1,000 range. And did you guys get a bump either by adding sites or -- where is your penetrance now?

DR. BADHWAR: So it is a very interesting question. So you are talking overall in the database in terms of volunteer public reporting, correct? So if that -- or -

MEMBER EREKSON: Participation.

DR. BADHWAR: Yes, in public reporting.

MEMBER EREKSON: Participation in the database.

DR. BADHWAR: Oh, in the entire database?

MEMBER EREKSON: Yes, in STS -- in this measure.

DR. BADHWAR: In this particular measure? So there -- several things. The answer is two-fold. One is, programs are growing still. Second is that approximately five years ago, we had an alignment with U.S. News and World Report and that the process of a program voluntarily publicly reporting through a clinical registry was credited with a small number -- a numerical credit, and the U.S. News and World Report Cardiovascular and Heart Surgery Guidelines score, and that was communicated to all participants and perhaps -- though we don't know for sure exactly what changed -- it further

encouraged in our pathway towards voluntary participation on all of the measures. Second part of that -- third part of that is that in this particular isolated AVR is relatively still newish compared to CABG. And it is a three-year rolling participation and so programs that are early on, it takes them three years to accumulate data to actually participate, as opposed to the CABG one.

MEMBER EREKSON: So the question is how many -- we have heard before that it is 95 percent of hospitals performing these procedures are participating in this database. Is that your current estimation as well?

DR. BADHWAR: Yes, that is correct. It is approximately 95 percent overall. Some of the VAs don't participate, but essentially nearly everyone else, because it is an important measure to participate in a clinical registry. And so it is actually higher than 95 percent.

MS. SKIPPER: Okay, are we ready to vote on gap for 2561? Okay.

MS. KOSURI: Voting is now open.

(Pause.)

CO-CHAIR FLEISHER: Put this on the record. So the question -- will you rephrase the question.

MEMBER YATES: Are TAVRs considered part of the measure?

DR. BADHWAR: No, currently TAVRs are not considered part of the AVR measure. These are surgical AVR measures. We are working collectively with our cardiology colleagues at the American College of Cardiology in developing a risk model for public reporting with TAVR, which is ongoing right now. And that will soon be presented to NQF.

CO-CHAIR FLEISHER: Results?

MS. KOSURI: For measure 2561 for performance gap, we have 14 people voting moderate and one person voting high out of 15 people who voted.

CO-CHAIR GUNNAR: So the measure passes on performance. Shall we go to --

MS. SKIPPER: Composite construct.

(Pause.)

MS. SKIPPER: So you all are voting on the quality construct and rationale -- is it explicit -- explicitly articulated and logical? I think we talked a little bit about that this morning. Is there anything else, Bob, you would like to add? Or anyone else in the room?

MEMBER CIMA: No, it is -- it is based off the STS. So we have done this multiple times. The methodology is sound and very reliable.

CO-CHAIR GUNNAR: Right, I was -- just as an aside, I was going to streamline that for future discussion. If there is no new evidence and the evidence is still the same basis for which it -- then we can sort of state that and move on. Same here. I mean, we have seen this measure before. So I -- any other discussion?

(No response.)

CO-CHAIR GUNNAR: If not, we will vote.

MS. KOSURI: Voting is now open for the composite portion of measure 2561.

(Pause.)

MS. KOSURI: We are still waiting for one more vote.

(Pause.)

CO-CHAIR GUNNAR: But to be clear, 14 is quorum. So if we hit the 14, we are good, right? Fourteen apparently is just fine. We have 15 that are voting.

MS. GOODWIN: Just to put it on the record, the reason why we went from 15 to 14 is because we are not counting Dr. Grover in the denominator because he is abstaining.

CO-CHAIR GUNNAR: Correct. So we are there now. And it -- and it looks like it passes. Okay, sorry. Go ahead.

MS. KOSURI: No worries. With -- for measure 2561, for the composite portion, we have 8 people who voted high and 6 people who voted moderate with a total of 14 people. And so this passes.

CO-CHAIR GUNNAR: And next is reliability. Anything you want to say about reliability, Dr. Cima?

MEMBER CIMA: No. I thought we were streamlining these.

CO-CHAIR GUNNAR: Yes, I agree. But anything to add?

MEMBER CIMA: No. The data speak for themselves. The association and the analysis shows that it is very reliable over time.

CO-CHAIR GUNNAR: Yes, let me rephrase the question. Since our last discussion of this, is there anyone who has changed their mind on reliability?

(No response.)

CO-CHAIR GUNNAR: Seeing no comments in the room --

MS. MARINELARENA: Actually this measure was retested. They have to test the composite construct. The methods panel reviewed the measure and did provide some -- their ratings. For reliability, you can choose to

accept that rating and not vote. There were some recommendations that they made here and some issues that they want the committee to discuss. I don't think the -- it is on page 8 of the measure -- of the preliminary analysis. The full specifications were included. So I don't think that was an issue.

The subgroup members have concerns about the inclusion of SDS doctors -- well, that is validity. But in the risk adjustment model the standing committee should review and determine if SDS factors are appropriate. You do have to vote on validity because they did not reach consensus on validity of the measure. So you need to discuss validity and determine if the results are acceptable. And then again, like I mentioned earlier, we need to seek clarification from the developer on the levels of analysis that were tested and which testing results correspond with each level of testing -- whether it is facility level or clinician level. So the methods panel did not have a problem with the

reliability. You can choose to accept the moderate rating, but we do have to discuss validity and then vote.

CO-CHAIR GUNNAR: So we are on reliability at this point. And we will discuss validity next. So any other further discussion on reliability?

(No response.)

CO-CHAIR GUNNAR: Go ahead and vote then.

MS. KOSURI: Voting is now open for reliability.

(Pause.)

MEMBER YATES: Just for the record, the SDS question -- I mean, all that was included was race and gender. And I am not -- I am not aware of gender being an SDS question or -- and race is separate.

CO-CHAIR GUNNAR: Hold those thoughts.

MEMBER YATES: Okay.

CO-CHAIR GUNNAR: Hold those thoughts for just one moment. Let's get to -- that's

next, though.

MEMBER YATES: I saw it in front of me.

(Laughter.)

CO-CHAIR GUNNAR: We are just shouting out things now. All right. What do we have for this one? Reliability?

MS. KOSURI: So for reliability we have five people who voted high, and ten individuals who voted moderate, with a total of 15 people.

VICE-CHAIR GUNNAR: Very good. So it passes. Now we will go to validity. And Dr. Cima then Dr. Yates.

MEMBER CIMA: Well this -- this sort of went to my question about the star ratings because I also was wondering about that. I think the group that reviewed this was also concerned about the star rating as being the publicly reported aspect of this. And that's what I was getting at but I was -- I was -- you know, I think the developer needs to discuss that if that

is a valid way of doing it, given it's just sort of so generic. But the -- that's what they asked us. And I also wasn't quite clear on what they -- why they wanted to separate out some of those risk factors. So I was going to defer to the group because I am not sure we have actually good measures to inform the database on this. But those are the two issues that came up under validity. And then my -- my main focus was on the star rating.

VICE-CHAIR GUNNAR: So Dr. Yates and then the developers?

MEMBER YATES: Correct me if I am wrong, but it's -- the two things they found disparity in was that higher rates of -- for women and for blacks, or African Americans, and the -- the logic being that that's a physiological or genetic issue as opposed to SDS/SES per se. And you know, people don't like including SDS in some things because they feel like it is going to reward hospitals that are performing poorly as opposed to being an actual

risk factor. But I think for the purposes of this specific technical measure, it really doesn't make a difference. And we are not talking about poverty per se. People use African American race as a surrogate for poverty, but it is not necessarily, you know, a risk factor in this particular case. I would be interested in how the measure developers feel about that.

VICE-CHAIR GUNNAR: STS? Do you want to ---

MR. ANTMAN: So if I may confirm, I believe our lead statisticians are on the line with us, Sean O'Brien and Maria, and they can speak to specific questions related to reliability and validity.

DR. BADHWAR: While they are chiming in, I will just comment that there's obvious validity -- there's statistical validity, of course, but then the face validity of this measure, I think everyone is aware -- in terms of how we approach people -- to address your other question on the social disparity issue. As we go

through model development and the amount of hours that go into parsing out each risk factor and the coefficient of how that impacts the overall model, each of these items are addressed. And we go through one at a time -- age, sex and race. And to this date, those haven't parsed out to be significant yet. We are continuing to observe those. So hopefully that addresses your question, particularly as you made those mention. But I will have -- and I think Sean O'Brien and he can comment on the statistical validity calculations.

MR. O'BRIEN: Sure. This is Sean O'Brien from Duke University. With respect to the race and sex question, I had not remembered that that was raised in the context of validity that it's -- as it relates to the risk adjustment procedure and the rationale for having factors in the model that are associated with socioeconomic status and disparities issues. And our framework for approaching the risk adjustment is to -- is to think about minimizing confounding in the

sense that we would like to -- it is not practical to think about a randomized trial, or randomizing patients to go to a particular provider in order to obtain an unbiased assessment, but we would like to ask the question, how would results at one particular participant -- how -- what would those results be hypothetically if the mix of patients was similar between that participant and the reference population that they're being compared to -- which is the overall STS?

And you know, a requirement for unbiased estimation is that all the -- all the factors -- you have measured and accounted for all the factors that are potential confounders. So we are kind of agnostic with respect to what variables are potential confounders and to satisfy the assumptions that are required for unbiased estimation. You are more likely to meet those -- I feel like you are more likely to meet those assumptions, which inherently can't be verified, if you adjust for a large number of

factors. So it wasn't -- they were not in there or out, specifically, because they were somehow connected to STS-related variables therein, because they are one of the dozens of factors that are measured on patients, but in principle we would like to balance all of the potentially important factors across participants when they are being assessed in comparison to the -- to the benchmark. And I will keep going unless there's -- unless I am interrupted for questions. I am happy to be.

CO-CHAIR FLEISHER: So to be clear, race is or is not in the model?

(No response.)

CO-CHAIR FLEISHER: Race is in the model, which presents a problem.

DR. SHAHIAN: Can I respond, Dr. Fleisher?

(No response.)

DR. SHAHIAN: We -- you know, putting -- putting on my NQF hat now for a second, I think NQF is now open to considering

socioeconomic or socio-demographic factors when there is a plausible association between those factors and the outcome. And for example, we did a very extensive study of socioeconomic factors in readmission because we think there is a very clear potential there.

On the other hand, for mortality, we see no particular reason why there should be an expectation that Medicaid or dual-eligible status or any of the other markers for socioeconomic status should affect mortality. Now, why is race in the model? I think race is in the model because historically we have had it. And we can't parse out the genetic or physiologic aspects of being, for example, African American. That population has a higher prevalence of hypertension. Hypertension happens to be a risk factor. So -- so there are potential socio-demographic factors associated with race, but also genetic factors. And we have just taken the agnostic view that, you know, this has an impact. We know it has an impact. We are not considering

it because of its socioeconomic association, but because of the potential for racial differences in risk like hypertension and things like that. So that is kind of the rationale for that.

MS. MUNTHALI: So thank you for clarifying that. And I just wanted to clarify for the committee, Dr. Shahian is right that we are looking at socio-demographic factors and social risk within risk models differently than we did before. So prior to three years ago we prohibited the inclusion of social risk factors in our risk models. We also have said that race should not be used as a proxy for social risk or socio-demographic factors. But it can be used to make some genetic inferences. So that clarification was important for us because we would have had some challenges trying to look at the measure as it currently is specified.

MEMBER YATES: And I would argue that it would be better if they use the AHRQ ZIP code poverty index, or dual-eligibility as an index. Because I don't think people have more

complications because of the color of their skin, per se. Now, maybe the association of somebody's skin color with -- as was put items such as -- or, conditions such as hypertension and the like -- would be important. And from a genetic perspective, then I think it is reasonable from a medical risk adjustment to conclude it.

MS. MUNTHALI: And I think you're right there. And one other thing I did not mention is, as we are asking developers about the connection of these socio-demographic factors and risk factors to the outcome, we are asking them to lay out a conceptual model of what that would look like and to then validate it with empirical testing. So that's -- and -- and because of the testing, or whether or not there is a conceptual relationship, the data or the evidence may not show that. That is why we are asking them to kind of lay out this entire rationale in their submission to us.

CO-CHAIR GUNNAR: So with that background, any more discussion on validity?

Amy?

MEMBER MOYER: Yes, I guess have a question conceptually about if the reason for including race in the risk adjustment model is the higher prevalence of these clinical factors, wouldn't those clinical factors already be accounted for and taken into account in the risk-adjustment model? Wouldn't that perhaps -- is that like a double impact potentially? I am a little confused by the separate influence of the two factors.

VICE-CHAIR GUNNAR: Comments?

MEMBER SAIGAL: It sounds like it to me. I mean, basically it's -- intent is not to have race in the model. Race is in the model. I mean -- you know.

(Simultaneous speaking.)

DR. SHAHIAN: I used hypertension as an example, but -- Sean may want to expand on it -- but I think there are potentially numerous genetic factors and racial factors that we may not account for in the model. And given its

impact on -- in previous analyses on outcomes, we decided to leave it in, again, without using it as a proxy for socio-demographic status -- or socioeconomic status, but using it because of potential genetic and physiologic factors. And Sean, you may want to expand on that.

MR. O'BRIEN: Well, I mean, our approach to assessing its inclusion or exclusion was empirical. So the model does adjust for a very large number of risk factors -- close to 40 -- but empirically, in the case of the morbidity component -- and there is a clear association even after adjusting for those variables -- between race and the morbidity outcomes. So it was included. I think that it was not included in the mortality portion of the adjustment for the mortality component in the case of the ADR measure. But we see empirically -- we don't know the underlying reasons for the association, but we know that race may be associated with, you know, any number of unmeasured factors. And so the idea is that we -- to the extent that there

is unmeasured factors that could be explaining the observed differences between the participant who is being evaluated and other participants, we need to try to maximize -- measure those. And so to measure -- to measure variables that are proxies in the sense of being associated with -- with underlying causal factors is adequate to minimize confounding that would be there if you didn't at least adjust to the variables --

(Simultaneous speaking.)

MEMBER SAIGAL: Can I follow up on that? Because -- so are there measures of SES in the model? Dual-eligibility? Is that right?

(No response.)

MEMBER SAIGAL: No? So then basically, I mean, there known risk factors that are associated with race are already in the model and you are saying something else is accounting for the impact of race, and it sounds like it is SES. I mean, genetic factors are undefined. So logically that sounds what it is. So now -- and basically it is the intent is not to have race

used in this way. It is being used in this way in my mind. So I would say it is a problem unless you can change it.

DR. BADHWAR: Can I make a comment just to add to the genetic -- just to give a clinical, real-world example? So, East Asian Indian males of lower than average body weight but higher preponderance of small-vessel coronary disease, just to give an example -- patients such as those have a little bit more difficult coronary anastomoses to be done. There is no way of grading that in -- in -- using the current risk models that exist. Difficult small-vessel coronaries -- that is not a factor that we have on our database. But an example of that -- and there are many other that -- why race can be important and used in a -- in an instructive --

MEMBER SAIGAL: But in India, if all those guys have small vessels -- in India, all the Indians in India are having this problem? I mean, it doesn't make sense to me. How do we know that that's true across the entire

population of Southeast Asians? It is an impression.

DR. BADHWAR: I will respond by saying just from -- I am just using that as one example. It's a fairly well accepted -- within our -- within the cardiac surgical community, when you see someone of a certain ethnicity, you are expecting a certain variable in the operating room. It is not measurable by current ways -- we don't have a defined way of measuring X, Y and Z, but it is an inference. But that is why race is in the model. Not by any kind of socially -- social inference, but by other factors that we can't account for.

MEMBER SAIGAL: Right, I understand the intent is that. But the impact, actually, it sounds like they're measuring SES, especially if they are African American. I mean that's -- that's who you are loading on there because you already have the other factors incorporated that we know matter in terms of the outcomes here. The --

(Simultaneous speaking.)

DR. BADHWAR: That is assuming that you have in fact included every other potential confounder in your model, which -- I am sure we haven't. There are -- you know, we can only account for a certain number of confounders that we routinely measure. I would not assume that all the delta between what is in the model and not is SES. That's our point. And I think that's what Sean was saying as well.

MEMBER YATES: Yes, in defense of STS, they took a lot of different variables and then they threw them into multi-variable regression analysis and they came up with things that made a difference. And race happens to empirically make a difference. And that -- unfortunately, I have to mix two meanings of the word black, but it's a black box in the sense that you don't know what is causing that difference to the Nth detail. But there is enough multiple different things that are different for that to make it an important risk factor and it has to be included.

I mean, in -- I have to use an orthopedic analogy, but, you know, we would need to use gender as a risk factor for hip replacement because it is -- we know that the shape of their acetabulae are different and they have a smaller pelvis. And we know that they are more osteoporotic. And -- but there's also data to show that there is a different ligamentous -- a different tension to their ligaments, a different propensity to ligamentous tears and the like, and so we are -- we are left with some things that we -- are so poorly defined, we can't use them as individual risk adjustments, but when added together they become a composite risk. And I think it is reasonable when it is used for this purpose -- and especially for black and white, again I am using the word black in a different context, in a black and white outcome of these type of morbidities and this -- and mortality, I think it is an important thing to include. And it's not going -- it is not expected that the socioeconomic component of race, or the overlay

of that, is affecting these very clear-cut outcomes. And that's where I would argue --

(Simultaneous speaking.)

MEMBER SAIGAL: Well, why is it not expected? Why is -- how can it not be expected? It should be expected.

MEMBER YATES: The -- because, again, I don't think that -- I don't think that the -- necessarily the definition of race in and of itself is dragging poverty into this particular question. You know, people use race as a surrogate for poverty, and I don't think that is an appropriate use for it in this case. That is my opinion and I think it would be justified by statistical analysis.

MEMBER SAIGAL: Well in my view it has not been justified by this analysis. I think -- I mean, I think this is a fantastic registry and project. I don't think it should go down for this, but I think it is notable that this has not been addressed in a meaningful way. So I would just make that in a note on this measure that it

has not been addressed.

MEMBER EREKSON: I would like to support what you just said. And I would also like to say that as we consider these socio-demographic risk factors, there's just better variables to measure it than race as a surrogate. And so, if we are accounting for all the knowns and we are putting race in the model, then it would really be nice to see these other things considered or not considered as -- as a note for the record and for the future that, you know, all of the things that we've already discussed -- the dual-eligible, the Medicaid -- all of the things that we can easily obtain out of billing data that doesn't include adding more variables to an already cumbersome data collection, would be helpful for us to be able to review in the future.

CO-CHAIR GUNNAR: Dr. Stein?

MEMBER STEIN: I am not arguing against -- I think it is fine to keep race in the model, but an easy solution is just also do a

stratified analysis by race, right?

DR. SHAHIAN: We have those data available. And it -- and it is not being -- just again to repeat -- we are not using it as a socio-demographic variable or proxy. That is not its purpose in this model.

MS. MARINELARENA: I just wanted to point out too that Hispanic is part of the model. Hispanic is not a race even from -- for a clinical perspective, Hispanic is not a race, it is an ethnicity. So it is not clear what clinical issues we associated with Hispanics.

DR. BADHWAR: Just a final response to Dr. Saigal's comment that -- again, that this is -- it is definitely not a surrogate for socio-demographic profiling. That is not at all the intent of this. But certain races or demographics do have a higher preponderance of diabetes -- a higher preponderance of poorly controlled diabetes and secondary effects that are covered in the model, but it is -- as Dr. Yates mentioned, it's a multi-variable factor

that came out in the calculations.

MEMBER SAIGAL: I totally get that that is -- not the intent. I completely hear everyone's comment -- not the intent. But logically, it is the impact. I mean, Indians don't show up as a special risk model -- factor in the risk model, although they have a smaller set of coronary arteries that apparently, according to consensus. So I mean, something else is driving it. Logically it has got to be SES. In other models, that is what it is -- that is what it loads on. So I would just say it has to be looked at. And as -- I think the suggestions on the committee are really important because I don't know if we should or shouldn't put this in as a -- as a -- in the risk models, but NQF says we shouldn't, so as a committee we have to sort of discuss that. And that's just -- I think it is a glaring example of what has not been happening.

MEMBER YATES: I would argue that there would be an unintended consequence of

dropping it from the model. If it is thought to be an important risk factor, you run the risk of patient avoidance and not -- you know, lemon-dropping by surgeons. And there would be less access to care. And I -- at this point in time, I really think that we are arguing apples and oranges. Not to mix it in with the lemons.

MEMBER STEIN: I would just encourage the developers to put in the material they are sharing with us the stratified results. I think that would help.

MS. MUNTHALI: And I would second that because that is actually a requirement. So we ask for not just the risk adjustment based on the SDS, but also -- or SES -- but also stratification so that we can see the differences between groups.

CO-CHAIR FLEISHER: So my question is actually hypothetical depending on where this goes. But if you wanted to show us that inclusion in the model did not have a negative effect on adverse selection, and actually

institutionalizing any kind of racial profiling, or -- who would you do that? Have you thought about that? Because that's the concern is that it is institutionalizing differences. So have you seen over time -- has the race Hispanic or -- since that is not a race -- or African Americans have changed and the influence of that? Or -- and as well as the influence of detecting the extent of disease? Because what I am hearing is disease is here and your argument is actually severity rather than presence. Because you have a lot of these kind of risk factors that are associated with differences between different racial groups. So is it severity of detection?

