NATIONAL QUALITY FORUM

+ + + + +

SURGERY PHASE 3
STANDING COMMITTEE

+ + + + +

WEDNESDAY
AUGUST 17, 2016

+ + + + +

The Standing Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:00 a.m., Lee Fleisher and William Gunnar, Co-Chairs, presiding.

PRESENT:
LEE FLEISHER, MD, Co-Chair; Professor and Chair of Anesthesiology, University of Pennsylvania/American Society of Anesthesiologists
WILLIAM GUNNAR, MD, JD, Co-Chair; Director, National Surgery Program Office, Veterans Health Administration
KARL BILIMORIA, MD, MS, Director, Surgical Outcomes & QI Center; Vice Chair for Quality, Northwestern University and Northwestern Medicine
RICHARD DUTTON, MD, MBA, Chief Quality Officer, United States Anesthesia Partners
ELISABETH EREKSON, MD, MPH, Dartmouth Hitchcock Medical Center
FREDERICK GROVER, MD, Professor of Cardiothoracic Surgery, University of Colorado School of Medicine

JOHN HANDY, MD, Thoracic Surgeon, American College of Chest Physicians

CLIFFORD KO, MD, MS, MSHS, FACS, Director, Division of Research and Optimal Patient Care, American College of Surgeons/Professor of Surgery, Department of Surgery, UCLA School of Medicine, American College of Surgeons/UCLA School of Medicine
BARBARA LEVY, MD, FACOG, FACS, Vice President, Health Policy, American College of Obstetricians and Gynecologists
BARRY MARKMAN, MD, Senior Medical Director Medicaid, Aetna
KELSEY MCCARTY, MS, MBA, Senior Manager, Quality and Safety Program, Department of Anesthesia, Massachusetts General Hospital
LAWRENCE MOSS, MD, Surgeon-in-Chief, Nationwide Children's Hospital
AMY MOYER, Manager of Value Measurement, The Alliance
KEITH OLSEN, PharmD, FCCP, FCCM, Professor and Dean, College of Pharmacy, University of Arkansas for Medical Sciences
COLLETTE PITZEN, RN, BSN, CPHQ, Clinical Measure Development, Minnesota Community Measurement
LYNN REEDE, DNP, MBA, CRNA, Senior Director, Professional Practice, American Association of Nurse Anesthesiologists
CHRISTOPHER SAIGAL, MD, MPH, Professor, UCLA
SALVATORE T. SCALI, MD, Assistant Professor of Vascular Surgery, University of Florida-Gainesville
ALLAN SIPERSTEIN, MD, Chairman Endocrine Surgery, Cleveland Clinic
LARISSA TEMPLE, MD, Colorectal Service, Department of Surgery, Memorial Sloan-Kettering Cancer Center
BARBEE WHITAKER, PhD, Director, American Association of Blood Banks*
A.J. YATES, MD, Associate Professor and Vice Chairman for Quality Management, Department of Orthopedic Surgery, University of Pittsburgh Medical Center

NQF STAFF:

HELEN BURSTIN, MD, MPH, Chief Scientific Officer
ELISA MUNTHALI, MPH, Vice President, Quality Measurement
MARCIA WILSON, PhD, MBA, Senior Vice President, Quality Measurement
JASON GOLDWATER, MA, MPA, Senior Director
KAREN JOHNSON, MS, Senior Director
MELINDA MURPHY, RN, MS, Senior Director
DESMIRRA QUINNONEZ, Project Analyst
CHRISTY SKIPPER, MS, Project Manager
KATHRYN STREETER, Senior Project Manager

ALSO PRESENT:

JAIMO AHN, MD, PhD, Orthopedic Trauma Association
VINAY BADHWAR, MD, FACS, FACC, Society for Thoracic Surgery
JULIA BERIAN, MD, American College of Surgeons
MICHELLE DARDIS, MSN, MBA, RN-BC, The Joint Commission
KATHY DOMZALSKI, MHA, BS, RN, The Joint Commission
BRUCE HALL, MD, PhD, MBA, FACS, American College of Surgeons
GAETANO PAONE, MD, MHSA, Society for Thoracic Surgery
DAVID SHAHIAN, MD, Society for Thoracic Surgery
ARYEH SHANDER, MD, FCCM, FCCP, Englewood Hospital and Medical Center*
JONATHAN H. WATERS, MD, Magee Women's Hospital, University of Pittsburgh
ANN WATT, MBA, RHIA, The Joint Commission

*** present by teleconference**

A G E N D A

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

8:00 a.m.