DR. PAONE: I am not exactly sure how to answer this, but you know -- considering socioeconomic status has become in vogue over the past few years -- and I don't say that lightly -- it is clearly not an easy thing to define. I think -- depending upon what you look at, it becomes controversial in one way or another. You look at geography -- and if you look at

geography, is their ZIP code enough? Or a city? Or do you have to go down to the street or apartment level, right? Before you actually get reliable data that's usable?

You know, the concept of dual-eligibles I think is a good one at the patient level because I do think there is some evidence that at the individual patient level, the outcomes are a little worse for dual-eligibles than for others. But if I remember correctly, we did a pilot study of this at STS a few years back, and the numbers are relatively small, so that it really doesn't drive any changes in the star ratings overall from that perspective. And so --

CO-CHAIR GUNNAR: So let me be very clear, you have analyzed one star -- one-, two-, and three-star facilities and socioeconomic factors are not driving those ratings?

DR. PAONE: I think we have -- I believe we have done that, including dual-eligibles as an SES variable -- in a pilot study.

David, is that not correct?

(No response.)

DR. PAONE: I am sorry? Oh, that was for readmissions. I apologize. But -

CO-CHAIR FLEISHER: I didn't look -- how strong is -- what is the coefficient of the influence of African American race as well as Hispanic in your model? Is it, you know, 1.02? Or 1.0 -- you know --

(Simultaneous speaking.)

DR. PAONE: I don't have that information. I would ask Sean, he is on the line, if he knows --

CO-CHAIR FLEISHER: Yes, I am asking --

(Simultaneous speaking.)

(Pause.)

PARTICIPANT: Sean, are you on the line? The question was, what the coefficient was for race, if you are looking that up, perhaps.

MR. O'BRIEN: Apologies, I was on mute. So -- so in the morbidity endpoint

component the odds ratio comparing patients with black race compared to other races including white was 1.27, in other words a 27% increase in risk. And as I mentioned -- race was not a factor in the mortality risk adjustment.

DR. PAONE: Let me just add a more personal sort of practical approach to this is -- I don't know any surgeon who has ever looked at a patient and said, their risk is higher and therefore I am not going to operate on them -- because they're -- because they're black which means they are likely poor. It just doesn't happen in the real world that I am aware of.

MEMBER YATES: On the contrary, saying any surgeon would be fine if you were saying any cardiothoracic surgeon, because you are dealing with situations that aren't truly elective. Whereas, when there is the opportunity to perform elective surgery, unfortunately there is biased lemon-dropping and cherry-picking in elective surgeries that leads to loss of access. So I -- in the real world, that really does happen. And

I can tell you people that -- there are surgeons that just refuse to see Medicaid, and not just for medical -- not just because of reimbursement.

DR. PAONE: Well then I will accept that and understand it, but change my comment to say in the real world of cardiac surgery, then. And I would believe that very strongly.

MEMBER YATES: And to STS's credit, the case complexity mix, which is another way of looking at risk avoidance, has gone up slightly in the STS database. And to the STS database credit, partly because you guys are risk-adjusting, which does make it an even playing field for all surgeons, so they don't feel like they are being unduly measured for taking on complex cases. So kudos to the STS for doing that.

MEMBER JARRETT: This is Mark. I raised my hand a little while back, but I guess that's part of the punishment for not physically being there. You know, I am listening to this discussion, you know, our health system is very

much involved now through one of our leaders in looking at social determinants and all of these issues. And I think -- we all agree that I think STS needs to move to, you know, a better mechanism to include social determinants. And I don't think we know what the impact of race is -- whether it is genetic, whether we are really looking at a proxy for access and other social determinants. But we have to realize, we are really at a point in the science of this that we really don't have great social determinants. I mean, you know, dual-eligibility is a poor proxy for it, quite frankly. Doesn't that account for a lot of the population that has no insurance? And I think that we really need -- you know, need to sit back and say this is the status of where we are today, and it may not be ideal, but it does provide information that, if it is publically reported, does help the public, which is where we are going -- as well as help performance improvement. But saying that we expect that STS will come out -- because I -- I

had the same concerns about the measures that I looked at, that they need to come out with a -- with a viewpoint on how they are moving forward with social determinants and other things more quickly.

The other thing -- I will make a comment -- also, depending on what's written in the medical record as race and ethnicity -- when we did an analysis of this and an audit of this, it is kind of frightening. And I have a feeling in a lot of hospitals and other places around the country, this is often got by a registrar who usually has no concept, is afraid to ask, or puts down, you know, miscellaneous or other -- things like that. So I worry, also, about the validity of what -- of actually what is written, especially as you move from white versus African American to anything else.

DR. SHAHIAN: Could I respond to the question about where we are heading in exploring this at STS? We are very close to signing a contract with an academic institution that will

provide us with the ability to geocode every single patient in the STS database. And for those of you that aren't familiar with this, most people think this is probably the single best way to assess overall socioeconomic or socio-demographic status -- gets you down -- it takes the patient's address, converts it to a very highly specific latitude and longitude, and gets you to a block-level, or Census-track level socio-demographic status. We will do this with every patient. And then we can explore some of these questions that have been raised empirically. So that's -- that is something that we hope to have within a year.

MEMBER YATES: And ideally you would cover that as a continuous variable across the entire population as opposed to using a cutoff, which cuts out the community effect because -- you know, my observation is, it's the social topology of the neighborhood that makes it a tough deal in terms of getting good healthcare and getting access. And there is a big

difference between living in some parts of the Bronx versus living in Palo Alto.

DR. BADHWAR: Just a comment on -- not representing STS, but just as a healthcare researcher in general -- I mean, as all of you -- that the data is just not there yet, but these types of issues like geocoding are evolving rapidly. Another one is distressed community index and what that -- how that would influence outcome. And so, as these types of variables are becoming more accessible for our healthcare analyses, perhaps one day -- to address Dr. Saigal's earlier comments that -- which are very valued, that yes, it's incumbent upon us as a society, regardless of what specialty it is, to have a more robust analysis before making a conclusion of connection.

CO-CHAIR FLEISHER: So I actually have a question, Lisa. I am going to put you on the spot in that, as we think through this and try to make determinations -- and as the committee, which owns this portfolio, weighs a decision on

this particular issue, is this a up-or-down because of what we think about how they included those particular variables? Is this a clear message, then, when this comes back -- do we have the option to say if this comes back, we have to change to what we hope will be the next -- next time, with a clear signal? Is there any clarity from your perspective on how the committee is clearly concerned about the inclusion of African American and Hispanic in the model?

MS. MUNTHALI: Yes, it is a good question. We are struggling with that now. I mean, as it clearly is included in the specifications, it does go against our policy as part of this SDS trial. It is a five-year trial. We are finding the developers, to the point that is just made, are having a hard time getting access to the right data -- merging those SDS data with clinical data. It has been a challenge. So they are using perhaps less than optimal, less than desired data to make inferences about social demographic factors and

their relationship to outcomes.

With that said, I think the committee needs to send a strong message because it is part of our criteria. We are in -- we are all in a period of discovery with the influence of risk factors on health outcomes and the impact of those factors that are outside the clinical care delivery system. But we would like to see that variable not in the risk model as a proxy. If it is indeed -- I was getting a little confused about whether or not it was and wasn't. I think the CSAC will have some challenges with that if it is used as a proxy for SDS. So I think with the comments that were made -- and I can't remember who made them down there -- about just very strong caution to the developer, to -- I think somebody said they wouldn't vote it down, but very strong caution that you come back soon with --

CO-CHAIR FLEISHER: So the irony is, if we weren't in the SDS period and this was approved years ago, this was probably in the

model, but because we are in the SDS period, it has actually popped up as a potential surrogate -- although we heard David's comment that this was not originally put in the model for surrogate purposes. So it is a -- just to articulate -- it is a confusing question of how much do we buy into the genetics or severity of disease argument. But we are now in SDS where there has been a -- in a trial period where there is a very clear signal that that is not acceptable, but it was when we first looked at this measure because it wasn't considered an SDS variable. It was considered --

MS. MUNTHALI: Right, because we thought at the time that we would be masking disparities, as many do on the other side. So -- but we are in this period now. So this is what we hold. This is the standard we hold all outcome measures to that come through NQF.

CO-CHAIR FLEISHER: So what I am hearing is that if we decide to approve it, that we are giving a very clear signal that it

probably wouldn't be approved in the next time, and that is only an if. That we want them to look into this issue.

MS. MARINELARENA: Sure. So when you are considering risk adjustment, you are looking at the threats to validity. So does -- do these two risk factors, are they a risk to -- a threat to the validity of the measure?

DR. BADHWAR: May I raise one point? Just to add to the confusion slightly. That if we consider, say, African American race to the odds ratio of 1.27 -- I think as Sean had said -- it is a statistical input. It is not being used for racial profiling or that type of more egregious type of effort. However, if you think about it in a different way, if the concern is that a provider may see an elevated risk for a -- a race and maybe have risk aversion, which is the issue, technically if that is a higher risk for a certain race, it actually advantages by risk adjustment to take care of that person. So --

CO-CHAIR FLEISHER: That is actually

-- we should stop, but that actually is not the argument of why we said it shouldn't be in. What it says it is masking that a hospital can do worse in an African American patient -- for example in this case -- because it gets risk-adjusted out. It is not the argument that somebody -- so, I think you are confusing why we made a clear, concerted effort at the CSAC of saying why we didn't want race placed there --

MEMBER YATES: In the past.

CO-CHAIR FLEISHER: In the past.

MEMBER YATES: Right, that was the paradigm of the past and now the paradigm is, at least present data to show that it shouldn't be a risk adjustment.

CO-CHAIR FLEISHER: But it should be stratified --

MEMBER YATES: Right.

CO-CHAIR FLEISHER: To show us. We should vote.

CO-CHAIR GUNNAR: All right, can we move on to validity? And just to torture this

one a little bit, how do -- does this come to a note to our decision? Whatever the decision is? Let's vote, and then we will get to that next step. So we are voting on validity.

MS. KOSURI: Voting is now open.

(Pause.)

MS. KOSURI: We are still waiting for a couple more votes.

(Pause.)

MS. KOSURI: Okay, we have the 14 votes. One of our members had to step off, so -

CO-CHAIR FLEISHER: Can I make a motion -- I don't know how the committee -- that we get stratification by race in the maintenance at one year? Because I think there is enough concern ---

PARTICIPANT: Was consensus reached?

MS. KOSURI: No, it is 64 --

CO-CHAIR FLEISHER: It is reached, but there is lots of concerns.

(Simultaneous speaking.)

MS. KOSURI: We beat 60 percent. Do

you mind if I announce it for a second?

MS. MUNTHALI: To Lee's motion, the reason he mentioned a year -- it is because it is the annual update of the measure, so we want to make sure it is within process. We will work with the SDS to make sure this happens -- and looked at by the committee.

CO-CHAIR FLEISHER: Does the committee agree with my motion?

MEMBER SAIGAL: Second.

CO-CHAIR GUNNAR: Just -- for those on the phone, if you would like to vote no, just email your vote in. Any -- just hand vote here -- anybody in opposition?

Seeing none, I think we have -- the recommendation from the -- in within a year, back from the developer.

MS. KOSURI: So I just wanted to --

(Simultaneous speaking.)

MS. KOSURI: So for measure 2561 for validity we have nine votes with moderate and five votes with low for a total of 14 votes. And

that is above the 60 percent. It is 64 percent, so it passes.

CO-CHAIR GUNNAR: So this passes on validity with, in my opinion, clear question to it by the fact that it is 30-some percent low. And a request -- a strong requirement that -- appended to this decision that the developer come back with how they are going to address this issue going forward. So we can move on to use -- is that is what is next? If I am --

MS. KOSURI: Composite construction.

CO-CHAIR GUNNAR: All right. Dr. Cima, any --

MEMBER CIMA: No.

CO-CHAIR GUNNAR: Any comments from anyone?

(No response.)

CO-CHAIR GUNNAR: Can we carry on to the vote? And this should --

(Pause.)

MS. KOSURI: Voting is now open for the composite construction.

CO-CHAIR GUNNAR: Yes.

(Pause.)

MS. KOSURI: Voting is now closed with -- with six votes for high and eight votes for moderate, for a total of 14 votes. The composite construction part of measure 2561 has passed.

CO-CHAIR GUNNAR: So the -- you announced, it passed. We move on to -- now do we get to use?

CO-CHAIR FLEISHER: Just for the -- to help the room think about for the -- for all the other measures we are going to look at, we discussed the fact that any measure that is built on this measure, if the votes are going to be the same, we will be able to ask if anyone wants to pull the measure for individual voting, which there is no problem doing -- that is actually the approach we take for the math. But if you think you are going to have the same questions about any other of these measures, we use these votes and carry them to the next discussion.

CO-CHAIR GUNNAR: That was jumping

ahead.

CO-CHAIR FLEISHER: Yes.

(Laughter.)

CO-CHAIR GUNNAR: Let's -- let's -- we are just trying to get through our first measure by 11:30.

(Laughter.)

CO-CHAIR FLEISHER: And our second measure by 11:40.

CO-CHAIR GUNNAR: All right. So where are we headed next? Feasibility?

MS. KOSURI: Yes, voting is now open for feasibility.

CO-CHAIR GUNNAR: Dr. Cima, any comments before we move on to ---

MEMBER CIMA: No.

CO-CHAIR GUNNAR: Any comments in the room? Dr. Cima, we can carry on and vote.

MEMBER CIMA: Vote -- let's vote.

CO-CHAIR GUNNAR: Yes.

(Pause.)

MS. KOSURI: Waiting on one more vote.

(Pause.)

MS. KOSURI: Okay, with seven votes for high and seven votes for moderate, with a total of 14 votes, we have -- for measure 2561 for feasibility -- has passed.

CO-CHAIR GUNNAR: Very good. Next? Usability, right?

MS. KOSURI: And now it is usability.

CO-CHAIR GUNNAR: So, use -- any comments, Dr. Cima? Anyone in the group?

MEMBER CIMA: No, it just goes back to the question that was raised quite a while ago about, is it being used the way we want it to be used? The public reporting aspect of it -- that was my main concern. That -- is it enough to just be what it is -- voluntary? Although a lot of people are doing it, and the star rating -- as we discussed, we are not here to comment on it -- but is that useful information? And we had that long discussion about it. That was my one concern as an individual, is -- are we moving the bar for usability? Or is it the same thing we

have been doing for the last decade with these measures?

(Pause.)

CO-CHAIR GUNNAR: Any further discussion? Impressions? Elisabeth?

MEMBER EREKSON: I would just say, to add to the conversation that we have already had, is perhaps the developer could go and actually conduct patient focus groups to get a little bit more information on what patients want to see on the website. And I think -- you know, I have been on the website this morning already, and there is definitely columns, and it is very clear to understand. You can definitely make the morbidity -- the composite measures look bigger, but still include some of that other data. The other comment I would have is that I feel that all surgeons feel that their outcomes are better than they are when those are good outcomes, and they have less of the bad outcomes than they actually do. Just like all surgeons, if you ask them what their volume is, tend to overestimate

how much they are doing. And so, to -- and that is just a truth in surgery because we all want to do a good job by our patients -- I think to have the onus for public reporting go to the patient level where the patient has to ask her or his surgeon what their -- their data is, takes away from the public reporting. So I just would say that I like some of the pathways that have been proposed. I think we have other measures that are QCDR measures, and that is how we are talking about public reporting. But it is something that -- but I don't think the onus should be on the patient and the individual surgeon because those data are not as good as what a transparent public reporting is.

CO-CHAIR GUNNAR: Dr. Stein and then Rick.

MEMBER STEIN: I have a question for the developer, did -- for the 67 percent that are publically reporting, since you -- since you know where they lie in the zero to 100 percent, are they -- do they tend to be the highest 67

percent? Or -- have you looked at that?

DR. PAONE: So the answer to that is we see what the public sees. The STS is a voluntary -- participation in the database is voluntary, public reporting is voluntary. The data is housed at the Duke Clinical Research Institute, not at the STS. And so we don't know of the 1,091 centers, which are three stars or two stars for any particular outcomes. That is not information that we have. That information is obviously with the institution. So of the -- so, on the website, 67 percent report -- or 69.9 percent now report. The star ratings are available for them individually. If you go to that website, you can actually organize them in -- depending on how you click the categories, you can have them pop up in order from three stars for mortality or three stars down -- or click them again and the one-star programs will pop up to the top for the individual categories. But if they are not publicly reported, STS has no way of knowing. In fact, I can tell you a year and a

half ago there were -- when we added the five -- when we added the composites for the surgery, there were three programs in the country who were three stars for all five categories. We had no idea who they were and had no way of finding out unless they came forward -- and one did and so we knew who one was. That is just -- again, the nature of the voluntary aspect of this.

DR. SHAHIAN: Let me expand on that question, though. We did actually look at this as a research endeavor. And I presented this to the American Surgical Association about three years ago. We had access on a research basis to all the programs in the country, and we looked at the performance of publicly reporting programs versus non-publicly reporting programs. You can go to PubMed and you can see the abstract. But very striking that publicly reporting programs uniformly for all nine reporting periods that we studied between 2010 and 2014 -- every single reporting period -- publicly reporting programs had lower risk-adjusted mortality and higher

composite scores for every rating period. And again, you can go to PubMed and see the detailed results.

MEMBER DUTTON: So obviously you can slice that either way around, but it is interesting. I have a different question for the developers about the science of this. Obviously STS existed for a long time before the public reporting started as a quality improvement tool for surgeons and hospitals and you have -- I believe you have documented substantial improvement over that period with the measurement and reporting as a quality improvement tool. How did that change when you started public reporting? Did the vector of change defer? Do -- did it cause a drop-off in overall improvement, or an increase? In other words, did the public reporting help?

DR. SHAHIAN: Behind that question is the presumption that the main goal of public reporting is improvement, and some people would argue it is the only way to improve. In fact, I

think there are many examples of improvement without public reporting -- one of the best of which is the Northern New England Cardiovascular Disease Study Group. And there are many other examples.

We think the main reason to do public reporting is public accountability and transparency. I think there are many ways to improve. And we have -- you know, we were improving before public reporting, and I don't -- as I recall, for most of the measures, the trajectory for improvement did not change dramatically when we started publically reporting. Because we have been working on this for years before and after. And those graphs are actually also in that article if you want to look at it. But in most cases, you know, we have just been improving steadily over the years.

DR. BADHWAR: If I can comment to enhance Dave's comment. At the institutional level, if you think about the overarching objective of what public reporting is -- and the

aspects of, once you sign up to voluntarily public report -- the arc of improvement is increasing because everybody is trying to get to that two- and three-star rating. And by being able to compartmentalize each of the subcomponents of where you are not three stars is the, you know, issue du jour at every cardiac surgical program that is publicly reporting -- to define, how can you improve? And so lessons through some of the task forces -- one of which that Gae chairs -- is how do we take those lessons from those three-star programs and share it in terms of quality improvement? So I would -- I would say to your question that yes, there is a gradual improvement at multiple sites that are participating in public reporting.

CO-CHAIR GUNNAR: So we are at use. Any other discussion or comments around use?

(No audible response.)

CO-CHAIR GUNNAR: If not, then we will vote.

MS. KOSURI: Voting is now open.

(Pause.)

MS. KOSURI: Okay, for use, with a total of 14 votes, we have 12 who voted to pass and 2 who voted to not pass. So this -- this -- it passes for measure 2561.

CO-CHAIR GUNNAR: Very good. Our next is usability. Any discussion about usability?

(No response.)

CO-CHAIR GUNNAR: Hearing none, it is open for a vote.

(Pause.)

MS. KOSURI: Waiting for one more vote.

(Pause.)

MS. KOSURI: Okay. With a total of 14 votes, we have 2 votes for high, 10 votes for moderate and 2 votes for low. So for measure 2561, usability, we have passed -- the measure has passed.

CO-CHAIR GUNNAR: Very well. Now we move to overall endorsement. Any further discussion, comments before we vote on this

measure for endorsement?

(No response.)

CO-CHAIR GUNNAR: recognizing the discussion and caveats that we have already come to -- very good.

(Pause.)

MS. KOSURI: We are still waiting for two more votes.

CO-CHAIR GUNNAR: He will certainly present the amicus brief at the --

MS. KOSURI: For measure 2561 for overall suitability, this measure has passed with 14 votes for yes out of the total of 14 votes.

CO-CHAIR GUNNAR: Okay, and -- and I think we have given appropriate direction as to what needs to be the addendum to this decision? All right, very well. Now, our next is virtually the same. So the next is 2563. It is STS -- is that correct? STS aortic valve replacement and coronary bypass graft composite score. And the discussants are -- Allan is not with us, right?

MEMBER SIPERSTEIN: Yes, I am here.

CO-CHAIR GUNNAR: Allan, great. So let me ask, is there anything you would like -- I mean, it is virtually the same -- it is identical to the process we just completed, right? It -

MEMBER SIPERSTEIN: Correct, yes. So -- but first I just want to put in my two cents on a lot of that last discussion. I think the whole discussion on race and socio-demographic is very, very important. And just to editorialize, I think it just reflects our maturation in terms of measures and outcomes. Whereas previously we were ecstatic to be able to risk adjust mortality. And it simply shows the sophistication in what has gone on. And I think the real importance for the patient is to identify modifiable risk factors, both at a personal level and a societal level, and so are quality efforts to move that forward.

So that -- now that I have gotten my two cents in -- so exactly. This measure is exceptionally parallel to the measure we just discussed, simply for aortic valves with CABG

being done. My question -- I will try to focus this for the developer -- is obviously we are dealing with a lower-volume procedure with a higher overall mortality and complication rate. And I just want to get their input on how they think this impacts the validity and usability of the measure.

DR. PAONE: So I think that is a reasonable point to make. However it -- AVR CABG is -- fewer numbers than isolated AVR, although that may not be the case a few years from now with the advent and continued exponential growth of trans-catheter aortic valve replacement. But having said that, it is still the third most common procedure. It is -- and its complexity is such that it is a different procedure from the aortic valve replacement. And so I do think it is very reasonable to have a separate composite score for this. We have provided the graphs that show that -- from the composite have been updated and I believe they have been sent to the committee -- that demonstrates that there are

differences in the spread of morbidity and mortality outcomes for these two measures. And so I do think it is very reasonable to have this be a separate composite. And it is part of a group of outcome composite scores, which frankly are part of, now, the individual surgeon composite. So we need these mortalities particularly -- as well as for the hospital-based composite, which we are in the process of developing.

DR. BADHWAR: And one additional point to Dr. Paone's is that for isolated aortic valve replacement, these are often and increasingly more commonly performed as a minimally invasive procedure, and not an entire sternotomy, whereas AVR CABG -- it is -- it necessitates an entire sternotomy. So there might be some inherent differences, and hence the reason why these two measures are independent of each other. They are obviously parallel, but that should be a stand-alone measure.

MEMBER SIPERSTEIN: But again, with

the construct of those -- the procedure, the risk adjustment, the data collection -- exactly parallels what we have just been discussing for the last hour and a half.

CO-CHAIR GUNNAR: So do we -- can we -- do you want us to run through all the votes? Or does anybody think that they are going to change? Or would like the opportunity to change their vote from the previous measure? And is that -- are we allowed to do that, or not? Can we -- can we just move to pass or no pass on the -- or do you want us to run through all of it.

MS. MARINELARENA: So you don't have to vote for evidence because that's a maintenance measure. We are okay with that. Gap, if there is any issues with it -- because we do require that they submit new performance data. Staff did provide a rating of moderate. You can choose to accept that. You don't have to vote. Composite construct as well, you can choose to accept the rating that staff gave it. You don't have to vote. Reliability -- there was a rating of the

methods panel -- rated it as moderate. You don't have to vote. Consensus was not reached on this one as well -- same issues as the other one. We can carry the vote over and include it in the discussion. Feasibility -- you have to vote, but it is the same issue. We can carry it over. Same for use and usability. And then do a -

CO-CHAIR GUNNAR: I guess I am asking, is it -- it is probably easier just to forge ahead and run through the voting line. And then you've got it documented. And then same -- attachment or addendum to this decision as --

CO-CHAIR FLEISHER: I will ask it different. Does anybody want to poll any one aspect of what was just discussed by Melissa?

CO-CHAIR GUNNAR: But she just said we can't do that.

CO-CHAIR FLEISHER: No, she just said we could.

MS. MARINELARENA: You don't have to vote on every individual -- you can choose to accept the ratings that we provided.

CO-CHAIR GUNNAR: No, I want to accept the ratings that we -- I -- is there anyone who would change their vote in any -- can we go and carry the same votes forward? Not -- not -- not the -

MS. MARINELARENA: I think they are the same. Yes, they are the same.

CO-CHAIR GUNNAR: But -- but you only have the -- you will actually have the vote count then.

MS. MARINELARENA: Yes, okay.

CO-CHAIR GUNNAR: See what I am saying?

MEMBER SIPERSTEIN: Yes, Allan here, I feel comfortable using the same vote.

(Simultaneous speaking.)

PARTICIPANT: Agree.

CO-CHAIR GUNNAR: Okay, let's just see if we can run through this. Let's go through the first -- what do we have to do? Vote first is what?

PARTICIPANT: Thirteen.

CO-CHAIR GUNNAR: No, no, no. What's the first thing we vote on?

MS. SKIPPER: The first criteria is evidence.

CO-CHAIR GUNNAR: Okay, would anyone change their vote on evidence if we take that -- the answer is no. So can we adopt that what we did for the previous measure -

MS. SKIPPER: I just want to make sure everyone in the room understands what we're -- and it's okay? Okay. I am just looking at -

(Simultaneous speaking.)

CO-CHAIR GUNNAR: Anyone who has a -- has a -- if there is one vote who is not comfortable with this, we will go -- start from the beginning, okay? I don't have a -- all right.

(Simultaneous speaking.)

CO-CHAIR GUNNAR: Yes, exactly. Next, so evidence passed same vote percentage carried forward. What is next?

MS. SKIPPER: Gap.

CO-CHAIR GUNNAR: Gap. Anyone changing their gap vote for this measure versus the previous one?

(No response.)

CO-CHAIR GUNNAR: None? Or would like to take that vote?

(No response.)

CO-CHAIR GUNNAR: No, hearing none, next. Adopt that.

MS. SKIPPER: Composite construction.

CO-CHAIR GUNNAR: Composite construction. I don't see any -- so we will just adopt whatever the vote percentage and say -- every -- no one would change their vote. Next?

MS. SKIPPER: Reliability.

CO-CHAIR GUNNAR: Reliability. Any new discussion on reliability? Anything different from what we've --

(No response.)

CO-CHAIR GUNNAR: Very good. Next?

MS. SKIPPER: Validity.

CO-CHAIR GUNNAR: Validity.

(No response.)

CO-CHAIR GUNNAR: Hearing none, move forward.

MS. SKIPPER: The composite. Validity composite.

CO-CHAIR GUNNAR: Oh, yes. It -- same vote.

MS. SKIPPER: There -- you have to take three separate votes under scientific acceptability for a composite.

CO-CHAIR GUNNAR: Okay.

MS. SKIPPER: So now we are looking at the composite construction.

CO-CHAIR GUNNAR: Anyone want to change their composite construction vote?

(No response.)

CO-CHAIR GUNNAR: Or would like to re-vote on it. That's a bad way to put it. Anyone would like to re-vote on that -- on that measure?

(No response.)

CO-CHAIR GUNNAR: Okay, next?

MS. SKIPPER: Feasibility.

CO-CHAIR GUNNAR: Feasibility?

(No response.)

CO-CHAIR GUNNAR: Hearing none, carry one.

MS. SKIPPER: Use.

CO-CHAIR GUNNAR: Use?

(No response.)

CO-CHAIR GUNNAR: Hearing none.

MS. SKIPPER: Usability.

CO-CHAIR GUNNAR: Usability?

(No response.)

CO-CHAIR GUNNAR: And we will -- just for -- to make all of us better, let's go ahead and actually vote for whether we would endorse this. So let's take that vote.

(Laughter.)

CO-CHAIR GUNNAR: Yes. A.J., that's as good as it gets. I think we're -- we're -- as long as -- as long as we can do that, we are going to see a lot of that today. We are going to be done pretty -- there is so much redundancy in what we have to do for -- and we have seen

these measures twice over -- we have lived with these measures now for 12 years, right? All right. So here we are. I didn't get a -- I didn't -- it didn't sign me in.

(Simultaneous speaking.)

PARTICIPANT: If you could refresh, that would --

CO-CHAIR GUNNAR: Oh, here it goes. It finally came through. Very good.

(Simultaneous speaking.)

MEMBER JARRETT: And while we are voting I am just going to make an off-hand comment that anybody who said that they want to go through everything again would be like somebody from the old movie 12 Angry Men.

(Laughter.)

MEMBER JARRETT: For those of you who remember it.

CO-CHAIR GUNNAR: Actually, did it win that year -- Academy -- I think it actually won that year, didn't it? Did 12 Angry Men win Best Picture for ---

(Pause.)

(Laughter.)

MEMBER JARRETT: It was nominated for three, but it didn't.

CO-CHAIR GUNNAR: It did not, okay. There you go. Please -- it's been a couple years since I saw it.

MS. KOSURI: We are still waiting for two more votes.

CO-CHAIR GUNNAR: We need a couple more votes.

PARTICIPANT: Are we taking a lunch break?

CO-CHAIR GUNNAR: No, no we are going to keep -- we are going to -- lunch is going to be very brief, okay?

PARTICIPANT: Well, it is a working lunch.

CO-CHAIR GUNNAR: All right, we've got 14 votes. Very good. It passes and we can move -- do -- do you want a quick break?

MS. KOSURI: We are just going to read

the votes off for one second. So for measure 2563, the committee has voted to pass this measure for overall suitability for endorsement with 14 votes out of a total of 14 votes.

CO-CHAIR GUNNAR: Go ahead.

CO-CHAIR FLEISHER: So we are going to -- what we thought is, we would take a break till 12 -- or a little -- and bring -- five after 12? Take like a 20-minute break? Twenty-five minute break? Let's just -- we can do 12, okay. Amy, you were getting me concerned. We are going to take a break till 12 to bring our lunch back, stretch and we can have a working lunch so that we can keep moving, since this is the committee of surgeons and anesthesiologists. Sorry for the non-surgeons and anesthesiologists. We are used to this.

(Whereupon, the above-entitled matter went off the record at 11:43 a.m. and resumed at 12:05 p.m.)

CO-CHAIR FLEISHER: So we are rejoining the discussion and we are going to skip

a couple of measures because of the availability of someone to lead the discussion, correct, Christy? So do you want to tell us what's going on and keep us on track?

MS. SKIPPER: Okay, so we've moved the discussion of 0122. There is a discussant who has to step away for a surgery, so we'd like to discuss this measure. Also we'll have quorum, so it just made sense to go ahead and move it around and discuss this measure, so thank you all for your flexibility.

And I think Dr. Handy, I don't know if you can come back to the line, but you, Dr. Yates, and Dr. Scali are discussants for 0122.

MEMBER YATES: Are we waiting for Dr. Handy?

MS. SKIPPER: But if one of you could go ahead and start, and he'll chime in?

MEMBER YATES: Well, how about before we start, should we let the developers present it? Isn't that usually what we do, or is it because it's for re-endorsement and we're going

to just press ahead?

MS. SKIPPER: That is usually what we do. Thank you. So Mark and team, could you tell us about 0122?

MR. ANTMAN: So I'm happy to introduce this measure briefly. Again, 0122 is the risk-adjusted operative mortality for mitral valve replacement plus CABG.

This is the operative mortality measure that constitutes the mortality domain for the mitral plus CABG composite, which is that mitral plus CABG composite is NQF endorsed and it is publicly reported.

I'll note that just skipping ahead a bit, this committee will also be looking at 01, sorry, 1501, I'm sorry again, 1502, which is the operative mortality for MV repair plus CABG.

The MV replacement and the MV repair components are actually combined in the mortality domain of the mitral valve and CABG surgery composite, hence our reference to that composite as the MVRR composite, so just to explain why

mitral valve replacements and mitral valve repair are separated out.

I'll defer to my surgeons if they want to add anything further. Okay, thank you.

MS. SKIPPER: All right, so if there's nothing else to add, then our lead discussants, Dr. Handy, Dr. Yates, Dr. Scali, you can start with the measure discussion.

MEMBER SCALI: Yeah, I'm happy to take us through it. So thank you for that introduction. Just to clarify, the level of analysis for this measure is clinician groups and facilities. This is a maintenance evaluation. It was originally endorsed in 2007, re-endorsed in 2015.

The review of evidence was not required. However, the developer did provide ample references from the literature that documents that this particular procedure has one of the highest known mortality rates for all surgical procedures.

There was a question for the committee

to consider in the pre-work up by the reviewing group from NQF. The question for the committee is, "Is there at least one thing that the provider can do to achieve a change in the measure of results?"

And I think that sort of goes to the fact that as they're getting feedback on these measures quarterly, and so I think that they can obviously act on these results to sort of improve patient restratification and patient selection, and certainly the patient counseling.

And the second question to the committee was, "If derived from a patient report, does the target population value the measured outcome and is it meaningful?" and I think the answer to that is yes given the fact that it's looking at mortality.

So I don't know if there's anything else that you wanted to add, A.J., in terms of evidence?

MEMBER YATES: I have little to add. One of the comments was that someone had

mentioned the fact that it presents small numbers making it harder to detect real differences in performance.

However, the mortality of the rate gives it gravitas enough for the work group, the people who were doing this basically a recommendation, endorsement.

MS. SKIPPER: And I believe Dr. Handy joined the line. Is there anything you'd like to add, Dr. Handy, to evidence, or share of the evidence for this measure?

MEMBER HANDY: I just logged on, so I don't know what A.J. said, but I'm assuming you guys are cooking through these things here since my schedule had that this was to be talked about around 10:00 my time here, so you guys must have considerably abbreviated the discussion a la some of the preamble. Is that true?

MEMBER YATES: You have no idea.

MEMBER HANDY: Good, I'm not against it.

MEMBER YATES: I move to vote.

MS. SKIPPER: All right, did I miss someone? Okay, we're now voting on evidence for measure 0122.

MS. KOSURI: Voting is now open for 0122.

MS. SKIPPER: And we see that we have 100 percent votes. We just want to read the number for the record if we can, so just bear with us for a second.

And so I suppose while we figure that out, if we could move onto discussion of gap.

MEMBER SCALI: So pertaining to gap, the developer did provide comparative data from the STS database between two different reporting time periods of 2011 and 2014 compared to 2014 and 2017.

They describe the odds ratio greater than one being undesirable and that it implied participant increases of patient risk for experiencing the outcome if they have a value greater than one. There was a value of eight to 13 percent on average with a median of 9.1

percent in the era of 2011 and 2014.

This was relatively static in the era of 2014 and 2017 still demonstrating that there is evidence for improvement and there is variation in the outcome, and I think the committee had given a moderate gap rating, which I think was appropriate.

MEMBER YATES: Nothing to add.

MEMBER HANDY: I agree. That second to last comment is mine. The mortality disparities also were evident across sex, age, and race/ethnicity.

MS. KOSURI: We're just going to announce the evidence, but for one second. So with 14 votes, out of a total of 14 votes, the evidence criteria has passed for measure 0122.

MS. SKIPPER: And now we'll take a vote on gap in just a moment for 0122.

MS. KOSURI: Voting is now open for measure 0122 for performance gap. Okay, with a total of 14 votes, we have four votes for high and, oh, 15 votes now. We have four votes for

high and 11 votes for moderate for the performance gap criteria for measure 0122, and so it has passed.

CO-CHAIR FLEISHER: I think we're going onto reliability next. So for reliability, the score is the rate or proportion of cases that are meeting the criteria of the specifications and numerator statement.

The numerator statement is the number of patients age 18 years or older undergoing combined mitral valve replacement and CABG who die, including both all deaths occurring during the hospitalization in which the operation is performed or even after 30 days, and then two, deaths that occur after discharge from the hospital, but within 30 days of the procedure.

The denominator statement is all patients undergoing CABG MVR. The exclusions are none. The data source is a registry, the STS registry.

The question that was posed to the committee was, "Do we have any concerns about

whether the measure can be consistently implemented?" Personally, I do not. It's already being used and it's already being implemented at centers for quality improvement.

Relative to the reliability testing, this was a maintenance measure, so less emphasis was on this, but there is no score level testing that was done, so the highest achievable rating was moderate.

This is a mortality end point, so certainly it's a highly repeatable outcome measure. It's very clear and very unambiguous.

The developer provides agreement rates from a 2014 audit of 108 STS adult cardiac surgery participants and demonstrated an overall agreement of 95.73 percent.

Again, the level of analysis was there were two potential levels and there was a question about clarification from the developers about clinician group versus facility.

This was a question that was brought up in the morning discussions about how the

overwhelming majority of centers are now at the facility level because those are the people who are in the STS database. It's really the minority of clinician groups that are participating. It's really at the center level.

And then the panel's recommendation was moderate reliability, which again seemed prudent given the caveats already given.

MEMBER YATES: Again, it's a measure that's known to us and we've seen before, and for re-endorsement, I don't see any major challenges to the reliability or validity, and I will defer. I would argue that the debate we had previously on STS should hold for this as well and not be reopened.

CO-CHAIR GUNNAR: So I have a question though. Have you examined the 31-day, you know, this concept that you have somebody alive until day 30, but only to discharge them from the hospital and get them home to beat the metric?

Have you examined post-30 days and post -- particularly if somebody has now lingered

in the hospital because your definition isn't just 30 days. If they stayed 45 days and then died, you're still counting them. So the question is are they discharged? Is there evidence that you discharge people past 30 days?

DR. BADHWAR: So if I can just interpret your question, I think the clarity is that the measure is 30 days or in hospital. So if a patient is there for three months and they die in hospital, that still counted.

PARTICIPANT: Correct.

CO-CHAIR GUNNAR: But are you finding evidence that people, to game it, then discharge somebody off the acute care ward to beat the mortality? Do you have any evidence? Have you ever looked at that?

PARTICIPANT: We don't have evidence that further clarifies that type of gaming if that was to occur, at least in these measures. If you're asking in general across the board, that's a more detailed question, but at least in this special, we don't have that information.

MEMBER JARRETT: Yeah, this is Mark Jarrett. I had a question and on my measure, it's the same issue on mortality, that it's not even trying to gain the system, but the reality is the patient may spend two months in the hospital.

They're really a train wreck. They're not getting better and it's decided to move them to, you know, a nursing home, and, you know, and they succumb two weeks later, but now it's more than 30 days after the surgery and it's right after their discharge from, you know, and they've already been discharged from the hospital, but clearly their death is related to their complications following the surgery, and that's a reality that can really happen. So it's not a matter of gaming. It's just missing mortalities.

DR. PAONE: I don't know if we have a way of capturing that information specifically. I think we have in the past looked at the number of deaths and the reliability of in hospital 30-day mortalities versus the percentage of

patients, of data sets that had increasing numbers that were not reported for the out of hospital deaths.

This is an issue for certain for every single mortality measure, the possibility of, as you say, either gaming or just the natural history of some very sick patients, particularly with these LTACs where patients get transferred, as you say, 40 days, 50 days, 60 days to a facility and then die a week or two later.

I would suspect, and I think we've looked at this, that the number of patients that fall into that category is quite small, so even if we're not capturing them, I don't think it changes the outcomes to any specific degree, although it's an ongoing problem and it's an important issue to pay attention to, and there is the need to follow up to 30 days post-discharge, and so that would capture some of those that don't have that information.

Now, you need to have 98 percent capture on that metric in order to be able to

receive a star rating, so there's been some significant improvement in how many of those are captured.

MEMBER YATES: It wouldn't answer the question directly, but as a surrogate, do you keep track in the registry as to discharge disposition?

DR. PAONE: The answer is yes. There is a category in the database for where the patients are discharged.

MEMBER YATES: Because you would then be able to assess the risk of loss information by looking at discharge after 30 days to a skilled nursing facility or hospice as opposed to home, and if that occurs only one percent of the time, then the data, the bleed out on the data wouldn't be that much.

MEMBER JARRETT: Well, this is Mark. I agree with that, but I'm still concerned because even though the numbers are small, the numerators for especially some of the mortality statistics on some of the other measures are also

small.

So adding one or two to a given facility or a given group might really truly impact their standing. So I'm just saying I think that's something you need to look at going forward.

DR. BADHWAR: So this is a very important constructive comment. A couple of things just as a broad issue, as Guy just mentioned, we're really trying to police the mortality variable at 30 days for all participant sites to make sure that they have contributed that, and almost making the record incomplete in some of the measures if they don't provide a mortality number such as missing this.

The second thing, and this is not particularly relevant to this right now, but we're in the process of looking at NDI through a research mechanism and mapping that as part of a research enterprise to see if we can use that to validate 30-day outcome to make sure that it's accurate.

MEMBER GROVER: Let me just say at one point we were hoping to use the Social Security, and an attorney in the Social Security is dealing with another issue not related to any specialty or trying to get data has blocked that, and that way we could have 100 percent. I mean, if you wanted to, you could have a 60-day and six months even to verify it.

So if there's anything that, I can get you more details on that, or if there's anything NQF could do as a group to lobby to reverse that decision, that would be very helpful because that's really stupid and it's hurting, I think, every database in terms of seeing how the outcomes are.

DR. PAONE: Perhaps unrelated to this, but I think we all recognize the growing need for, you know, value in our outcomes, and so there's a lot of work going on now in trying to figure out how to do one year outcome not just 30-day outcomes, and that question goes back to the Social Security Death Index and access to it,

so.

CO-CHAIR FLEISHER: Right, so I actually want to commend STS because we had a debate several years ago and our concern about the CMS mortality measure was that it was not in hospital plus, and for those who were here, we actually preferred this over the CMS, that we need Bill's paper and continue to do that research, but I acknowledge without that Social Security link, that makes it difficult, so I want to say thank you.

If you get that data, it's great, but it may be worth scanning the literature, including my colleagues' papers, to put it into the data given that you're not linked to any kind of death registry.

MEMBER JARRETT: This is Mark. Just on the issue, just for everybody's information on death registries, we've been looking at it for our own mortality around CMS stars and everything else.

Unfortunately, it turns out, New York

state has not been reporting for over three-and-a-half years to the Social Security Administration their deaths.

Now, how Social Security knows how to stop the checks, don't ask me, but they do not report it officially into Social Security, and I don't know if any other states do the same thing.

MEMBER DUTTON: I just wanted to throw out, Fred, this would a great issue for the NQF to take on because everybody would benefit from it. I've tried to get Social Security master death file information as well, and it exists for credit card companies, not for science.

CO-CHAIR GUNNAR: Any further discussion? If not, we can --

MS. MARINELARENA: Before you vote, because the question was asked through multiple, on all of the measures and it was posed to the Methods Panel, does the committee have any issues with the fact that the measure is specified at two different levels of analysis?