CO-CHAIR FLEISHER: Okay. Good morning. It is 8:00 according to my iPhone clock, so we are going to get started. I want to thank everyone for coming on time. We have one new member who has joined us this morning.

And can you identify yourself and any disclosures?

You did the disclosure yesterday. I'm sorry.

MEMBER WHITAKER: I did my disclosures yesterday. I'm Barbee Whitaker from the American Association of Blood Banks.

CO-CHAIR FLEISHER: Great. Well, welcome.

And we do not have Dr. Cima on the phone today. Is there anyone else? Are the phone lines open? Yes?

OPERATOR: The line is open.

CO-CHAIR FLEISHER: Great. Do we have anybody from the standing committee on the phone?

OPERATOR: No, we do not.

CO-CHAIR FLEISHER: Great. Thank you.

Incredibly productive yesterday. What was most important is we actually touched on a lot of cross-cutting and difficult issues. And Barbara and I were talking about the need --

(Recorded music playing.)

CO-CHAIR FLEISHER: Yes. As long as it's classical.

Some of the issues that were touched on, we think the Committee did a fabulous job outlining the issues and some of them have actually brought up issues that we'll take to the CSAC for more strategic questions. For example, the last measure we went over yesterday.

We have two hours to get through our colleagues from STS' measures. I would ask -- we did a great job yesterday, but even today to be more effective in limiting any additional comments to not that I agree, which -- but that I have a new area to actually question with regard to the measure developers or for the discussion.

So if we can do that, I think we can get through the entire day. And we must finish by 3:00 since I know lots of people have planes, trains and automobiles even before that.

So nothing further? Who's coming to the table?

CO-CHAIR GUNNAR: Christy, did you want to say anything?

MS. SKIPPER: No, just good morning and welcome again to our second day of the meeting and looking forward to an efficient and productive meeting.

CO-CHAIR FLEISHER: And as always, we want to thank NQF staff both for the preparation before, which was outstanding, and all the work done here.

So if you can introduce yourself and we'll go through the measures. And someone has a conference call. And, Larry, so we'd like to put that first. So if you can identify yourselves. And then if we can start with 134, that would be fantastic.

DR. BADHWAR: Sure. Thank you for your accommodation. My name is Vinay Badhwar. I'm a professor of surgery at WVU, West Virginia University, and I'm humbled by being the Chair of Public Reporting for the STS.

And we thought, just because there's a few new members, that a 90-second just preamble on the STS Database, that it was established in 1989. We now have 6 million patients and over 95 percent of cardiac surgery performed in the United States is registered by the database. The database is a qualified clinical data registry by CMS and for PQRS, and we've got a portfolio of measures that are risk-adjusted outcome model-based and long-standing NQF relationships, as you know.

We have 36 measures, all of which are used for ongoing feedback to our participants for the purposes of quality measurement and quality improvement, and many of these are included in the PQRS system. The three areas to just briefly touch on in 30 seconds are penetration,

completeness and accuracy and the suitability for public reporting.

So the penetration, the database compared to CMS has now 95 percent penetrance and it's been implemented through a robust audit process over the last decade. And now we have an audit of 10 percent. And in the history of our audits there have been no egregious absence and essentially 100 percent, particularly on the most important indicators such as mortality.

And then finally, all the six measures we're going to discuss today are designated for public reporting. And it's such an important element in the STS Database that it's evidenced by -- currently we have in the adult cardiac surgical section 49 percent of sites are participating in public reporting and 57.9 percent of all congenital sites are public reporting voluntarily, which we believe is among the highest in specialties in the United States.

And so today we'll touch upon six measures: three process measures for CABG and

three composite measures, and Dr. Paone, Chief of Cardiac Surgery at Henry Ford and our leader and Chair of the Adult Cardiac Surgery Task Force is going to kick us off.

DR. PAONE: Thank you. Good morning. As Dr. Badhwar said, I have the honor of serving as the Chair of the Task Force on Quality Initiatives for the STS, and I guess we'll start with Measure 134. It's a little out of order, but I will start by saying that there have been and continue to be significant advances in both medical therapy and use of advanced percutaneous catheter-based approaches for treating patients with ischemic heart disease.