I know that STS provides an

explanation of why, because of, you know, most surgeons are in the hospital. Is that sufficient for you? And in the future, we can ask them to, you know, clarify that up front, or do you recommend in the future, do you think that they need to provide an analysis at the hospital level?

CO-CHAIR FLEISHER: Fred, this may be actually the time to talk about the influence of the hospital CEOs that you were discussing because I think reporting at the hospital level probably has the most profound effect on improvement, correct?

MEMBER GROVER: Is this on? Yeah, they're kind of two different issues, but when we originally set up the STS database back in 1989, it was -- and it's different from the ACC because the ACC is with the hospital or the institution. We made it for our members.

And at that time, there were a fair number of hospitals that had more than one group of cardiothoracic surgeons operating in them,

mostly obviously in the private area, not the university area, but, so for that reason, we made it group specific, so it goes to the surgeon.

Now, the hospitals see the data. They're all shared because they're obviously picking up at this point the cost of the data manager, so it's all shared now, but that's how we got on that.

But in practicality, I'm not sure what percent of, how many hospitals now have more than one group. That would be something we could try to figure out.

I don't know that I -- I think it's probably so small that it's probably immaterial. I mean, it's probably not worth fretting over at this point. There are bigger issues to take care of, I think.

DR. SHAHIAN: I think this is largely a semantic issue, Dr. Fleisher. It's a surgical group, and in some cases, the surgical group is organized within a hospital. In some instances, it may be a private group in multiple hospitals,

but it would probably help all of you if we were simply to say, "This is a measure designed for a surgical group."

And I think in terms of testing, it doesn't make, frankly, it shouldn't make any difference whether that surgical group happens to be organized within a hospital or a private practice group working in multiple hospitals. We get all of their cases, and that's the critical issue.

MEMBER YATES: Just to follow up on that for clarity, earlier we heard that it's possible for MIPS reporting for several of the measures. Do you capture the surgeons' MPI and/or their 10?

DR. SHAHIAN: Yes.

MEMBER YATES: All right, well, if they have the MPI and 10, then the 10 would give you an idea as to what groups are which within the institution and across institutions if they go to more than one hospital.

DR. SHAHIAN: And if our composite

report, if you go to our website, you'll see that even if a group is reporting as a, and Mark can correct me on this, but even if you're reporting in one or the other of those two methods that I just indicated, you have the option of actually getting --

We've actually gone through that exercise of taking multiple groups and combining them at the hospital level for the purposes of reporting, so we give you the option both ways.

DR. PAONE: If I could just sort of try a real world quickly here, I'm from Michigan. I participate in the Michigan Society of Thoracic and Cardiovascular Surgery, which is a quality collaborative of 33, all 33 hospitals, all 33 public, non-federal hospitals that do cardiac surgery in the state of Michigan.

And two examples, there's one hospital that has two separate groups operating in it still, and they report as one hospital, and then there are several hospitals, there are two surgical groups that operate at several different

hospitals and their results get reported to the state collaborative per hospital, not for the surgical group.

Now, I suspect there may be circumstances still, although very few, where a surgical group may operate at three or four hospitals and collate all of that into one group. I don't know that for certain, but I would, as Dr. Badhwar said earlier, I think that would be decidedly unusual at this point.

CO-CHAIR FLEISHER: I'd actually be curious, both Elizabeth and your comment, and to mine and yours, from a patient perspective, and I'm not talking about today, but from a patient perspective, do they want to know is that group going to that hospital, how you make a decision getting back to the focus group that was previously suggested, or is it you go to the surgeon and you don't care which hospital they take you to?

So, you know, I know that's a -- your database was set up for one reason. We're now

asking for public reporting. So any thoughts on -- because it's not testing in the way I think people are asking, which is which hospital should I go to and surgical group?

MEMBER EATMON: From a patient's perspective, I know when I was looking for my surgeon actually in Atlanta, I chose a particular doctor, but did not realize that I had a breast surgeon and a reconstruction surgeon, and one surgeon could operate in one hospital group and then other surgeon didn't have the credentials to operate in the same one.

So then at that point, from a patient's perspective, who do you drop off? Who do you need the most? But overall, it's about accessibility.

If you don't have the accessibility to make that decision, which I found through a lot of other groups that most people don't have that accessibility, and then when you start talking about people that are not from metropolitan Atlanta that are in rural Georgia, that

accessibility goes even further down.

And just like I spoke about, we're going on getting research. We just want the doctor, but then when we find out that they may not be part of that group, that has a different effect on us.

And it kind of brings on some other stressors and things of that nature because you think you're going to this one group that's in the hospital with your example where there could be two separate groups, but when you get there, they're like, "Oh, on Wednesdays, only group B operates," and you need the surgery.

So it is imperative or important rather, from a patient's perspective, and accessibility is definitely key with some of those decisions that we make.

MEMBER MOYER: I would just throw out there we had a centers of excellence program around CABG, and the discussions that we had with our membership and with patients was they're really interested in a surgeon at a hospital

together.

Because they kind of felt like, "Well, I don't know what I should look for in a hospital or what I should look for over here, but I want the whole group working together and I want to know how they do together."

And for a lot of our groups, it is one group at one hospital, but I think we've got a fair amount of hospitals where someone, say, from an academic medical center comes in every once in a while, and I think that's a little different situation than that surgeon operating in their home academic institution.

CO-CHAIR FLEISHER: So can I make a suggestion independent of the vote that you may want to test it at the hospital level for any future --

Correct, it's not tested at the hospital level as was suggested, but it would be great for any kind of maintenance or certainly for the update to see the data that essentially, Amy, you're asking for. Are people in agreement

with that for --

DR. BADHWAR: Perhaps it's a comment to those questions raised. Obviously for cardiac surgery, it's a team sport and that's the whole principle behind that as opposed to gall bladder operations and things like that may not necessarily rely on critical care and nursing in a more robust way, and so that's the basis of all of our models.

And so just from a restoring faith in the process from a patient perspective, we as a society truly feel that it's yes, obviously the surgeon is of great importance, but so is the entire team, and that really drives almost all of our models and the principles in which we base them as a concept. I just wanted to get that out there from a patient advocacy perspective.

MEMBER GROVER: There are some old articles, I don't know whether they still apply, from the Northern New England Group back in the late '80s and early '90s where they found that, I think it was a small number, but some surgeons

got better results when they operated at one hospital as compared to another hospital.

And I don't know whether that's still true because there's not that many that do it now, but that is an interesting question because just as you said, it's a heart team. It's your anesthesiologist, your perfusionist, and the whole bit unless you carry them with you, unless you take them all with you to these hospitals, which would be awfully expensive for a practice in this day and age.

MS. MARINELARENA: So just a better clarification, from the NQF process perspective, the measure has to be the specifications and the testing have to match.

If the measure is endorsed in the way that it looks now, it would look like it's endorsed at the facility level and at the clinician level, so the public thinks that this measure, or will interpret it that it is endorsed at the facility level.

It is not because it's not tested at

the facility level. We can take that off and then it's implemented. The way it's used, it could be used -- you know, you could say it's reported this way, you know, and the explanation that STS has provided, how it's used with surgical groups is fine, but it cannot be specified at the facility level until it is tested, and, you know, we've captured that recommendation, and STS can, you know, look into that if they'd like.

MEMBER YATES: I move that we vote.

CO-CHAIR FLEISHER: Any further discussion?

MS. KOSURI: Voting is now open for measure 0122. Okay, voting is now closed for measure 0122 for the reliability portion of the measure criteria where 13 people voted moderate and two people voted high. This means that 87 percent have passed the vote. With 87 percent, the vote has passed. I apologize.

CO-CHAIR GUNNAR: So validity next?

MEMBER SCALI: So with respect to

validity, I echo Dr. Yates' commentary about the validity discussion and some of the sociodemographic factors that are implicated in these models just like they were in the previous things that we did this morning, so I think we've had that discussion.

Nonetheless, there was new face validity and empirical validity testing of the measure that was performed by the developers. However, the Methods Panel expressed concerns about the face validity testing that was submitted since it did not meet NQF criteria. So I don't know if there's any additional comment about that.

MEMBER YATES: My only comment is that we have endorsed this before with the same questions and I don't see any reason to not endorse it this time or at least vote for this being acceptable.

CO-CHAIR GUNNAR: Voting is open.

MS. KOSURI: Voting is open. Voting is now closed with a total of 15 votes. Thirteen

people have voted moderate and two people have voted high for the validity portion of the measure criteria for measure 0122. This is 100 percent who have passed the vote.

MEMBER SCALI: Does that mean feasibility is next?

MS. KOSURI: Yes.

MEMBER SCALI: The data elements are routinely generated and collected by the healthcare team. The measure has already -- it's a maintenance measure. The data abstraction is done by a provider other than the person who gets the original information. Most if not all of the variables are available in the electronic health record.

There are moderate costs associated which we've all been aware of in the past, and again, the data collection is already in use, so I believe it's highly feasible because it's already happening.

CO-CHAIR GUNNAR: So my only comment is I know in the interim, since 2014, you've

added like a considerable number of data fields, particularly for the mitral valve, right? Is that correct?

Yeah, and so the question, are you getting any feedback from the poor individual who has got to put all of that information in because it's now pages, particularly for the mitral valve?

DR. BADHWAR: It's actually mitral valve and aortic valve. There are multiple fields that were added initially for research purposes, but I'll just give you an example within the details of the field.

It really surrounds around anterior leaflet and posterior leaflet repair, and for the field, I know Dr. Gunnar knows this and Dr. Grover, but there's substantial evidence to note the level of complexity when the repair involves both leaflets versus just one leaflet, that the outcomes are fair to be worse, either longer operative times --

CO-CHAIR GUNNAR: I guess I'm not

asking specifically about, you know, justifying why you did it. It's really about the, in a feasibility perspective, it's gotten to the point where any time studies on what it takes to actually complete one of these records now, because they're, I forget the number of pages and the number of fields. It's pretty stunning.

DR. BADHWAR: Yeah, there's about 400 data elements, and so it does take several hours for one data manager to enter the fields. So we have received feedback, frankly, and your impression is correct.

However, we do have fairly good adoption of those fields, particularly those that are doing these types of operations.

MEMBER STEIN: I have a follow-up question. When some of the sites are entering this information, I understand the electronic health record, some as a hybrid and some manually, is that correct?

DR. PAONE: The desire certainly is at some point going forward is to be able to

directly download as much information as possible from the electronic health record into the database. I think that's obvious as to why that would be.

Some data elements obviously lend themselves more to that than others, and we had this conversation on a conference call about a week-and-a-half or two weeks ago talking about the future of AI and the ability to intelligently read the reports and even download most of the information.

I suspect that's true to whatever extent various institutions use the health records that they have and the ability to transfer that information, so I don't know the specifics of that and I don't know if any of my colleagues do as to what percentage of centers are doing that.

Right now, what's most likely to be able to be downloaded from the electronic health record are sort of the easy things, which helps a little bit, but not a lot.

In reference to the question about increasing the number of data points in the recent spec upgrade of the database, the extra work is case specific. If you're doing a coronary bypass operation, there hasn't been much change in what's been added or subtracted.

On the other hand, the issues that are more technical related to aortic surgery, aortic valve surgeries, atrial fibrillation and maze procedures, and complex mitral valve repairs, and truthfully, the extent to which that causes more or less work is really specific to the institution depending upon the relationship between the data manager and the surgeon, how much data the surgeon enters directly into the worksheet and how much time the data manager may have to spend looking through the record or tracking down the surgeon.

So it's quite variable. It's an issue that we're well aware of and paying great attention to in trying to simplify the data requests going forward and also automate as much

as possible, but the truth is that's a little bit a ways yet before it becomes, I think, done in a manner that will actually make a difference.

MEMBER STEIN: So one of the other measures that I was reviewing, it said that some were, the data was coming in through EMR, some as a hybrid, and some manual, and I guess my suggestion would be that you may want to put in your risk adjustment model how the data is being entered because that may impact performance or proposed performance.

MEMBER YATES: As someone who is really concerned about data collection within my own institution and who values registry data far more than administrative data sets, a lot of it has to do with how you set up the data fields in your EMR.

You can have EMRs that contribute nothing to the STS or you can have EMRs that as things are entered from the outpatient arena or the inpatient arena, then automatically go into the STS database because it's been set up the

right way.

So a lot of that is going to be variable, but it doesn't mean that the data is not legitimate because it was hand entered, if you will, or physically entered at the point of contact, but those were many, many points of contact rather than somebody retrospectively trying to capture that.

The fact of the matter is that this is a tiered problem, and tautologically speaking, it is what it is because it happens, so it must be feasible because all of these hospitals are replying, but there's going to be, you know, there's going to be plateaus that you're going to reach, and right now having one person who is totally dedicated to making sure that your hospital's database is up to date for STS, that's one FTE that you're paying for.

If they were to double the amount of data that has to be collected, you're then hitting a different plateau, and so far, I don't think they've hit the second level of plateau and

they're fully utilizing the person that's already been assigned to make it happen.

So I would argue that we could argue about the fact that yeah, it's expensive to be in it, but the benefits from it have obviously been costed out by the hospital and they've decided that it's worth doing.

MEMBER STEIN: I think my concern is that the ones getting three stars, it's because they're entering the data in a certain way versus the two stars that are using some other approach, and that's why I'm suggesting to at least explore it, if not adjust for it.

MEMBER YATES: Well, that's why they have the 10 percent audit.

DR. BADHWAR: Yeah, I was just going to mention I think the discussion is interesting, of course, but however, we have a fairly robust audit process, and if the issue is going to be is the data accurate, there is a 10 percent audit process, and the outcome of that is in the 98th to 99th percent accurate in general.

CO-CHAIR GUNNAR: So let's hold that for validity. Let's vote on feasibility and carry forward, right, because we're still on feasibility, right?

MS. SKIPPER: We're on feasibility and we already voted on validity.

CO-CHAIR GUNNAR: Yes, sorry.

MS. SKIPPER: So, yeah, you're correct we're on feasibility.

CO-CHAIR GUNNAR: So we already voted on validity, so we can bring it up next time when we want to talk about validity again on the next measure 12 more times. Let's get to feasibility. Any other discussion on feasibility? And if not, then let's vote.

MS. KOSURI: Voting is now open for feasibility for measure 0122.

MEMBER STEIN: While we're voting, I just want to follow up. It's not a matter of accuracy. That's not the concern that I'm having. It's the ability to capture everything at granular detail of someone manually entering

in these in the EMR, entering it in, and whether that might be affecting the parameters.

MS. MARINELARENA: Use your mics, please.

DR. BADHWAR: Thank you. So honestly, personally, I hear you completely. You know, I think you raise a very important point. This is something that the data managers sweat about on a daily basis and communicate to us.

The real issue is the precision of the data collection is what you're commenting on, so I totally respect what you're saying.

In fact, STS as a whole is actually undergoing a full review right now of our data warehouse mechanisms and how can we make the user experience better, and this gets to Dr. Gunnar's point of the large number of variables and how we refine the process.

So, yes, some vendors are actually extracting data from the EMRs and supplementing it. However, some do not use that process. We hope to evolve the database much like this whole

discussion all day today as how are we evolving to make it more precise, and more accurate, and better for the user experience.

And we're actually working on that very diligently and actively now, including looking at the whole technological experience of data precision. So your point is very well taken and actually being addressed.

MS. KOSURI: Voting is now closed for the feasibility portion of the measure criteria for measure 0122. We had a total of 15 votes where 11 people voted moderate and four individuals voted high, which means that the committee has passed this portion of the criteria.

CO-CHAIR GUNNAR: We'll move onto use.

MEMBER SCALI: So with respect to use, there was some information that was provided by the developers about improvement over time. They developed reports on operative mortality rates. They were saying that they were trending slightly lower. The overall event rates in the last three

12-month periods were 9.9, 9.3 and then 9.45 percent.

They did not address potential unintended consequences such as sicker or higher risk patients being potentially shifted to other institutions, but did state that the benefits of tracking the measure outweigh the theoretical risk.

And then with respect to the usability, again, we hit the same snag we discussed this morning about the public reporting caveat. Technically in the write up, obviously the public accountability program was thought to be unclear, but public reporting is planned as of January '19 for the developer.

With regard to feedback, all of the end users get reports quarterly and they can provide assistance with interpretation, and users give feedback to the developer and the feedback is considered when there is changes that are incorporated on the measure. There is no facility level testing that was provided, so the

measure can only be used for the provider level.

And preliminary rating of usability was a no pass similar to what we saw this morning as it was originally endorsed in 2007, but, since it has not been an accountability application or in public reporting, but I believe it's part of a composite measure elsewhere, so again, this is the same discussion from this morning.

CO-CHAIR GUNNAR: Correct, so this is the first test that will carry us through all of the other, any other -- it's not part of a composite or is it?

MEMBER HANDY: It is.

MR. ANTMAN: And to add to what Dr. Scali said a moment ago, yes, the public reporting of the mitral composites was planned for January '19 and it did go live last month.

CO-CHAIR GUNNAR: So to bring back this morning, this decision here will set us on a path of one or two options? This one is being public, so we're good here, got it, yes, no challenge to this one.

MS. MARINELARENA: No, this measure individually is not being publicly reported. It is part of a component. It is a component of a composite. The composite is being publicly reported, correct?

CO-CHAIR GUNNAR: No, no, no, to clarify so everybody is on common understanding here, this, what we're voting on is a measure that when it was endorsed initially, knowingly was done so with the understanding it was going to be publicly reported in three years, correct?

When it was endorsed initially, the understanding between NQF and the developer was that there would need to be, this would have to be publicly reported within three years, is that correct?

MS. MUNTHALI: For six years or a plan for public reporting.

MS. MARINELARENA: So our criteria says --

DR. PAONE: Can I clarify just to make sure everyone understands what this measure, how

it's reported? It's reported as part of -- there's a mitral valve repair or replacement composite which includes mortality and morbidities for isolated mitral valve repairs and mitral replacement.

There is a second mitral valve composite for mitral valve repair or replacement with coronary bypass, which is what this MVR CABG mortality is reported as part of the mortality component of that. It's not separately reported.

CO-CHAIR GUNNAR: Correct, and this is a theme that's going to exist for all of the next measures that we approach. And as we discussed earlier, so this is a part of a composite that's reported, but not independently reported.

MEMBER YATES: But for clarification, it's only two components within the composite, the publicly reported composite.

DR. PAONE: The publicly reported composite for mitral valve repair or replacement/coronary bypass surgery, for instance, includes data for isolate mitral valve,

I'm sorry, for mitral valve repair/CABG and mitral valve replacement/CABG.

The mortality component of that composite includes mortalities for both procedures combined, not separately. It's probably not something I should be saying so clearly, but given this conversation, but it's the fact. The volumes are required in order to be able to become more reliable and more valid.

MEMBER YATES: So it's one degree of separation from being reported separately.

MEMBER HANDY: So this is John Handy and I'm going to show my enduring confusion. So when you go to the STS site and you look at the mitral valve CABG composite score as was mentioned earlier by my colleagues in the room, it's listed in three parts, overall score, absence of mortality, and then absence of major morbidity.

And I thought our sticking point this morning was the absence of major morbidity, each one of which individually, all five of them, are

NQF approved end points, but we're conglomerating them all together.

So actually if you look at the website, there is public reporting of the absence of operative mortality separated out from the overall composite score, and the absence of operative mortality as we're talking about right now is an NQF endorsed measure.

CO-CHAIR FLEISHER: Can I ask a question? So I heard a statement that's important, whether or not the robustness and the stability of individually reporting this metric is sufficient. If you report it, you're worried about the stability.

There is one thing to say that statement. There is another thing to actually document it because my question to you is we've heard today from our patient phasing that it would be great to report this out in a year.

And I'll actually make a proposal that within the year, can we get a plan and maintenance back that either you do report this

out or you give us good documentation of why there's unintended consequences of publicly reporting this independently that we could take back?

I don't know if that's a fair question because it says if you have a plan for public reporting, that's good, but I really -- I want more than --

CO-CHAIR GUNNAR: That's a new credible plan.

CO-CHAIR FLEISHER: Well, that's why I'm saying it's one year to sort of come up with documentation of why there's unintended negative consequences of publicly reporting this.

DR. SHAHIAN: Could I respond to that, Dr. Fleisher? We actually, in 2007, when we started doing composite measures, we had a two-part article called quality measurement in adult cardiac surgery 1 and 2, and that paper was for CABG, but similar arguments would apply to these other procedures.

One of the problems is as mortality

rates have fallen for almost all cardiac surgical procedures, trying to differentiate quality based just on mortality is very difficult.

We found that using mortality alone, that you could identify about one percent of STS providers as being outliers. When you used a composite measure, it went up to about 20 percent that you could identify as outliers.

That's right in the paper and that's 12 years ago that we showed that. I have no reason to believe that would be any different today.

That is, I think, the main rationale in my mind or one of the main rationales is just to have a better chance of discriminating levels of performance. It's a statistical issue related to sample size and number of end points available.

And then the other thing is that, as somebody alluded to earlier, if you're just looking at mortality, that's only one aspect of being a survivor. Most people want to know not

just if they're going to survive, but am I going to survive with a life altering complication like dialysis-dependent renal failure or a debilitating stroke?

So the composite provides a more multidimensional approach to quality, so if there's a statistical and a sort of qualitative reason for reporting the two together we think.

CO-CHAIR GUNNAR: So let me add to that again going back to the strategy here. The strategy here is if it is, my esteemed colleague, is passed with an addendum with a clear understanding within a year, that it's the message to STS.

I will propose the second strategy which is to fail. Let STS go back. And this is just not this measure. It's the other measures that make up. Come back to the committee with a justification and a plan forward that we can review and vote on before we endorse or make our decision forward to NQF, CSAC , and above, right?

The benefit of my argument for the

latter is STS may come back and say, "You know, let's rethink all of this. Maybe we don't need to report one and two or whatever publicly. Let's back off. Put our focus on the composites. We've justified their use and people are actually relying on them," as opposed to from a public facing point of view.

Alternatively, they could say, "Here is our plan and this is our plan going forward to immediately begin to publicly report, for those who are willing to publicly report the individual pieces of this, and here is our strategy to do that."

So those are the two options, I think, unless someone else has a different option, option three, but we're going to know, I think, if there's any -- is there any further discussion or a need for understanding? Amy?

MEMBER MOYER: I guess the only thing I'd throw out there, so for instance, I have the anti-lipid treatment discharge. I would not mistake anything on the STS site for a single

reporting of that measure.

CO-CHAIR GUNNAR: No, no, no, we're doing just this one.

MS. MARINELARENA: Let me finish. I do this for a living and I would not have noticed that extra R in your online reported measure versus what's NQF endorsed if we hadn't had this discussion. That's really confusing. I just wanted to throw that out there.

CO-CHAIR GUNNAR: Any other comment?

MS. MARINELARENA: This is Melissa. Just a process issue. The composites that this is a part of are not up for endorsement right now; you endorsed those in 2017. The measure under discussion is the one before you right now.

MEMBER SAIGAL: Can I clarify what we're talking about? Basically the issue is that these individual measures are being reported directly or rather as part of a composite that doesn't meet the criteria and what do you do about that; that's the issue. And personally I think that if they're being reported in some way

as part of a composite, that's good enough for me. So I accept that as a form of public reporting of them, even though I specifically call that individually.

DR. BADHWAR: May I just respond to this, because it's somewhat of a similar argument. And Amy, I hear your comment. The MVRR just for clarity because I think you had some blank looks, that's mitral valve repair or replacement, but to Melissa's point that that's a previous measure, that is being public reported. To Dr. Saigal's point, these are being reported to the individual sites, and like the previous argument to give a very clinical point, to improve quality at the site level, they need that information and that's how they're being much more focused on their replacements versus the repairs. So it's the same argument before, it's still applicable with this measure.

CO-CHAIR GUNNAR: Any other -- oh, Joshua, anything?

MEMBER STEIN: I don't know. I think

your last comment goes back to my earlier question which is if we want the developer to make changes to our suggestions, do we have to vote it down to be able to get that versus accepting it if they do make those edits.

MS. MUNTHALI: So you've voting on the measure as it's currently specified. So I don't know if that helps.

CO-CHAIR FLEISHER: And when does this measure lose endorsement?

MS. MUNTHALI: Now. If you vote no -- or yeah, do not pass, it would lose endorsement.

MS. MARINELARENA: But just to be clear, the composite that this is a part of will not lose endorsement, because the components do not need to be endorsed.

CO-CHAIR GUNNAR: But the developer can come back to the committee based on this non-endorsement decision with a rebuttal that we can then review and make a new vote on, yes or no?

MS. MUNTHALI: They could, but I think this is where we have to have dialogue with the

developer. So how soon could you come back to the committee? We do have a path forward if the measure is not passed or recommended for passing, because it has to be upheld by the CSAT. And Amy and both Lee are on the CSAT. So we do have recourse if you want to reconsider, if you want a reconsideration request of the standing committee's decision today prior to the end of the comment period. But it all depends on how quickly revisions can be made.

DR. BADHWAR: May I just ask a question just for point of order; this is the same argument from the previous discussion, and so if a decision on this is different than the previous discussion, does it not go back and it's still applicable to all of our measures, if you then separate the individual elements from the composite and vote against one or any of them? Speaking in broad generalities because the validity, specificity, face validity of these items and the statistical aspects of all these are all similar to -- the same thing for the

aortic valve sub-components. So I just wanted to be clear on how we would approach things.

MS. MARINELARENA: No, so this is different criteria. The conversation started before the composite, but it was just the general conversation. The composites refine 2561 and 2563 because they are publicly reported. We just got a little bit ahead of ourselves talking about this. Again, this is an individual measure that is not publicly reported by itself, which is the issue.

CO-CHAIR GUNNAR: And the time from its initial endorsement has passed, from which there was an agreement at the get-go that it would be publicly reported.

MEMBER YATES: And I would argue that SDS is within their rights to feel like they are using public reporting in a sense that it's part of a composite. And I agree with Dr. Fleisher from his earlier comments that I don't think that the Central Committee would actually consider this scenario when they started to -- or when

they decided to start enforcing the public reporting as being a requirement for re-endorsement. And I think this is a very different and separate environment, and I think we're well within our rights to prove it and say that it's -- or say yes to this vote and yes to endorsement. And with the proviso that it's with Dr. Fleisher's addendum and also with the proviso that it's not what they intended, or it wasn't what we think they meant. I would, however, point out that I think Dr. Fleisher has moved into a living constitution type of philosophy and away from originalism with his comments.

DR. SHAHIAN: I could respond from a developer standpoint that I think you should vote up or down based on the measure as you see it. I would not anticipate that there are going to be significant changes on our part. We've been doing this now for, this is our ninth year public reporting. There is no organization in the United States or perhaps worldwide that does more exhaustive public reporting than we do. We

report these measures as part of a composite, we publicly report them. We think for the public that it's much more useful to have the measure as we currently provide it. So like it or don't like it, but I think this is what it's going to be.

CO-CHAIR GUNNAR: Any other discussion?

All right, vote.

MS. KOSURI: Voting is now open for the use part of the criteria for Measure 0122.

I think we're still waiting for one more vote.

Oh, so we have 15.

Okay, for the use part of the measure criteria for Measure 0122 we have a total of 15 votes with 8 voting to pass and 7 voting to not pass. And the percentage -- give me one second -- is 53 voting to pass and 47 voting for no pass which indicates consensus not reached.

CO-CHAIR FLEISHER: Which is actually back to my comment that that means by the post

call we have to reach consensus, so is there any information that people want from the developer? We heard David talk to help them make a determination by the call, because we continue to talk, right, with consensus not reached?

PARTICIPANT: Mm-hmm.

CO-CHAIR FLEISHER: We continue to vote on the rest --

PARTICIPANT: We don't get a final vote.

MS. MUNTHALI: We don't do a final vote.

CO-CHAIR FLEISHER: But everything else --

MS. MUNTHALI: Yes, and part of the reason we wait for the post-comment call is to see if the public comments will help to inform your final vote, to help you get over the consensus not reached decision here.

MR. ANTMAN: May I just ask what is the threshold for consensus, what percentage?

MS. MUNTHALI: You needed to get 61

percent or more.

MR. ANTMAN: Thank you.

MEMBER YATES: I would ask STS -- this is a reasonable ask for the next two weeks or whenever we have our post call; I would ask that they provide a written statement just explaining why there would be adverse outcomes for reporting it separately as opposed to continuing to use it primarily in a composite.

MS. SKIPPER: And our post-meeting call is next Wednesday, so if you are able to come up with that, we would need it for the call on February 20th.

MS. MARINELARENA: Actually, what I would recommend is this is consensus not reached, we're not going to have an overall vote. This will go out for the 30-day public commenting period. That would be when STS should submit that comment to the public and then we ask for comments from all of the public. And then after the post-comment call that is when you will vote again and consider STS' statement and any other

public comments that we receive. So it won't be on the next immediate call, but in 30-some days. We will outline all of this for you.

CO-CHAIR FLEISHER: So the important part about that is since it's the public issue that we have here, the clearer the explanation and if there are other groups that either support or refute your explanation, that informs this discussion at the call. So we take all public comments very seriously, so.

MS. SKIPPER: Call your friends.

CO-CHAIR FLEISHER: Well, no. So if there is a group that actually can talk to how these measures are utilized, including -- well, we can do that offline. That would be very helpful.

CO-CHAIR GUNNAR: I think we're -- what did we do -- we just did vote overall, right, to endorse or not endorse; is that what we're doing?

MS. SKIPPER: No, not for this one.

CO-CHAIR GUNNAR: Not for this one.

We'll just stop and move on, right?

MS. MARINELARENA: So we'll go back to --

CO-CHAIR GUNNAR: Can we go --

MS. SKIPPER: Yeah, we do usability but not in overall recommendation.

CO-CHAIR GUNNAR: Okay, from a strategy point of view now there's a line of measures that are going to get us to this decision point for all of those measures. How do you want to handle this? You want to just go through it like we normally do and just march through them?

MS. SKIPPER: For the record.

CO-CHAIR GUNNAR: For the record we're going to march through -- and where would you like to go next? That leaves the lipid, the process measure for last -- unless you want to do the process measure now and then bunch all the other ones later. We either got to --

MS. SKIPPER: So we can have the discussion on usability if there's nothing else

to add for 0122 and then we'll move on. And then we can start with the 0114.

CO-CHAIR GUNNAR: So is there any further discussion on usability --

MS. SKIPPER: Mm-hmm, for 0122. So you can have the discussion now; we're not going to vote on it. But if there's anything else you'd like to add for the record now, that will be put out in the report for public comment.

CO-CHAIR GUNNAR: Hearing none, carry onto the next.

MS. SKIPPER: Yes, Measure 0114 and the developer will introduce it. I'm sending an email now with some additional material the developer did provide and I'll also be screen-sharing it. So, Mark?

MR. ANTMAN: Yes, thank you. While we're waiting for that screen-share to come up -- Measure 0114 is the STS risk-adjusted postoperative renal failure measure. So this is the first of five component measures of the morbidity domain for the CABG composite to be

reviewed today. And again, this individual measure along with the other four components of morbidity are statistically rolled up together to create the morbidity component that goes into the CABG composite. Now that the edit is on the screen I'll explain that -- and I apologize for not being able to get this into the materials for the committee review earlier. It's just within the last month, approximately, that based on some feedback from STS database participants we realized that there was a need for clarification in our exclusion for this measure. Christy or other staff can -- thank you. The previous wording of the denominator exclusion, as you see here, was patients with documented history of renal failure baseline serum creatinine of 4.0 or higher. Some STS database participants questioned how they can account for this because there isn't actually a history of renal failure element, data element in the STS database. Participants have all along been simply reporting the data elements that exist, baseline serum

creatinine or patient undergoing dialysis, specifically currently prior to surgery undergoing dialysis. So this edit is simply intended to match the language of the exclusion to the data elements that exist in the database; it doesn't represent any significant change to the measure; it's simply an edit to the exclusion, as I said, to match the data elements.

And Christy, if you'll scroll down to the next page, then you can see under the denominator exclusion, so there's the edit. We've eliminated documented history of renal failure, baseline serum creatinine remains as it was is simply edited to say greater than or equal to 4.0. And again, adding the phrase "Or currently prior to surgery undergoing dialysis," because that in fact is the data element that's collected.

So we're happy to answer any questions about that.

MEMBER JARRETT: This is Mark. The other Mark. The baseline is collected considered

at what time before the surgery -- excuse my ignorance on that -- and does that represent, what about somebody with acute kidney injury where their baseline is normally 0.9, they come in with some heart failure and other issues and they have some renal insufficiency and their creatinine is high, but in reality that does not represent major intrinsic renal disease?

DR. PAONE: I'm sorry; could you just repeat that again? I only heard half of it.

MEMBER JARRETT: I'm sorry; I apologize. So I assume that 4.0 represents the last baseline creatinine prior to surgery, but how does that reflect the delta if somebody had, came in needing surgery but had acute heart failure associated with, had some pre-renal disease or acute kidney injury, and let's say even from dye or whatever and really does not reflect their intrinsic renal disease but reflects either pre-renal or acute kidney injury? Does this really differentiate between those, or have you evaluated whether that makes a

difference?

DR. PAONE: No, that's a good question. The intention is to eliminate patients with chronic renal failure of significance so that you're not -- and someone with a creatinine over 4.0 is very likely to go on and require dialysis, and so that's intended not to include those patients, although they're very small in the outcome measure. But you're right; I don't know how to account for that. I suspect the number is fairly small, but I don't know that we've looked at that specifically. It is something to think about as we go forward in terms of trying to make sure that we capture all those who need to be captured with this. But I don't know that we've actually looked specifically at that, it is a very good question.

DR. BADHWAR: I'd add to that, though, and say that the concept of the actual domain is to be, again, more stringent to detect anyone because from a patient advocate or a hospital, whether you get renal failure post-operative, if

you go on new dialysis, that's bad, regardless if you had pre-renal disease or not. And so we're actually being more stringent or penalizing sites for the negative outcome. And so you can look at it in two ways; I would advocate that looking at it from the benefit of how to avoid this measure and if you had that patient that had AKI preoperatively from all the data that comes out, so from a dye issue, those are the patients you should probably wait and not actually operate on right away. And so I would just advocate that the reason this has been edited like this, it's actually making it more stringent.

CO-CHAIR GUNNAR: All right, our discussants, are we headed to evidence now -- hm?

PARTICIPANT: Discussions with Dr. Saigal --

CO-CHAIR GUNNAR: Dr. Saigal, Stein and Barbara's not on, so. So evidence?

MEMBER SAIGAL: Right, so this is outcome measure and, you know, I guess evidence requirements aren't that significant, but there's

a spread in performance 0.8 to 9.94 percent. It's a significant problem, there's at least one thing a surgeon can do to mitigate the outcome, so it's pretty reasonable. There's a gap in care, the report what's mentioned they also talk about the risk for this outcome being worse among black patients and a little worse in men. So that's the evidence.

Would you like to vote on that?

CO-CHAIR GUNNAR: Yes, we need to go through the votes. So, evidence. Any other discussion on evidence -- Dr. Stein, anything further?

MEMBER STEIN: I think the developer -- I thought the developer provided extensive evidence supporting the rationale for the measure and there are ways to reduce this measure during the perioperative period and I'm unaware of any new studies to change the evidence, so I think there's sufficient evidence.

CO-CHAIR GUNNAR: Can we vote then?

MS. KOSURI: Voting is now open for

the evidence portion of the measure criteria for Measure 0114.

I think we're waiting for one more.

Okay, so I think we'll -- voting is now closed for the evidence portion of the measure criteria for Measure 0114 where a total of 14 people voted to pass this measure, so the committee passes the evidence portion of this measure.

CO-CHAIR GUNNAR: I guess we'll go to gap.

MEMBER SAIGAL: I mentioned gap already.

CO-CHAIR GUNNAR: So we can proceed, unless there's any other discussion on gap, let's go and -- Joshua?

MEMBER STEIN: Actually, this is the first time I'm looking at this measure and I thought the information the developer provided in their Table 1B4 both for this measure and the other one that I reviewed, I think you need a statistician to look at this information. I

thought it was incredibly confusing and I think it's very difficult to interpret without additional information, at least for me to say whether there's gaps; specifically there are a whole bunch of odds ratios, there's no information as to what it's comparing the odds of, is it males versus females, one time period versus another time period, there's no confidence intervals provided. Odds ratios without confidence intervals to me are completely useless, so both this measure and the other one I reviewed, and probably all the measures it seems like the data's being presented the same way. And I think if a statistician looked at this, they could clarify and present it in a manner that we could make use of.

MEMBER SAIGAL: The reporting odds ratios that are significant, that was my assumption, and I don't know if that's true. Good question.

MR. O'BRIEN: This is Sean O'Brien; do you want me to jump in and try to give an

explanation of what was presented?

MEMBER SAIGAL: Yeah.

MR. O'BRIEN: So with respect to disparities, the intent there with that question, the measure designed to assess performance for participant level, and as it's designed it's already estimating outcomes in a way that's averaged over different types of patients. But the question is raised, is it possible to look at performance across participants in sub-groups of participants who may be subject to disparities. When we talk about stratifying results to look at disparities, these measures aren't designed to compare performance across the sub-groups; they're designed to compare participants across participants in comparison to the national benchmark. So when it comes to stratifying, rather than saying here are outcomes for black patients and white patients and males and females, we're asking the question is there variation across participants in their performance when you look at different sub-groups

of patients. So one participant in theory could have excellent outcomes for one type of patient or one set of, based on characteristics but different performance for different types of patients. So we've done an analysis, how much variation do we see when we look at patients separated by males and females, black and non-black race, et cetera just as an attempt to basically provide information that was requested by NQF related to disparities.

MEMBER SAIGAL: But you didn't do tests on those variation data to see if they were significant, did you report whatever you came out with? You didn't actually see if the confidence intervals were crossing one, or what was the --

MR. O'BRIEN: So, each participant, anytime we'd ever report a measure of participant level performance, that's always accompanied by a measure of uncertainty like a confidence interval. But now we're talking about summarizing results for 1,000 hospitals, and so what we're doing is saying each hospital gets a

point estimate of their estimated performance and how much variation is there across participants. So each person has an odds ratio and they're summarizing the distribution of those point estimates of odds ratios, so there's not a great way to summarize -- we could summarize the distribution of the lower limit of the confidence interval or the upper limit of the confidence interval, but basically -- yeah, we're basically just saying how much do we see differences in our estimates of performance across participants. That's why there's no confidence intervals associated with them.

MEMBER SAIGAL: That doesn't make any sense to me. I don't know if anyone follows that. Can you explain it to me?

MEMBER DUTTON: They're not actually doing a comparison; they're showing the distribution

MS. SKIPPER: Can you speak to the mic, please?

MEMBER DUTTON: I'll try that in

English. They're not actually doing a comparison, so there's no statistical comparison of the point estimate, as he says; rather, this graphic on the screen is just a distribution of their sites. This is what the curve looks like for the variability in the data. I actually find that very useful to know how the percentiles lay out.

MEMBER STEIN: I really have no idea what they're showing. I mean, here it says 2016 or 2017, then there's information 2011 to 2012, then there's a bunch of summary statistics. And then if you scroll down more, there's a bunch of odds ratios, but there's no explanation as to what they are.

MR. ANTMAN: May I comment on that table, please?

Excuse me. This table was originally submitted in an appendix to our application because we realized that if we had to put this table into the submission form itself, the columns would get confused and it would be

impossible to read. So, NQF staff, if it's possible to pull up the appendix we originally submitted, that will show this data much more clearly. I was asked by staff to put this into the form itself, which I agreed to, but knowing that it would be very confusing without the columns of the table being clearly distributed. So in the appendix you'll be able to read this data much more clearly.

MEMBER SAIGAL: So just to clarify this; so basically there's a spread in performance, so there's a gap, the questions about disparities, and what I'm understanding from the statistician is that there is a difference in post-operative renal failure if you're black versus you're white, and that is a meaningful statement to make; is that correct?

MR. O'BRIEN: No, that's not what this is looking at.

MEMBER SAIGAL: Okay, that's fine. Thank you. So basically we can't say if there's a disparity in the outcomes as presented here, so

there's no comments you can make about it, I guess.

MEMBER STEIN: Right, that was the conclusion I came to as well. There's not enough information provided.

MEMBER SAIGAL: Okay, so not enough information for that.

CO-CHAIR GUNNAR: Are we ready to -- any other discussion?

If not, then we'll vote on performance gap.

MS. KOSURI: Voting is now open for the performance gap portion of the measure criteria for Measure 0114.

CO-CHAIR FLEISHER: Has everybody voted? We need one more. Did we lose --

MS. KOSURI: I think we have the 14 people we need. So with a total of 14 votes, 2 votes for high, 10 votes for moderate, and 2 votes for low, the committee has passed the performance gap portion of the measure criteria for Measure 0114.

MEMBER SAIGAL: Okay, the next one is validity.

CO-CHAIR GUNNAR: Reliability. I got my cheat sheet here.

MEMBER SAIGAL: Okay, so I guess these data have been contracted several times; there's like a 10 percent contraction rate that seems pretty reliable to me. I don't know if anybody has questions about that?

CO-CHAIR FLEISHER: So are we ready to vote? Any comments?

Let's vote.

MS. KOSURI: Okay. Voting is now open for the reliability portion of the measure criteria for Measure 0114.

Okay, voting is now closed with a total of 14 votes where 4 people voted high and 10 individuals voted moderate. The reliability portion of the measure criteria for Measure 0114 has passed by the committee.

MEMBER SAIGAL: Now validity?

Okay, so I think the only issues here

are -- one of the tests of validity was whether performance in one period predicted performance in a later period. And it did, but then again you would hope that there would be improvements over time in these scores, so it's a question about whether that's a good test of validity or actually it's just showing you nothing's happening in terms of quality improvement. The other issue is that it does include race in the risk adjustment model; we already discussed that, so same I guess caution is there. But overall, I would not mark it as valid personally.

MEMBER STEIN: Yes, I had a similar concern about comparing past to current performance; I think in a way that's good, but then it also suggests providers can't do better. I also had some questions about whether the developer looks at outliers. There was someone who had like a 20 percent rate of acute renal failure which to me seemed very high to see whether that was like a problem with data input versus the fact they really had that kind of

rate.

Do you guys look at outliers to see whether you're confident that it's not just how it's being reported?

DR. BADHWAR: Your question's about outlier determination as a whole or just for this particular measure in terms of the creatinines of 10 or something like that, right?

CO-CHAIR GUNNAR: So this gets to audit question about -- so the validity and how do you go back and ensure that outliers or anybody is inputting valid information?

DR. BADHWAR: Sean can probably chime in to just talk in concept of the variable of say a creatinine is 111, something like that; that's immediately deleted as an obvious one. I'll let Sean answer overarching issues with the data entry in that large determination.