And as a result, the number of coronary bypass procedures has decreased significantly over the past two decades, but yet continues to be the most common procedure that we perform with over 150,000 procedures a year. I might also add that in the opinion of many of us it's by far the most highly scrutinized surgical procedure in history.

The first measure that we'll discuss is the use of the internal mammary artery in coronary bypass surgery. This measure captures a percentage of patients aged 18 and older undergoing CABG who received an internal mammary artery graft. There are reasons to exclude patients from consideration, and these include the subclavian artery stenosis, previous cardiac or thoracic surgery, previous mediastinal radiation; both of those may make harvest of the artery difficult or impossible, patients who undergo emergent or salvage procedures, and finally if there is no significant or bypassable disease of the LAD artery itself.

The IMA use has previously been endorsed by NQF and is a very significant component of our STS CABG composite score. Use of the IMA to bypass the left anterior descending artery, coronary artery during coronary bypass surgery is a Class 1B recommendation of the 2011 ACCF/AHA Guidelines for Coronary Artery Bypass Graft Surgery. And in those cases where for some

reason the left mammary is either unavailable or unsuitable for use, it is then a Class 2C recommendation that the right internal mammary artery is probably indicated to bypass the LAD.

The superiority of internal mammary arteries over saphenous vein grafts as coronary artery bypass conduits and the evidence supporting the short and long-term survival benefits of using the left IMA specifically to bypass the left anterior descending artery are exhaustive and well-documented over what is now approaching 30 years.

Long term patency rates of 10-plus years for the LIMA graft when placed in the LAD are routinely in excess of 90-plus percent compared to saphenous vein patency, which approximates 50 percent.

Most importantly, perhaps more than any other --

CO-CHAIR FLEISHER: I think we're very comfortable with the evidence of the superiority.

DR. PAONE: Okay.

CO-CHAIR FLEISHER: So what we're most interested in; and thank you, is the gap and any new evidence, because as you -- we can accept the evidence as it is --

DR. PAONE: Excellent.

CO-CHAIR FLEISHER: -- unless there's anything new that's come out.

DR. PAONE: Okay. Well, in fact then there really hasn't been anything new that's come out except for the fact that this remains really the most important part of a coronary bypass operation and we feel very strongly that it should be continued to be endorsed because we don't want to send a message anything other than that.

CO-CHAIR FLEISHER: And anything on gap that you can comment on? If not, I'm sure the -- Larry, do you want to start --

MEMBER MOSS: Actually John is the primary on this one.

CO-CHAIR FLEISHER: Okay. John?

MEMBER HANDY: Well, there isn't much

of a gap and it's a little bit tough to criticize this particular process measure. It's a maintenance process measure, so we don't even really have to vote on the evidence, frankly, just discuss the gap. The gap is listed in --

CO-CHAIR FLEISHER: Why don't we just very quickly -- would anybody like to vote on evidence?

(No audible response.)

CO-CHAIR FLEISHER: No? Okay. Evidence passes.

MEMBER HANDY: Good. So the gap is listed in a couple of different places. A little bit variously on page 6 it's listed as a low performer as 96 percent, high performer is 99 percent. In other words, the amount of volume left internal mammary arteries used. Page 3293, 100 percent. So the worst is 93 percent. So it's got a high penetrance in cardiac surgery. It's adhered to. And that's true across the SDS groups, too. The lowest was 98 percent. So there isn't much of a gap.

And so, back to what I was getting ready to say before we went and talked about evidence is that it's so central to coronary bypass it's a little difficult to discard it.

CO-CHAIR FLEISHER: Larry?

MEMBER MOSS: I agree completely. My only comment or question was whether we should consider reserve status, and if we did, whether that would negatively impact the composite measure that's so important.

CO-CHAIR FLEISHER: Yes?

MS. MURPHY: With respect to gap, one of the things we talked about yesterday is for some of these measures where the importance of the measure in terms of precipitating or contributing to an adverse event would cause you to think potentially about it differently in terms of retaining one in active endorsement with a low gap.

The other thing is that if you make the determination to put it into reserve status, it is still an endorsed measure, still therefore

eligible to be an integral part of any composite.

DR. PAONE: Well, we would suggest that -- we understand that putting it in reserve does not take it away from the significance with regard to the composite, however, we're not sure that everyone understands the subtle difference between being endorsed and being on reserve.