Sean, can you chime in?

MR. O'BRIEN: Yeah, the participants receive a very detailed data quality feedback report that's separate from the feedback report

reporting outcomes. I don't really remember the exact length and all the content of the report, but it's several dozen pages and very detailed. So missing values, but the raise checks on values are incorporated up front in the harvest process as well. So there are checks in place. When it comes to looking at outlier values for outcome rates, I think that most sites are paying very, very close attention to their outcome rates, and so have an opportunity to detect if their performance is in line with what they think it should be. So that outlier -- an outlier renal failure rate is not in the data quality report but is in their feedback report. So beyond that I'm not sure I have data to address the question about outliers.

CO-CHAIR GUNNAR: Let me ask a specific question; do you impute values for empty data, missing data fields?

DR. BADHWAR: Not for an outcome measure like that. If it's, for example, for renal failure -- this comes to the audit process

-- it's actually a chart examination of the data field. So for example, did they have renal failure or not have renal failure, did they go on dialysis or not go on dialysis. That is part of the audit process.

MEMBER STEIN: So if someone has a renal failure rate say of 20 percent, because that was I think the highest reported -- unless that provider comes back to you and says, "Hey, I'm surprised that my rate is so high," like do you guys go and look at the ones on the extremes and see whether they legitimately are on the extremes versus some sort of inputting error with how the data's being entered or something?

MR. O'BRIEN: We do with internal process of review, but we don't have a set criterion for assessing outliers like that with respect to renal failure and other outcome rates.

MEMBER STEIN: One other comment and this is not just to the specific measure; in terms of your risk adjustment methodology, it seems like you adjust for many different medical

conditions, and I know we already talked about things like race, but do you guys consider putting into your risk adjustment things like whether the patient's from an urban or rural location, whether the care's at an academic or tertiary facility. There are a bunch of non-medical factors that it would seem to be useful to build into risk adjustment models, so I'm wondering, are you planning on putting those in or are they being put in?

DR. BADHWAR: That's a very good question and this is a general question; those are fields that are actually in the database, and so they are being recorded. Whether they parsed out depending on the risk model as relevant in the multi-variable regression, it changes, but those are captured. So that's a good question.

DR. SHAHIAN: I can expand on that a little; I don't think those would be appropriate for a risk model being used to assess the performance of an institution. That kind of risk model -- and Sean, you can chime in as well --

that should be based on patient characteristics, not on the characteristics of the institution or where it lies. A patient may be -- well, I'll just leave it at that -- I could go into a lot of detail but historically and traditionally those are not included in models used to assess performance. Now, there are people that say, and this is being argued in some of the hospital compare measures, that hospitals should be compared to "like" hospitals; that's a different question. But in terms of adjusting procedural performance, I would argue that those things don't belong in a risk model, even though they may have an impact on it, they don't belong in a quality performance model.

Sean, do you want to comment on that?

MR. O'BRIEN: I think just leave it there, but I can answer questions.

CO-CHAIR GUNNAR: So we are ready to vote on validity, unless there's further discussion?

MS. KOSURI: Voting is now open for

the validity portion of the measure criteria for Measure 0114.

Waiting on one more.

Okay, with a total of 14 votes we have 2 votes for high, 9 votes for moderate and 3 votes for low. The validity portion of the measure criteria passes for Measure 0114.

I'll check that -- and the percentage is 78 percent; it passes by 78 percent.

MEMBER SAIGAL: Okay, so then feasibility. I mean, proof's in the pudding; it's widely used, people are doing it as expensive to the hospital, I think. I don't know if we need to belabor that more?

MEMBER STEIN: Yeah, I thought the feasibility was fine. You know, I think the developer should consider risk adjusting for how the data's being inputted in, whether it's via EMR versus hybrid versus manual entry, because I think that probably is affecting the outcomes, as I mentioned earlier. But besides that, I think it's otherwise feasible.

CO-CHAIR GUNNAR: Any other discussion?

Let's go ahead and vote.

MS. KOSURI: Voting is now open for the feasibility portion of the measure criteria for Measure 0114.

Voting is now closed with a total of 14 votes where 6 voted for high and 8 voted for moderate. Measure 0114 passes for the feasibility portion of the measure criteria.

MEMBER SAIGAL: Okay, then usability; so I think this is the same issue we're discussing in terms of whether one of the single measure is not directly reported counts; under public reporting my personal view is that it does. It may be used as part of a composite but others disagree, so I don't know if we need to belabor that more or more discussion is needed.

CO-CHAIR GUNNAR: Any further discussion?

Go ahead and vote.

MEMBER DUTTON: Actually, I did have

a comment on that. This is one of the measures that is reportable by individual physicians under MIPS, so in their QCDR, so this measure is publicly reported.

PARTICIPANT: But it's not the same measure through MIPS. It would be --

PARTICIPANT: That's only if those physicians choose to use it, right?

DR. BADHWAR: And the QCDR it is reported as part of CABG as a separate item for renal failure. It's on the website. It's --

MEMBER YATES: It's the reporting level.

MS. KOSURI: Could you put your speaker on?

MEMBER DUTTON: So I see what Dr. Yates is saying that it is the -- as presented here it is a hospital or group level of reporting as opposed to a physician level of reporting. So that does not count as publicly reported even though it's the same data being calculated in the same model?

CO-CHAIR FLEISHER: So that's your determination.

MEMBER DUTTON: Okay.

CO-CHAIR FLEISHER: Remember, they have not defined what accountability and public reporting is in the way which the line is written.

MEMBER YATES: Right, but we haven't had any -- we're not being presented with a distribution of results and outcomes for surgery specific data which may be very different than what were being presented for site specific, so we can't really use that as a surrogate for reported. What goes out in MIPS is going to be different.

CO-CHAIR GUNNAR: The other is at the time when the measure was initially endorsed. It wasn't a common understanding between NQF and STS developers that we'll rely on a second set of individuals to report this publically and then I get to bootstrap that.

DR. BADHWAR: So just for clarity,

it's on the STS website.

CO-CHAIR FLEISHER: Thank you. So we should just vote.

CO-CHAIR GUNNAR: We're voting.

MS. KOSURI: Voting is now open for the use portion of the measure criteria for Measure 0114.

I think we had an additional vote who voted early, so -- a voter who voted early, so make sure you can check and try again.

Okay, voting is now closed. For the use portion of the measure criteria for Measure 0114 there were a total of 14 votes where 6 voted to pass and 8 voted to not pass, and this is consensus not reached with 43 percent.

CO-CHAIR FLEISHER: So we continue, because consensus is not reached.

CO-CHAIR GUNNAR: We continue with the voting on usability.

I thought we went to -- oh, no. We go to --

MS. KOSURI: No voting.

CO-CHAIR GUNNAR: Any further discussion on this measure?

Hearing none --

CO-CHAIR FLEISHER: I urge a five-minute sun salutation, whatever you feel to get up and move around.

MS. SKIPPER: And to correct for the record, the reliability on this vote, 4 high, 10 moderate, because the measure score testing was not completed, high should not have been an option given to you all. We apologize for that. So for those of you who voted high for reliability for this measure, is there any protest if we move your 4 votes to the moderate?

MEMBER SAIGAL: I'm going to freak out.

(Laughter.)

MS. SKIPPER: Hearing no disagreement, we'll --

And the same for 0122, we'll move your high votes to the moderate vote, if there are no objections to that.

PARTICIPANT: Why?

MS. SKIPPER: Measure score testing was not completed. The highest rating possible in that situation is moderate.

MEMBER YATES: What if we're rebels?

(Laughter.)

MEMBER YATES: I've got to point out that we're about to wipe out about a half of the endorsements for the STS measures that make up 30 of our inventory. And we're now going to take 13 of 30 out of being endorsed, and I think we're doing it on the basis of semantics of what qualifies for public reporting. And I'm just going on the record that I find this to be very frustrating and I'm not sure that this should not have been foreseen and addressed in advance if communications at the central committee is an issue. It's somewhat mystifying to me.

CO-CHAIR FLEISHER: So we don't have a central committee.

MEMBER YATES: But it's starting to feel like one.

CO-CHAIR FLEISHER: So one of the questions is in the four weeks during public comment, right, this will still go out for public comment?

MS. MUNTHALI: Yes, this will go out for public comment which is -- we have quite a bit of time.

CO-CHAIR FLEISHER: So part of my question is can we get CSAC or some other body to weigh in on A.J.'s concerns?

MS. MUNTHALI: The CSAC would be weighing in. The CSAC takes in not just the standing committee's recommendations but also the public comments, so you don't want to unfairly influence them. We don't know if there's going to be an issue. These are discussions that are happening. This is going to be shared with the CSAC; they're going to review the report as well. Your co-chairs will be at the CSAC to talk about the discussion. But yes, you guys should talk about the portfolio at large and what this means to it, and where there may be measure gaps, do

you have the right measures in your portfolio; I think all of those discussions should be happening at the standing committee level.

MEMBER SAIGAL: I'm confused. So we're saying that because of the non-agreement on that one issue that we're not going to -- there's a follow-up call to this that we're all on that we can still further discuss it, or is that's it?

MS. MUNTHALI: All of the recommendations whether the measures pass or not are included in our report which goes up for public comment. And then you will meet on a post-comment call to reach consensus where you didn't, and then all of your recommendations from there will go to the CSAC.

MEMBER SAIGAL: Okay, I do second what A.J.'s concerned about; I think it sounds like a semantic issue and some very valuable measures are at risk, in my mind.

CO-CHAIR FLEISHER: So we have a chance in a month, post-public comment to re-vote and to get -- I mean, it would be useful to have

people after this is posted to discuss what public comment means including the public. So if there are public-facing entities that want to make comment if they are on this call or you know them, that would be helpful to guide us whichever direction.

Fred?

MEMBER GROVER: I would just like to see the complete NQF definition of public reporting before we leave today.

MS. MUNTHALI: We can share that with you.

MEMBER YATES: And given the fact that we're going to have a fairly long day or afternoon for the post-public reporting phone call. Just do me a favor, Christy and company, have you guys already put out the Outlook invitation for that?

MS. SKIPPER: Yeah, so the post-comment call is May 8th. Is that what you're asking?

MEMBER YATES: That'll help me to just

make sure that I block out my schedule.

MS. SKIPPER: Yes, and the report goes for comment March 21st.

MS. MUNTHALI: So the CSAC will meet in July and what we do prior to every meeting they have is prime them about the issues. So this will be something we talk to them about in addition to them receiving your report, and some members of the CSAC may join the post-comment call. We welcome that as well so that they can hear the discussion from you directly.

DR. BADHWAR: Respectfully, may I ask a question, Elisa, just from a process point? Would you desire a response from STS in the public reporting period or as a separate communication directly to NQF? How would you like us to respond to these issues?

MS. MUNTHALI: In the period, yes.

DR. BADHWAR: Okay.

DR. PAONE: If I can just echo what Dr. Yates said, I'm sure it's apparent at this point, but essentially 10 of the 15 measures

today will have been failed on the basis of this question as we go forward.

MS. MUNTHALI: So just because consensus wasn't reached, it doesn't mean that you have a final decision, it doesn't mean that they're failed yet. The committee has to vote on whether or not they pass during the public comment period, so I wouldn't take it as the last word.

DR. PAONE: I guess I meant for the purposes of today's session, but 10 of the 15 will all have that same issue as you discussed needs to be addressed going forward.

DR. SHAHIAN: Dr. Fleisher, can I make a comment? I apologize; I have to leave fairly soon for the airport, so I won't be able to join the rest of the discussion. But I'd just like to thank Dr. Yates for saying what I would have liked to have said; I think this is an unfair black mark against measures we have been using, and in our opinion in good faith publicly reporting for close to a decade. Based solely on

a semantic issue I've tried to outline the scientific reasons why we've done what we've done, and we will put this into a formal document, but to discredit these measures has implications that are far-reaching. These measures are components of our composites, probably the most widely regarded measures in, at least in surgery right now, and whether you no longer require NQF endorsement for the individual components of a composite, there's no question that when and if NQF endorsement is withdrawn from measures that are part of the composite, it will cast doubt on the overall system of measurement, and I think that's unfair. So I would like to express my strong discomfort with the approach that's been taken today. We will respond and I hope that folks on reflection will come around.

I will also distribute the original paper that we based this approach on, a set of papers so that you can read in detail the degree of discussion that went into whether or not and

how to include these individual components of morbidity. This was not some willy-nilly decision; this was based on a lot of analyses and I hope that you'll take the opportunity to read that.

So thanks for the opportunity to make that comment.

CO-CHAIR FLEISHER: Thank you. I'd actually like a brief discussion here to follow up on AJ's, and then let others. If we ignored what the words say but get to the intent, I mean, do people have strong feelings about whether or not these are publicly reported?

You know, because I'm curious about, are people voting no because of the way it's phrased, or people feel that the absence of a public reporting, independent of what the words are, is sufficient to vote it down? Could I get -- Amy?

MEMBER MOYER: So, I'm kind of thinking this through. To me, we're not saying those measures are failures, or that they

don't -- they're not acceptable.

I think, to me, it's somewhat a question of what types of measures need to be individually reviewed and endorsed by NQF. And so, it may be perfectly appropriate these measures aren't individually, publicly reported.

And so, then, do we need to review them as individual standalone measures that aren't going to be publicly recorded or used separately, because we're still looking at the composite. The composite would still be endorsed, it would still be something reexamined.

But instead of looking at, I don't know, 18 measures, maybe we have a really in-depth look at three composites. Because I think it's still important, but I'm wondering if the -- I don't know if the takeaway message -- I don't know that it's necessarily intended to be, no, you've got to like report all of these.

But maybe it's just a looking at, you know, what needs that individual level of review and that individual level of endorsement, for

things like --

CO-CHAIR FLEISHER: I'm going to get the CSATs. Did you have a comment?

MEMBER EREKSON: Yeah, I wanted to just comment on -- so, we're not in a point of consensus. And so, I think now is the time to say, well, what would help us change our route and what would help us get to consensus.

CO-CHAIR FLEISHER: Right.

MEMBER EREKSON: And I clearly heard the scientific basis of why the participants in the STS database want to publicly report in the way that they're publicly reporting. But what I haven't heard is how that public reporting is used by the public, or how that public reporting is used by the patient, or where the patient's voice is in all that.

And that would absolutely change how I'm considering this and how I think -- how we could even advise NQF on the public reporting measures over the month if I get more information on that, because that's where I'm really

struggling today. Not that I -- that's just where I'm struggling today.

CO-CHAIR FLEISHER: Rick?

MEMBER DUTTON: Amen, Fred. I feel as if NQF has made a major philosophical or policy shift by the seams here that we're just now catching up to.

I think NQF has had enormous value as a steward of measures for quality improvement, that would be the Q in NQF, and a shift to public, to emphasize in public reporting, is a somewhat different mission, and I believe there are potential and intended consequences there to how the measures are used and their value, honestly.

It's why I asked the question a day earlier, David, about what their experience with public reporting has been.

So, these are measures that have had a long history of really improving healthcare. They're about the best technical measures we have in our entire portfolio. I think they are still

enormously valuable for improving healthcare at the hospital and physician level.

Now, it seems like this isn't the kind of measure we want at all. And that's what's bothering me.

MEMBER GROVER: I was wondering for the SDS-advocated. We'd complied with NQF on this issue, plus the fact that I thought it was the right thing to do for our patients and the public.

Quite frankly, maybe we didn't read the fine print well enough in the definition, but we assume that since these were a part of the composite measures, they were participating in the public reporting effort.

This is a -- if this goes out for public comment, news in some very important opportunities to actually, potentially connect with CMS for physicians-based reimbursement, which would actually follow data instead of -- or follow the most accurate way, and this could be very, very damaging.

And if there's any way we can solve this issue, if it's solvable, before the public -- it's released to the public, it would be very helpful.

CO-CHAIR FLEISHER: Thank you. I'm going to speak on behalf of myself, not as Chair or a member of the Board, because I am trying to think through the implications of what we actually said at a high level, that it's okay to look at the composite, but not the individual measures.

And I don't know if we've ever looked close enough at the individual measures within a composite to actually say they're good enough, compared to the composite. We'd actually be just looking at the composite as whether or not that's valid, rather than looking at the measure of renal failure and individual mortality.

So, I'm actually concerned that I think Rick and others -- and AJ, you articulated -- the unintended consequences of making it easier not to endorse individual

measures could make the quality worse of the measures.

The positive of this is getting to measures that matter. So, perversely, I actually think what the STS did is the composites of the measures that matter.

The individual components, I wish there was some other means to say, these have been fully vetted by NQF, they're really important, they meet every criteria from stringency perspective. But they're not the measures that matter, because we actually agreed with you, or I do agree with you, that all that's important to put out is the ones you put on public reporting.

So, you know, I don't know where that goes within the organization, but it's good that potential -- I think when they thought of composite measures, they never thought of the positives of getting the composites proofed, but not since --

MS. MUNTHALI: Yeah. No, I think you

raise --

CO-CHAIR FLEISHER: -- and the components.

MS. MUNTHALI: The components. Yeah, you raised very good points. When we convened the composite framework technical expert can on 2014, use was not must-pass. So, these sort of discussions were -- they weren't on the table. Things, as quality measurement is evolving, we're evolving our criteria vows.

I know this doesn't give you much comfort. It is what it is right now. But this is where we're taking it to say maybe we need to re-look about -- look at how you squeeze in the individual components, vis-a-vis the entire composite.

And that's something we would do rather quickly. But it is, unfortunately, the criteria as it is right now. It is up to -- the degree to which, though, the measures and the composites meet the criteria, is up to you as a committee to decide. And I think you did take

the vote, but some are not -- there's no consensus that's been reached yet.

MEMBER YATES: I agree 100 percent with you in terms of the challenge to the quality of the composites is the elements of the composites are given a free pass on a regular basis, in that having the elements of the composites go through this process and be endorsed, only makes the composites stronger.

But I have two things to add, and they have to do with just basic fairness. These measure have already been endorsed before, and usability was part of that endorsement. And the potential target for its usability was presented probably before as being part of a composite measure.

And this committee and the NQF at large accepted that use as being acceptable at that time, and we're changing the definition. This is was similar to Lucy pulling the football out before Charlie Brown tries to kick it. I think this is, you know, sort of last second

changes.

The implications of this are that a whole bunch of things that are endorsed out there that include some of the Yale core measures that CMS uses, they're part of composites of reporting for value-based purchasing and for hospital compare, but not necessarily readily transparent or seen by the consumer or by the potential patient.

But we're not questioning those measures that CMS is using that are buried within complicated scores and outcomes. All of the HARQ hack-size scores would also fall out because those aren't transparent to patients. They get a star rating for the hospital and that's about it.

Now, I would argue that -- you could argue that maybe someone with incredible sophistication could dig into that and find it.

But they could also dig into the hospitals data by asking for it, with as much effort as it would take to go into those other measures that are still sitting out there that

really aren't being publicly reported.

But we've approved them for use within CMS's value-based purchasing, readmission rates, hack-side, all of those other ones that are out there.

So, this is opening up a can of worms as to what defines usability, and I think we gave people the reason to expect that a composite measure being reported was being used, or public reporting. So, I'm off my soap box.

CO-CHAIR GUNNAR: So, I just want to go on saying this. I was all in for connecting the dots between publicly reporting -- a composite measure is fundamentally publicly reporting -- until I understood, the way I understand it, is that I could still be a one-star in stroke if I could, if you will, two-star and the other four morbidity components, and be a two-star facility.

That means that fundamentally, I'm not publicly reported. I can't draw that association the way I understand it. Does that make sense?

If, as a consumer, I could go online and be assured that if I went to a two-star cardiac surgeon and whatever that -- I'm looking for an aortic valve enquiring bypass -- and it's two-star, then I can be assured that it's two-star both in mortality, as well as each component of that, that they don't have a heightened level of stroke or a heightened level of renal failure, that gets buried in that analysis.

And my understanding is that that's not true, unless I'm corrected.

DR. BADHWAR: May I just respond to that specific question first? So, again, this is an all-or-none phenomenon. If there is an outlier that has a very high level of stroke, that's not going to get buried statistically in a two-star issue. This is actually going to drag them down.

So, that's point one. Point two is more philosophical. Speaking on behalf of health systems here, as well as my role in STS as Council Chair, and speaking on behalf of STS, I

would say the answer is two-fold.

I would question, given the close alignment and sort of attached-at-the-hip partnership and public reporting transparency that STS and NQF have had over the years -- decade -- the risk of sort of disapproval of any one NQF-endorsed measure, as a hospital chair, when one goes through the quality aspects and sees that STS is no longer -- it's been sort of like the Moody downgrade -- that it's no longer that valued, and many of our quality officers sit on NQF or are on committees, including my own, was on just last week, how does -- what's the optics of that in terms of the reliability and the brand of STS quality?

The quality aspects of the database and its contribution is at the very center of what STS values as most important. And if we're a Moody downgrade because of this issue -- now, first of all, I'm saying, we would totally respect NQF and what you decide, but just remember the public trust and what that might do

in terms of optics, if we're now downgrading.

It's not maybe the most valued quality measuring tool, and does that actually help what we're all trying to achieve?

MEMBER CIMA: Can I make a comment?

CO-CHAIR FLEISHER: Yeah, make a -- go ahead.

MEMBER CIMA: We just --

CO-CHAIR FLEISHER: Before we keep -- I'd just like to get Rissa to comment, and then we'll continue the comments from the rest of the committee, because I think NQF should have a chance to respond.