One of our colleagues has stated that pretty much that probably the most important thing we do as cardiac surgeons and coronary surgeons is use the left internal mammary artery to bypass the LAD, and that goes to evidence. And I know we've accepted that, but we do think that this is probably the most important process measure in cardiac surgery and very strongly feel that we'd benefit from having it re-endorsed rather than just placed in reserve.

CO-CHAIR FLEISHER: Thank you. Amy?

MEMBER MOYER: So looking at the public reporting, which I really appreciate you have online, I know that you take great pains to make sure that you're showing statistically

different -- statistically significant differences. And I'm seeing a range of performance among groups that are publicly reporting on this measure alone, which to me suggests there is gap that can be used to distinguish performance. So I would support that there be -- there's a meaningful gap there.

CO-CHAIR FLEISHER: Thank you for that additional piece of information.

DR. PAONE: I think the average use is very high, but there are still a not insignificant number of outliers. So I think that would be true.

CO-CHAIR FLEISHER: Helpful. Other comments? Christy?

(No audible response.)

CO-CHAIR FLEISHER: No, call for a vote.

MS. WILSON: Desmirra is going to be helping with the vote.

CO-CHAIR FLEISHER: Oh, okay.

MS. QUINNONEZ: Okay. We are now

voting on Measure 0134, use of internal mammary artery in coronary artery bypass graft. Voting is now open for the evidence of 0134.

Sorry. Gaps. Yes, sorry. Voting is now open for the gaps of Measure 0134. Option No. 1 is high; option No. 2 is moderate; option No. 3 is low; and option No. 4, insufficient.

(Voting.)

MEMBER GROVER: Just for the record I want to state that I'm abstaining on all the STS votes.

CO-CHAIR FLEISHER: Thank you. I just wanted to ask staff if we think the gap is low but we still want to -- but we don't want to go to reserve status, do we -- how do we vote? In other words, if we think that it's important to not put it on reserve status. Elisa, if you can help us.

MS. MUNTHALI: So you would vote on performance gap regardless of your desire to have the measure go into reserve status because for a measure that's topped out, there's no opportunity

for improvement. It still needs to be -- all the other criterion still need to be passed. So what would happen is -- let's say you voted low or insufficient. You may say should we consider this measure for reserve status and vote on the rest of the criterion and then come back to answer the question on reserve status. Does that make sense?

MEMBER KO: I just have a quick question. Are there objective definitions, criteria for reserve status? I know we did some last time or a couple times ago, because like, oh, that's topped out, that's -- but what -- is there a definition for topped out?

MS. MUNTHALI: We do, and I don't -- Katie, could you pull that up? I think it would be easier for everyone to see it.

MS. MURPHY: And while she's pulling that up, there is one thing I wanted to say because of something I heard. If a measure goes into reserve status, number one, it is endorsed. It is an endorsed measure in reserve status, so

it does not lose endorsement. In fact, as Elisa just said, it goes through the entire process and it is an endorsed measure placed in reserve status because of a low gap.

MEMBER HANDY: So then what is the natural history of a reserved measure? We've talked about this a little bit in the past, too. I mean, does it get periodically reviewed? I know it doesn't have to be continuously supported by data.

MS. MUNTHALI: So it doesn't go through the three-year maintenance cycle, but because you have ownership, the Surgery Committee has ownership over the portfolio. You may want to revisit it if in the particular topic area you see that there might need to be a performance measure to address certain elements to improve care. So it's something that we will bring in front of you again to look at the reserve status list to see if any of those you would like to revisit given new data, new evidence, anything that you may have that would help to push that

forward.

MEMBER HANDY: And is there a sunset?

MS. MUNTHALI: So reserve status is almost like a sunset. We're seeing the first class of them. These are -- most of them are process measures that have come through our process early on. And so, but there may be an opportunity for them to come back if we see that improvements are going down. Maybe behavior has changed. And so, yes. So we -- but we haven't gotten there yet. This is like the first major class of reserve status measures that have gone through the NQF process.

MEMBER HANDY: Sure.

DR. BADHWAR: Brief comment. I appreciate that definition. It's very helpful. I just want to emphasize what Amy had mentioned, that there is still some practice variability that exists and this full endorsement is so important to practicing surgeons that take a shortcut. It is a little harder to do a mammary artery than a vein graft, and I just wanted to

emphasize that one point.

MS. MUNTHALI: And we just want to emphasize that the measure would still be endorsed. You're not losing endorsement. It's just that it wouldn't go through its periodic maintenance review.

MEMBER HANDY: Thank you.