MS. MUNTHALI: Yes, so I really do appreciate all of the comments, and I think it does show the strength of NQF endorsement. Really, what you're saying is, by having that signal, it signals something to the field. And that we take pride in.

But we also take pride in -- we're not making the decisions. We have constituted this multi-stakeholder committee with different

perspectives, and I don't think that every perspective is swaying in the same direction.

And so, this is not -- we are trying to be neutral in this, giving you the opportunity to see how you're interpreting the criteria and applying the criteria to the measures.

And so, this diverse committee is signaling that they're not quite sure where they're going, and this is -- I understand there's some strong viewpoints here.

I think we have to let the process weigh out. I think the ultimate outcome may not be as dire as you think. It is important for us to get, especially now, where we don't really have a definitive decision, to get others to weigh in, as well. So, I hope that helps.

We can't change, you know, the vote. The vote wasn't made by NQF, it was made by the committee members. So, I think what you started to do to get a sense of where people were feeling angst is a good thing, because I think it'll give you some general sense of the trouble spots and

what you may have to overcome by the post-comment call.

CO-CHAIR FLEISHER: Who would like to speak up from --

MEMBER JARRETT: Hi.

CO-CHAIR FLEISHER: Yeah, please.

MEMBER JARRETT: Hi. This is Mark Jarrett. So, first of all, following up with NQF's just, you know, what was just said, I think this offers up an opportunity rather than, you know, just a challenge. It's a really good opportunity.

Clearly, we're all -- you know, we all recognize that STS provides a great performance and quality improvement database that allows us to really move the needle.

The question really comes about of how we're defining publicly reported, which is an issue about composite scores. And quite frankly, here we're talking about this, but those of us who deal also with, you know, CMS stars, with, you know, three-year rolling averages that are

four years old, and other composite scores, like PSI-90, this is not an unusual discussion, so actually, I see this as an opportunity to bring this to, you know, on a national basis forward, because we have to make a decision, because we're either going to do, you know, follow science and say what's right and what really improves care, or get caught up in the lexicon of, well, if it's not defined publicly reported this way, then we don't want that measure.

Because that really kind of avoids the question of why we're measuring these things. And yes, transparency and giving the public the right information is really critical. But we have to make sure we're giving them the right information.

As I said before, if a composite score has invalid components to it, whether it be one or three out of seven or ten, then really that composite score may not be valid all together, and we need to address both the individual components, as well as the composite.

So, I don't -- I see this as more than an intellectual discussion. I see this as really the opportunity, maybe on a national basis, that we really address really basic issues going forward, because I'm sure it's not going to be just related to STS as NQF moves with other measures and other fields.

CO-CHAIR GUNNAR: Thank you, Mark. Any other comments?

MEMBER SAIGAL: Yes. Can I make one comment? And this -- Bill, what you were saying about -- I want to clarify that. The reason you changed your mind was because you're concerned that the two-star rating could cover up poor performance in one of the sub-measures?

MEMBER JARRETT: Correct.

MEMBER SAIGAL: And so, that an important thing to understand. So, basically -- but that's morbidity measure is a binary thing. So, if you fail one, you fail the whole thing. And is it possible to get a two-star rating if you fail morbidity?

DR. PAONE: Sean, can you answer that more definitively than I can. I would imagine that you can probably have a sort of high one-star rating and get an overall two-star rating. I don't know how often that occurs, if at all. Sean, do you know that from the statistical data?

MR. O'BRIEN: I don't know how -- I know for a fact that it can occur. I don't know how frequently. And I'd just say conceptually, we don't assume that a site's performance is exactly the same, that they perform well on mortality and exactly that much better on each individual endpoint.

The composite measures are averaging over the different measures, and of course there's some information lost when you try to sum up performance on multiple endpoints and multiple dimensions into a single number.

And so, there's not a -- so yeah, that's an area that's possible.

DR. PAONE: You know, I mean, I can certainly imagine that you can have every one of

your morbidities is in one or the other categories, and you would technically fail that if it were reported individually. But yet, overall, it becomes a two-star composite. I can imagine that that would happen.

I just -- what I -- I must admit, as I sit and listen to this, what I'm not quite so clear about is, we're not saying the measures are not important. We're not saying the measured outcomes aren't important. We're not even saying they're the wrong measures.

I think there's wide agreement that after all of the study that went into this, that these are the appropriate measures to be part of this composite, and many, many randomized trials and all sorts of reports. So, I'd use composite measures as outcomes for a variety of reasons, including statistical ones.

What's sort of, I guess, a little confounding to me is what we're really sitting here discussing is a changed definition of what public reporting is.

That seems to be basically the entire issue at hand, particularly for the mortality outcomes in the mitral categories, and all of the composite measures, because they're not specifically and individually reported. Rather, as part of a composite.

I just don't know, at least in my own mind, and I'm not the most experienced person in this group, but whether that should be reason to take these measures, which have been approved for a very long time, have been vetted and have been continuously upgraded, to change their status. It just -- I will admit sitting here, I just don't quite frankly understand that.

MEMBER SAIGAL: I agree with that and I. It's a great point, though, that those identified for the public's use. It isn't like a killer problem for the measure in my mind. It's just like a -- it's an important observation that can be used to improve public reporting, I think. But not necessarily -- it's not fatal to what is happening right now, I think.

DR. BADHWAR: I would say that these points are very valid in terms of the question. But if we look at the robustus (phonetic) of each of these measures, as you've attested to and voted on, these are not some random skin infection issues. These are major morbidity outcome measures.

And we're valuing each of them independently. And they all have importance. It's unlikely -- highly unlikely -- that any one site can be gaming and have a super high stroke rate and zero renal failure, to have a volume that would actually attest to public reporting.

Remember a couple of things. Just because you sign up for public reporting and you have -- you want to participate, if your volume thresholds are too low, you do not get public reported.

And so, those types of outlier determination questions that Dr. Gunnar raised, that's where those types of hospitals may come up, when they have to do ten of a certain

operation and they have a high stroke rate. Yes, that's a percentage.

However, they're not going to get public reported because they don't meet the volume threshold. So, the public trust is protected in that regard.

The second thing, let's raise one of the markers in the composite, and that's deep sternal wound infection. When this whole process started a few years ago, deep sternal wound infection was fairly high and relatively common.

And now, we learned about diabetic management and, you know, how you manage the sternum, preoperative antibiotics and all of these other quality instruments that we've done.

Deep sternal wound infection now is so negligible its weight in its contribution is so minimal. We're down to like the one percent range across the country because of the value of each of these measures.

And so, it gets back to the previous argument. If one degrades or takes away the

importance at a hospital level, there may not even be public reporting, or just learning or tending to how are they going to improve if it's not an endorsed entity.

DR. PAONE: If I could just take one more second and use an example here that Mark just pulled up from the website that may illustrate some of the questions about the value of the individual components I'm staring at for the AVR composite, there is one site that has two stars for operative mortality, one star for morbidity, who is a two-star program overall.

At least in this brief look there are four sites who have two stars for mortality, but one star for morbidity, who are one-star programs.

And so, I think that points to the effectiveness of the morbidity composite and being able to identify a lesser functioning or a core outcome facility, and I think actually confirms the value of the composite over individuals.

Because I would suspect the -- and I have no way of knowing from this data, and maybe that's part of the concern and I get that, but it's not unlikely that they -- each individual component may be a low two-star, but it's a one-star composite, as well as in the other direction. So, that's just an example of why I think that this is valuable.

MEMBER GROVER: Just a question moving forward. Would it help if -- and I don't know that this would be approved at the higher levels of STS -- if we did this, the morbidities and the incidents at each site and the mortality you've already got on there, is that what you all are after?

So, what if we go ahead, in addition to having the composite, do we know the results that go into that composite? What percent -- stroke percent -- renal failure percent mortality, that type of thing?

MEMBER EATMON: Yes, that would be something more aligned to what a patient is

looking for to kind of see, in terms of making the decisions, or just being educated themselves. And so, if that was a percentage number, they could have a realistic view of what the probability was.

MEMBER GROVER: So, was that implied -- was the NQF definition of public reporting for this?

CO-CHAIR FLEISHER: Yeah, that would be an example in the -- you know, the complicated question is, you know, if you look at CMS hospital compare, they don't -- I mean, you can get to some of the data, but a lot of times they say no different.

So, it might be, you know, whether STS comes up with a plan of how they want to present it. I mean, I'm just throwing it out there that, you know, the composite, if it fails, why does it fail? Where does, you know, if they have higher-than-expected stroke or something that's behind the numbers, how it's displayed is not what NQF endorses.

It's more of that they have a plan that we except to make this public reporting in some manner. And then, we vote on what that manner is.

DR. BADHWAR: So, if I might, I think the concept is an excellent one. The practicality is where I think we'll have some challenges. And I'm not speaking on behalf of STS yet, because we have to actually do the analysis.

But having been involved in a lot of these risk models and developing them, the good thing is that the incidents -- the actual hard incidents -- of many of these morbidity endpoints is so low that in order to develop a risk-adjusted model for an outcome, such as stroke or such as deep sternal wound infection, the actual data that's going to be required and the length of that, so right now we have a three-year composite for AVR.

If we develop a separate model to public report these items for a stroke, I mean,

Sean can potentially speak to this, or we can come back with an answer, I'm a little concerned that we might not have the actual sample size of the one event to be able to report risk-adjusted outcome. It might take like a five-year reportable issue.

CO-CHAIR FLEISHER: But this is an endorsed -- these individuals. So, all I'm -- again, as an individual asking, if you said this is an outlier, I don't know whether the committee would accept or not.

But it's really, you're asking for a plan of how you would publicly report without being specific in the way public reporting is defined for the individual components. AJ, did you have a comment? Or Chris, did you have a comment?

MEMBER HANDY: This is John Handy. So, I am kind of bugged by the fact that it seems like it's the prerogative of this committee to define what public reporting is.

I mean, we've had a definition in

front of us for about 30 or 40 minutes right now, but each of us is interpreting a little bit differently.

Some of us have felt that the public reporting is acceptable in its present form, and others feel that it needs to be more granular. Are we not working that out right now?

CO-CHAIR FLEISHER: So, I'm sorry, I didn't hear all, but what we were just discussing is, it is this committee's prerogative in the surgical space of what is acceptable public reporting.

So, NQF, at a higher level in the composite, did not say all public reporting is the same. It all must be hospital compare. What it's saying, in our domain do we feel that STS, for usability, has done public reporting that we find is a benefit to patients?

I know that's a really --- I need to -- I mean, but parsing how -- it's -- we are individually responsible in each of the standing committees.

MEMBER MOYER: So, I'm torn on this for so many reasons, because I think, well yes, we define it for the surgical area. I think consistency across the committees is a reasonable expectation as a measure developer.

And while we look at all the STS measures, if you're like an NC2A or a joint commission, where they're going to a bunch of different committees, I think it's important that you could have the same expectation of the same result.

As a purchaser, I'm always going to go for the composite. You know, I like it rolled up, I like the differentiation that comes with that. Gee, I'm really -- if I were to put my patient hat on, I like to see everything. I like digging into the details.

Whether that's actually useful I don't know, but I like to be able to see it, as a patient, which I don't know what -- I know I don't represent my organization, but I don't know if I officially represent my stakeholder ground.

So --

CO-CHAIR FLEISHER: So, we -- I know that my colleague is leaving at 3:30. Are other people leaving?

PARTICIPANT: I'm leaving at 3:00, but I'll be on the phone.

CO-CHAIR FLEISHER: Yeah, and I'm leaving at 4:15 on the phone. But -- so, we need to be done between -- in the next hour or so. Elisa?

MS. MUNTHALI: So, you know, I do want to clear the air. This decision to make this criterion of use must-pass was a very thoughtful one. Since you last met and maybe had a very substantive discussion about it, quality measurement has changed, accountability has changed.

The burden of measurement has increased, and one of the reasons we have made this must-pass is because we got so much critique from so many in Quality about just looking at the scientific merits of measures is not enough. You

need to look at the implementation of those measures, the context in which they're used.

And so, this is one way in which we're demonstrating that to say -- not to say -- to put limits on you, but to try to quantify all of this. That is quality measurement. Not just looking at the reliability and validity of measures. We understand that's important.

But if these are going to be put out as national standards that can be picked up by any entity, we need to make sure that to the extent possible, that there is transparency to the extent possible, and that is where the committee can decide that.

But Amy is right. Within all of this it is you as a committee determining the degree to which it's meeting that.

MEMBER YATES: I've got to add one thing. There's a consequence here, and the consequence is, is that as you become more granular, as you start to -- if you start to destabilize the composite measures by reporting

the individual parts of the measures, and scientifically if there's concerns from the collecting group or registry that if reported individually, they lose validity and reliability in terms of the risk adjustments and everything else that goes into the composite, you run the risk that people -- the surgeons, the providers, will no longer trust what's being reported.

I don't trust what's being reported in my specialty through two measures that were endorsed here, because of absolute lack of risk adjustment. But I'm held subject to it, regardless of the seats assessment being 0.67.

As that gets down to surgeon-specific, it becomes even less risk-adjusted and it becomes -- and the more granularity they get, the harder the risk adjustment and the harder the fairness.

And what you do is, you drive away care and access to care for those patients that are at the margin of acceptable risk because everyone is going to go through gang theory and

try to avoid risk.

And you're going to end up with less patients getting less care and it's going to hurt the most vulnerable patients first, and I think you have to be very careful how granular you want, because the public's crying for granularity, but the more granular you get, the more the people at risk are left out of the discussion.

DR. BADHWAR: To echo that interesting comment, I would also say that it's also applicable to the actual participant sites that actually are doing the voluntary public reporting.

And if there's a question on some of those validity issues, the subcomponents, which we will take back and we will respond as fully respectful to this process.

But I'd raise that secondary question. What happens if the sites say, well, I don't want to public report now? Then, it actually decreases the entire purpose of what we're trying

to do here together.

CO-CHAIR FLEISHER: So, we've got a lot of theoreticals. The question is, what's -- we've only got about an hour to an hour-and-a-half left of the meeting. Fred, you had a comment first?

MEMBER GROVER: I'll shut up because I'm not supposed to talk much, but I'm looking at this definition and nowhere in here do I see that it says that if you put -- if approved measure is part -- publicly reported by being a part of a composite measure, that that's not public reporting. And I think, in all fairness, we assumed it was.

And I think the definition lacks specificity and that needs to be considered. I really -- it really ticks me off.

MEMBER YATES: And what I just said isn't theoretical. It's actually happening, and I know that from polling within the American Association of Hip and Knee Surgeons.

And I also know that it's -- there are

surgeons in our community that scan who they're going to see in their office by their BMI. And if their BMI is over 35, they don't even get an appointment.

So, there -- and that has a lot to do with the question of reporting and perceptions of quality.

CO-CHAIR FLEISHER: So, are we still at the same place for everybody who's voted? Okay, so we're not going to call another vote unless somebody -- so, we need clarity and -- or we'll have more discussions for comment, as well as discussions internally, just so that --

(Off-microphone comments.)

CO-CHAIR FLEISHER: No, not criteria. But I think we should make the CSAC chairs aware of --

PARTICIPANT: They will be. Yes.

CO-CHAIR FLEISHER: -- what's going on. So, the question is --

CO-CHAIR GUNNAR: So we all know, there's one measure that's a process measure we

should go that's not -- doesn't line up under this.

CO-CHAIR FLEISHER: Right. So, we should -- people need to stretch? Or they just want to go to the -- go ahead.

PARTICIPANT: So, the process measure.

CO-CHAIR FLEISHER: All right. That process measure is -- it's 118, Anti-Lipid Treatment Discharge. And Keith's not with us today and it's Amy and TeMaya for discussions. So, the developers want to introduce this first?

MR. ANTMAN: Yes, thank you. As has already been said, this is the anti-lipid treatment discharge measure. It is worth pointing out, this is one of the morbidity components for the CABG composite. So, although this differs from the other measures under review today --

DR. PAONE: I apologize. It's not one of the morbidity measures. It's one of the perioperative medications which the CABG group has the two process measures, which are the use

of the use of the internal mammary artery and perioperative medications.

MR. ANTMAN: Yes. Thanks to Dr. Paone for that correction. Yes, it is part of a different domain in the CABG composite. The domain is labeled Receipt of Required Perioperative Medications.

But it is not immune to the concerns that were expressed before, because although when the CABG composite is publicly reported, there are domain scores and domain star ratings for each of the domains on the public reporting site.

You cannot drill down from the receipt of required perioperative medications domain to the individual, to the four or five individual perioperative medication measures. So, it is subject to that same concern.

That said, we'll welcome discussion of this measure.

CO-CHAIR GUNNAR: So, for evidence.

MEMBER MOYER: I don't believe there was any updates to the evidence since the last

submission, and that's fine.

CO-CHAIR GUNNAR: Any other discussion? And we vote.

MS. KOSURI: Voting is now open for the evidence portion of the measure criteria for Measure 0118.

MS. SKIPPER: And note, so those of you are voting high, recast your vote. The highest this measure can get is moderate, since no QQC was submitted.

CO-CHAIR GUNNAR: We may not have a quorum. Do we not have a quorum now? Do we kick this one down the road?

MS. KOSURI: So, give people another moment or two. And we're just waiting on two more votes.

(Off-microphone comments.)

MS. KOSURI: Okay, voting is now closed for Measure 0118. For the evidence portion of the measure criteria, 14 people have voted moderate out of 14 total votes. Thank you.

CO-CHAIR GUNNAR: So, we move on to

gap.

MEMBER MOYER: So, this presents a perennial topic and actually another interesting discussion around composites. The median value for participants on this was 99 percent.

So, I know my co-presenter isn't here, but I think both of our worksheets, we expressed a concern over the measure being topped out. That said, if it's part of an all-or-none composite, that could still be a meaningful component as part of that composite. So --

CO-CHAIR GUNNAR: Right. So, there's -- this gives -- to bring history to this committee, we were the first committee to recommend endorsed-but-reserved status on a topped out measure where the gap didn't exist anymore.

So, that's one possible direction for this particular measure going forward And I forget how we get there. We vote on gap and if there's no gap low, then we can make the recommendation for reserved status. Is that

how -- but endorsed? How do we do this?

MS. MARINELARENA: According to our guidance, this measure is not eligible for reserve status because reliability has not been demonstrated for the measure score. So, we only have data element reliability testing.

CO-CHAIR GUNNAR: So, the answer to this is, we follow the path through the voting process and if we get to a place where we stop, then that's where it stops, with no opportunity for reserve status, apparently. Yes.

PARTICIPANT: How did we do it last time?

PARTICIPANT: What's that?

CO-CHAIR GUNNAR: I can't remember the measures that we put in reserve, and --

PARTICIPANT: We put in the antibiotic skip measures, which obviously are extremely important measures.

CO-CHAIR GUNNAR: Right.

CO-CHAIR FLEISHER: I'm just curious about -- given the compliance. Does this make

any difference in your composite anymore?

DR. PAONE: I wouldn't -- statistically -- Sean, are you still on the line? Can you address that? Since it's an all-or-none, I would think just a couple -- I mean, overall it's a good thing that this composite has, you know, a very small gap. There is still a difference between the lowest and the highest.

I will point out in terms of the specifics of lipid therapy, that this is a 1A guideline for cholesterol in the recent cholesterol guidelines for patients with atherosclerotic disease, and certainly anyone having undergone coronary bypass surgery falls into that category.

So, it's an important measure, certainly from that standpoint. I mean, it makes as much difference as the pre-op beta blocker and the -- you know, pre -- I'm sorry, the post-op data blocker and the pre-op aspirin.

So, I don't know. Statistically, Sean, is there -- do you have any thoughts on

that?

MR. O'BRIEN: Yeah. I don't know to the extent which it drives the composite. And I think we don't assume that it's a major driver of the composite.

But I think we justify it based on it being a safeguard, whereby if a site did have substantially lower performance on this measure than that would contribute to the composite being lower.

So, it's not a major driver, but it's not adding noise to it, either, and it's there for accountability.

MEMBER YATES: One point of clarification. Ninety-nine percent -- a lot of hospitals are 99 percent compliant for all their patients. How many hospitals are not 99 percent? That's the question.

Because I see on here there's still some scores in the low 90, and that's the institutions' response. Being -- you know, a whole lot of hospitals being perfect at it

doesn't mean there's not a gap.

DR. BADHWAR: That's absolutely right. In fact, because it's in the composite and because we value that as important, I will flip around it.

So, you're 100 percent right in your statement. Flip it round the other way. If it wasn't in the composite, then human behavior takes effect and the likelihood of them focusing on discharging someone with high-dose lipid therapy may be skipped. And then, it does do -- may have a negative impact, since it is a guideline recommended point. But --

MEMBER YATES: So, what percentage of the hospitals fail?

DR. BADHWAR: Yeah, I think it's a low percentage. But since it's also an all-or-none --

MEMBER YATES: Right.

DR. BADHWAR: Then, we'd have to break down that data.

MEMBER YATES: Well, is it two, three,

four percent?

DR. BADHWAR: I'd say that's probably accurate.

MEMBER YATES: Okay. Well, two, three, four percent would be adequate gap for something. That's a -- I think a two or three, four percent hospital failure rate on this would be a major problem for a 1A recommendation.

Do we have that data, as we scroll back on the screen there?

MEMBER MOYER: Now to where? Where are we looking? I missed that.

MEMBER YATES: I'm not asking -- you know, the hospitals are performing 99 percent, but which hospitals -- what percentage of the hospitals don't do that?

CO-CHAIR GUNNAR: They want the data for that gap.

MEMBER YATES: There we go, for performance. So, there's still -- am I incorrect in saying that there's still a range there?

MR. O'BRIEN: The STS provided some

data looking at variation and performance across categories of performance.

So, we looked at categories of hospitals based on their past performance in one time period and look at their compliance rate for anti-lipid medication at a more recent period, and across the low-, mid- and high-performance groups, the percentage of anti-lipid medication increased from 89.8 percent to 99.8 percent.

So, that's the degree of differences across the performance groups.

MEMBER MOYER: It's actually further down in the document. It's another one of those funky tables. But if I'm reading this right, the bottom ten percent of hospitals has a 46 percent compliance with this. And then, the next decile starts at 94 percent. It reaches 99 percent out of like 50 percentile. So, for those bottom hospitals -- what page is this, it's page 19, there.

So, I mean, for those bottom hospitals, that is a pretty significant gap.

They're less than 50 percent.

MEMBER YATES: Yeah. And again, for a level-1A recommendation, obviously there's excellent evidence for it being beneficial. Any kind of gap in this one I would think still would warrant some consideration for being at least moderate evidence for gap.

CO-CHAIR GUNNAR: Any other discussion? Vote for gap?

MS. KOSURI: Voting is now open for performance gap for Measure 0118. Okay. Okay, voting is now closed. For a total of 14 votes, we had 12 votes for moderate and two votes for low.

For the performance gap portion of the measure criteria, for Measure 118 I will also apply the percentage. So, 86 voted for moderate and 14 voted for low. So, this portion of the measure is passed by the committee.

DR. BADHWAR: Drs. Gunnar and Fletcher, if I might, I just want to thank everybody. I have to depart. I'll leave

responses in the fine hands of Mark Antman and Gae Paone.

CO-CHAIR GUNNAR: Thank you for all your presence and your input today. Thank you so much.

CO-CHAIR FLEISHER: And your willingness to engage --

CO-CHAIR GUNNAR: Yes.

DR. BADHWAR: Anytime.

CO-CHAIR FLEISHER: -- in complex issues that we're still grappling with.

DR. BADHWAR: We're all on the same page. Thank you.

PARTICIPANT: Go Mountaineers.

PARTICIPANT: That's right, go Mountaineers.

MS. SKIPPER: So, we're ready for the discussion of the reliability for 0118.

MEMBER MOYER: The only concern that I had and that I saw mentioned was that potential level of analysis question, which seemed to be a kind of non-issue in a previous measure. So --

CO-CHAIR GUNNAR: TeMaya, any question? Anyone else? All right, we will vote on reliability.

MS. KOSURI: Voting is now open for the reliability portion of the measure criteria for Measure 118.

PARTICIPANT: So, that's the problem.

MS. KOSURI: Okay. Voting is now closed. For a total of 14 votes we have 13 votes for moderate and one vote for low.

For the reliability portion of the measure criteria for Measure 118, the committee has passed this measure with -- let me provide the percentage.

CO-CHAIR GUNNAR: Move on to validity?

MEMBER MOYER: It looks like there was some question around the method of validity testing used.

I think we discussed this on a previous measure, as well, that it looked at stability and performance over time frames where you might expect to see an improvement in

performance, although, honestly, given the performance distributions in the two time frames, I'm not sure how much improvement you would have seen. So --

CO-CHAIR GUNNAR: Any other comments? We'll vote on validity.

MS. KOSURI: Voting is now open for the validity portion of the measure criteria for Measure 118.

(Off-microphone comments.)

MEMBER CIMA: I don't have it active on my screen.

MS. KOSURI: It should be active.

PARTICIPANT: I had to refresh my screen, Bob.

MEMBER CIMA: I just did that and it didn't do anything.

MS. SKIPPER: If you're still having trouble, you can chat in your vote and we'll record it.

MEMBER TEMPLE: This can be a big component, I would rank it as a moderate.

MEMBER MOYER: Okay, thank you, Larissa.

(Off-microphone comments.)

PARTICIPANT: I'll put it as moderate. I just lost the screen now. I'm going to re-log on completely.

MS. KOSURI: So, voting is now closed. We have 14 votes for which -- for moderate for the validity portion of the measure criteria for Measure 118.

CO-CHAIR GUNNAR: Feasibility is next, right? Yep. No additional comments? It's feasible, all the reasons? All right. Vote.

MS. KOSURI: Voting is now open for the feasibility portion of Measure 118. And, Larissa, if you're still having an issue, feel free to chat. We can see how you vote via chat. Or, verbally is fine, as well.

MEMBER TEMPLE: It's working now. Thanks.

MS. KOSURI: Just waiting on two more votes. Voting is now closed for the feasibility

portion of the measure criteria for Measure 118. With a total of 14 votes, we have 11 voting for moderate and three voting for high, which means the committee has passed this portion of the measure.

CO-CHAIR GUNNAR: And now, to use and usability.

MEMBER MOYER: Really there's the same pattern issue we've been discussing. It's not individually reported but it is part of the composite. That composite seems to show a lot of variation on the website, so it's kind of -- seems useful to me, but the individual measure itself isn't listed up there.

CO-CHAIR GUNNAR: So, what's the composite that's -- can you explain a little bit more about the composite, how it's used and what --

DR. PAONE: The composite is the three medications. It's anti-platelet agents, it's postoperative use of beta blockers, and it's post-discharge. That's four. I'm sorry, it's

four.

It's preoperative beta blocker and discharge beta blocker, anti-platelet medication at discharge, and anti-lipid treatment at discharge.

CO-CHAIR GUNNAR: Can you talk about the -- and do you give star ratings for these, as well?

DR. PAONE: There's a star rating for the composite, currently two or three stars.

CO-CHAIR GUNNAR: And given what sort of gaps exist, I mean, we understand that gaps for this generally. Is there --

DR. PAONE: For the -- the gap overall is actually -- and I don't remember exactly. It's broader than the gap for any medicine individually, as you might imagine. Because if you're missing any one of them, it's -- again, it's a dichotomous composite.

You either get all of them or you get, you know, it's a zero or one report. I don't remember off the top of my head what the

difference is, but it is -- clearly it has to be larger than the gap, obviously, for any one of the medicines, because it's a summation, essentially, of the four.

And so, to the extent that, you know -- and this goes back to the discussion we've been having. All right? I mean, we've had this. If usability is defined as being individually publicly reported on our website --

CO-CHAIR GUNNAR: This one is not.

DR. PAONE: It's not. I mean, we've had that conversation. I can't make an argument that it is. It is not.

There are other ways that are important that the data is used. It's used individually by centers, it's used by hospitals to look at, it's used by surgeons.

As a division head I can tell you, for me this was an important domain that we worked very hard at getting to 100 percent on for many, many years.

And when somebody missed one, we

investigated why that patient didn't go home on lipids, or didn't go home on aspirin, because we know these outcomes are important to the subsequent longer-term benefits of the operation.

CO-CHAIR GUNNAR: So -- so --

DR. PAONE: But I can't -- I -- I don't mean to interrupt you, but --

CO-CHAIR GUNNAR: No, no, no, this is exactly right.

DR. PAONE: -- your point is -- what you're going to make is exactly -- is, if the argument or the discussion or the conversation is going to be, is it publicly reported, the answer is, individually it is not.

CO-CHAIR GUNNAR: No, actually, it wasn't, actually.

DR. PAONE: Okay.

CO-CHAIR GUNNAR: I don't mean to talk over you, Gae. I just wanted -- I want you to understand that if -- it depends on how this is -- you know, either the entire thing is topped out -- I mean, we already know that the variance

for this particular component of it is really bumping up against 100 percent. Right? I mean, you're at 99 percent. Your IQR is limited. Right?

And the piece is, is that it's like an assumption that that's going to be 100 percent virtually in the -- in a way, it's -- that I have a more comfortable feeling with saying, it's virtually publicly reported, but it's not -- maybe you just need to take it out of the whole thing. Right?

It's really not adding to the -- maybe it should just be the other three. I'm bringing this up just -- this gets back to either we have a demonstrable gap that we're trying to get to, one person or one facility not being 100 percent is a facility that makes it one star.

It must be, right? To be a one-star means that you're virtually not 100 percent. If I'm tracking this particular measure, you're not meeting it if you're not 100 percent, basically.

MEMBER HANDY: John Handy. When you

go to the website and actually look at this particular category, it's receipt of required perioperative medications, so the conglomerate that was mentioned before, that's got the most variability with regard to three and one stars, of all the four columns that are there, five columns that are there.

CO-CHAIR GUNNAR: And so, the question on the table -- that's great, John -- is, what impact did this particular measure have on that analysis?

DR. PAONE: I can't answer that sitting here. But I'm looking at a one-star program in medications that is at 59.2 percent of the patients receive all four of these medications.

The argument that I would make in terms of, you know, not taking this one out if it's topped off, is because we think they're all important. They're all equally important. And if you take this one out and then the next three get -- you know, you look -- or get better, or --

presumably, the composite would get worse.

But then, one more would get topped out, and now you've got a composite of two. And then, eventually, you've got one, you don't have a composite. Now, you're looking at an individual number.

Does this matter at all? Does this whole perioperative medication, is this important enough to be part of it?

Well, we obviously think so. It's been part of it. This was the -- one of the first -- right? -- was the CABG mortality composite. And it's publicly reported and, you know, it's -- I lost my train of thought here for a second. I apologize.

CO-CHAIR GUNNAR: So, we'll -- any other comments? Thank you for that, Gae. Mark, do you -- other question, or another comment?

MR. ANTMAN: Yes. If I may add, and I apologize, I think it was I who interrupted Dr. Paone's train of thought.

For what it's worth, looking onto the

website, again, it's not possible to drill down to determine how many, or, excuse me, which of the one-star hospitals, again for the perioperative medication domain, it's not possible on the website to determine exactly why they're one star, for which of the four medication measures they have failed, or potentially more than one.

But it may be worth noting that looking as -- after sorting by the star rating for perioperative medications, there are approximately 60 centers that scored a one-star for perioperative medications.

So, it's clearly a problem, even though it's -- the performance is 99 percent for this one medication.

Again, we don't know where the centers are failing for the other medications, but clearly, for an all-or-none domain, an all-or-none composite of perioperative medications, it's a significant quality issue that there are that many centers that are just getting one star.

If I may add briefly, Dr. Gunnar, this -- from the STS staff perspective, having heard the patient perspective today of what is valuable and what is understandable on the website or not, I will note that I have realized in looking at our public reporting pages, that although we've explained multiple times in this meeting what we mean by the major morbidity domain and the perioperative medication domain, and what goes into those domains, at the bottom of the page, Amy, I think you referenced what we say at the bottom of the page earlier.

We say, absence of major morbidity, patients who do not experience any major morbidity. We define major morbidity elsewhere on our website, but we don't define it here. The same with required perioperative medications.

When I go back home to Chicago tonight, and when I'm back in the office tomorrow, I'm going to make sure that my colleagues who work on the website will realize that we need that definition on this page, so

that, at least, is understandable.

So, in that respect this discussion has been very valuable for me.

DR. PAONE: Let me -- I got my train of thought back with the perioperative medications. One of the points I wanted to make was, it's a smaller component of the overall composite score, but it's not an insignificant one.

There are a fair number of programs, and I don't know exactly how many, but it is not -- it is very possible to be two stars in mortality, two stars in morbidity, three stars in perioperative medications.

And because of the small -- and this is one of the reasons for the multi-dimensional domain -- a composite score, that three-star in perioperative medications can drive you to be a three-star program overall for the composite.

Now, you can't be one star in mortality or one star in morbidity, and be a three-star program, because of your medications.

But it is another differentiator in an environment where we're trying, in as simplistic a manner with this three-star composite rating, not ranking, program, to differentiate programs that otherwise would be really gapped together quite closely.

MR. ANTMAN: Thank you. I actually -- sorry. Go ahead, Josh.

MEMBER STEIN: I was just going to say, in my field there are some quality metrics that are like never-events, and, you know, this kind of reminds me of some of those.

And it seems like we're being tasked to evaluate and show, you know, a spread in differences in performance on something that should be a never-event. So, you know, if all the hospitals are doing great at it, there can be consequences of taking that off.

I guess it puts us in a -- as people trying to review these measures in a difficult position if we're being asked to potentially ding a measure because it's not showing us enough

variability, when it's supposed to be a never-event type of thing.

MR. O'BRIEN: This is Sean O'Brien. Can I jump in and just ask whether the bar charts have been displayed that show the performance gaps across the low-, mid- and high-performance groups? Because my impression is that we provide a data that does demonstrate a performance gap.

On a relative scale, there's a 50-fold difference in the frequency of patients who do not receive anti-lipid medications across hospitals in the highest performance group compared to the lowest performance group, but I just want to make sure those data have been presented on everyone's radar when you're thinking about evidence of gap.

CO-CHAIR FLEISHER: Thank you. You know, Mark, I want to thank you for your last comment. And I'm thinking back to Amy and TeMaya and Elizabeth. And where I came from, this in a more -- specific to the language perspective.

But what I've learned today also, is

that David's comments that the public knows how to use this website, as well as how to ask the physicians for more data why they're an outlier, or why they're not a top hospital, doesn't exist.

And the question -- you've always been leaders. You know, it's interesting that there was a law that was before Congress to actually make transparency of quality metrics on individual hospital websites. And I asked Nancy Foster why HA was not supportive, and they said, there needs to be a national website.

So, I'm -- you know, dating back to what I asked in the beginning, but really reframing it from a patient perspective, if you can teach patients how to use your website on your website and how to ask the right questions and understand the components and why you should say, if you're a one-star hospital, or even a two-star hospital, ask them why, and you could get to that other data, that would be a huge advantage from a public reporting perspective.

So, I don't know how you get there

quickly, but that's what I just learned in your response to Amy's comment. And that -- then, to be honest, you would do a great service to the public and to this entire field.

MR. ANTMAN: And I'll add, if I may, the STS has a separate patient portal. I don't know offhand, we could look quickly, I'm not sure what is said on that patient portal with stars explaining these metrics and what goes into these metrics, but that's certainly one thing that I and my colleagues will be looking at closely, be sure that adequate explanations are provided there.

CO-CHAIR FLEISHER: Yeah, because that actually could be part of the plan.

MR. ANTMAN: Yes.

CO-CHAIR FLEISHER: Because, I mean, from my perspective, although I'm supportive, that would make me feel very good that the usability of your data changes overnight by making patients understand how to use the data and how to drill down, even if it's not -- even

for explanatory purposes.

And all the reasons that David nicely outlined and you did, why the composite's the right place to start. It's sort of, you start with the composite.

If you don't look good, you ask. But if you look great, there's actually no reason -- if you're a three-star hospital, there's no reason to go any further, I assume.

DR. PAONE: I just -- I have to comment on that. I don't think that because of a hospital -- did I misunderstand you? I hope I did, that if a hospital's a three-star program, that is absolutely not a reason not to go further in any variety of ways that -- I'm sure I misunderstood you.

CO-CHAIR FLEISHER: I -- yeah, I said it incorrectly. That doesn't mean you can't continue to improve. And one of the nice things is, you've shown how improvement -- all I'm suggesting is that the patient may have, from a patient perspective, of trying to dig down

further may be different than the provider always trying to get better.

MEMBER YATES: Question. Do you still affiliate with Consumer Report?

DR. PAONE: We are not, no longer.

MEMBER YATES: When did that stop?

DR. PAONE: I don't remember exactly. In the last year.

MEMBER YATES: It used to be that it was reported on Consumer Report.

DR. PAONE: That's correct. And it's not any longer.

MEMBER YATES: Because that interface was probably more consumer-friendly, no pun intended.

DR. PAONE: Yeah.

MR. ANTMAN: I can explain that a little bit. We heard from consumer reports in the spring of 2018. Not deciding to do away with STS star ratings specifically, but letting us know that they were closing their health rating center all together.

So, although they still publish information -- they still publish health-related information, they decided that they were no longer going to publish any health ratings, whether it be physician or surgeon ratings, or hospital ratings, or other ratings data that may have been provided by other organizations, because, as they explained to me, they decided that it wasn't of sufficient interest to their readership.

MEMBER YATES: US News and World Report is moving in to the surgeon-specific sphere.

DR. PAONE: Yes, we're aware of that.

MEMBER YATES: So are we.

DR. PAONE: I know we're running late, but I just want to follow up again on the three-star thing and make a quick comment. I -- even from the patients' perspective, because the program's a three-star program, they should still consider the website and its information as a starting point and not the end point, and should

be prepared to ask questions in any area that they wish.

I've got patients who come see me with a folder full of papers from -- and we all do -- from various websites on the Internet, most of which tell them nothing. And I oftentimes -- I've had patients come, and you all have, who said, I researched you on the Internet and, you know, absolutely that's why I'm here.

And when I asked them what they found, none of what they found is the reason that they should be coming to see me. So, we just implore patients in general to get as much information from as many sources as possible, starting with our website.

CO-CHAIR FLEISHER: Thank you.

CO-CHAIR GUNNAR: Any other comments? Discussion? All right. Usability and use.

MS. KOSURI: We are now voting on the use portion for Measure 0118. Voting is now open. The options are possible, no-pos.

(Off-microphone comments.)

MS. SKIPPER: And we're still waiting. Just two votes.

MS. KOSURI: Mm-hmm.

MS. SKIPPER: And, Chris, if you're on the phone and you can hear us, just type in your vote via the app, or you can speak it to us.

So, just one vote shy, so we'll need to reach out and collect the final vote for this measure. But we can still go on and discuss -- well, people are packing up. So --

DR. PAONE: I don't want to be out of place here, but as a point of order, the majority, there's a number of other measures. There are four more individual composite measures, there are three more mitral measures, that are going to have the same exact issue that we've been discussing all along, on use and usability.

So, I don't know that we need to go through this whole conversation, and I leave it up to all of you to decide. But there are -- the aortic and CABG mortality measures, which won't

suffer those, because they are individually reported mortality issues.

So, perhaps we could focus on those and get through those, and then work to the others, if necessary.

CO-CHAIR FLEISHER: That would be great, but we can discuss that.

DR. PAONE: We don't have enough votes. I get it. Okay.

CO-CHAIR FLEISHER: So, there's a couple of interesting things. By having a lack of a quorum, in effect the other measures don't go down at all. They have to go for a call in which we can get --

MS. MARINELARENA: We already have it scheduled.

CO-CHAIR FLEISHER: We already have it scheduled?

MS. MARINELARENA: Mm-hmm.

CO-CHAIR FLEISHER: So, at that call, if we get any insights from Mark, back from STS, where we get insights internally, we will have an

opportunity to vote again on usability, but there's no -- for the new measures we haven't discussed.

MS. MARINELARENA: We will be voting on the measures we have not discussed.

CO-CHAIR FLEISHER: Correct.

MS. MARINELARENA: On February 20th from 12:00 to 2 p.m., Eastern.

CO-CHAIR FLEISHER: That's --

MS. MARINELARENA: I don't know when it is. I have no idea when it is.

CO-CHAIR FLEISHER: That's next Wednesday. Okay, I'll be in California. I'll be on the call.

MS. MARINELARENA: Then, it will be 9 a.m. for you.

CO-CHAIR FLEISHER: Yes. I will miss one conference. Okay. So, do we want to continue discussion, or put it all on the call?

MS. MARINELARENA: Let's put it on the call, because if we continue discussion, the rest of the committee that is not here --

CO-CHAIR FLEISHER: Will not hear that.

MS. MARINELARENA: Yes.

CO-CHAIR FLEISHER: Okay. And, AJ, your comments are acknowledged.

MS. SKIPPER: So just as Melissa said, our next call is scheduled for Wednesday, February 20th from 12:00 to 2 p.m., where we will be discussing the remaining measures.

Following that, we'll be writing up a report of your recommendations. And that'll, again, be posted for a 30-day comment period, opening on March 21st.

And then, we'll bring you all back on May 8th for the post-comment call. Anything else to add, Melissa?

MS. MARINELARENA: No. I just want to thank everyone. I know these are really difficult discussions to have. I know internally as a team we have struggled with this. We've worked a lot with STS. So, we commend you for struggling with this, and, you know you'll help NQF work through these issues.

We appreciate STS being here today and we're glad that you were here in person. It makes that much easier to have these conversations around these really complex issues and really complex measures.

So, we really appreciate your time and your contribution, and we look forward to talking to you next week. And if there's anything we can do, you can resend the information.

If there's anything that you need from us to make it easier to get through the conversation, just reach out to the team and we're happy to put something together.

And, of course, thank you to Lee, who will be in California, so don't really feel bad for him, and to Bill, who just left, for leading us today. Thank you all very much.

MS. SKIPPER: And just one more thing before we close. If there is anyone on the line who'd like to make a public comment, we'd definitely like to give you the opportunity to do so now.

DR. PAONE: If I could, then, just take this opportunity, as well, on behalf of the STS and my colleagues, to thank everyone for their time put in before they got here, in addition to the conversation today.

We don't -- the fact that we had some disagreements on process doesn't mean that we're not completely committed to making this the best possible outcome for everyone involved, and we thank everyone for their efforts.

CO-CHAIR FLEISHER: And thank you very much. You did engage the group incredibly well, and appreciate that, and of course, always appreciate the NQF team for the outstanding work in preparing us and taking us through this difficult conversation.

We will huddle before next Wednesday to help with the discussion. Fred, thank you for not voting but being here to give guidance, as always.

(Whereupon, the above-entitled matter went off the record at 3:32 p.m.)

E

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Surgery Standing Committee

Before: NQF

Date: 02-13-19

Place: Washington, DC

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