CO-CHAIR FLEISHER: Rick? Comment?

MEMBER DUTTON: Okay. Just a comment on the reserve history. So we did this a couple years ago with some of the antibiotic measures. I don't know if it contributed or not, but certainly within the next year CMS moved those out of PQRS as being topped out. Is this still a PQRS measure for you guys?

(No audible response.)

MEMBER DUTTON: Okay. That they also told us, Anesthesia, that we had to take it out of our QCDR measure set for CMS, because we initially did that with it. And we did that for one year, but then they told us we had to remove it from that use.

We do still find that antibiotic measure very useful. A lot of our members still report it. It's useful for hospitals and insurance companies and sort of internal quality improvement, so lots of people are still capturing it. But that's been what's happened to the measures that we've --

CO-CHAIR FLEISHER: Thank you.

MEMBER DUTTON: -- seen on reserve.

CO-CHAIR FLEISHER: This is the endorsement reserve status, and why don't we vote? We're up to 20. Could people re-vote just to make sure we get --

(Voting.)

CO-CHAIR FLEISHER: Barbara?

MEMBER LEVY: So I actually have a question about our definition of a moderate/low performance gap. I mean, is that -- are there metrics that we're supposed to use? Because I think it's quite different for us to look at a gap in the STS Database that has penetrance of 95 percent versus a measure that's electively

reported by people who think they do a good job with it. And when we've got 95 percent penetrance and we do have a spectrum of people, I think that's a different discussion than a discussion about a topped out measure that's voluntarily reported among whoever feels like they want to report it.

CO-CHAIR FLEISHER: So I know of no definitions that -- so I -- are you aware? I mean, that's a good discussion that we'll put in our report to send back to us, to you and I to continue.

MEMBER LEVY: Thanks. We're really creating a lot of work for ourselves, aren't we?

CO-CHAIR FLEISHER: Right. Allan? Because I want to move forward.

MEMBER SIPERSTEIN: But I think relative to an individual measure like this, I think we really have to look at is there opportunity for improvement and is there variability? And we've had data presented that says yes.

CO-CHAIR FLEISHER: Thank you.

MEMBER SIPERSTEIN: Even though the number is high.

CO-CHAIR FLEISHER: So let's finish the voting.

MS. QUINNONEZ: All votes are in and voting is now closed. Okay. For performance gaps of Measure 0134 the vote reads 10 percent voted high; 52 percent voted moderate; 38 percent voted low; and 0 percent for insufficient.

CO-CHAIR FLEISHER: So it passes, correct?

MS. QUINNONEZ: Yes.

CO-CHAIR FLEISHER: Okay. So it's still going forward for full endorsement.

Next, John?

MEMBER HANDY: Well, the rest of it's pretty easy because the STS Database -- so reliability with the track record, the construct and the -- all the things that make the STS Database an enviable thing. The reliability is high.

CO-CHAIR FLEISHER: Can we vote?

MS. QUINNONEZ: Voting is now open for the reliability of Measure 0134. Option 1 is high; option 2 is moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the reliability of Measure 0134, 81 percent voted high; 19 percent voted moderate; 0 percent for low; and 0 percent insufficient.

CO-CHAIR FLEISHER: So I just -- like reliability will be the same for any of these measures that are from the database. I would like to -- no? I would like to know if anybody would like to vote separately on reliability for the measures. Or we -- as we come up to each measure I will ask if anyone would like to vote separately.

MEMBER PITZEN: I'm sorry. It was me kind of shaking my head. I mean, the reliability is dependent on the individual metric --

(Simultaneous speaking.)

CO-CHAIR FLEISHER: So we'll go by each measure and say do we need to vote again or can we take the previous vote?

Next.

MEMBER HANDY: So the same comments about validity. The penetrance of the database and the auditing habits or mandated auditing characteristics of the database make it very valid. And this is a very clear-cut. You either did or you didn't use an IMA.

CO-CHAIR GUNNAR: Larry? Comments?

MEMBER MOSS: No.

CO-CHAIR FLEISHER: Any other comments?

(No audible response.)

CO-CHAIR FLEISHER: Let's vote.

MS. QUINNONEZ: We are now voting on the validity of Measure 0134. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the validity of Measure 0134, 86 percent voted high; 14 percent voted moderate; 0 percent for low; and 0 percent for insufficient.

CO-CHAIR FLEISHER: Next?

MEMBER HANDY: Same is true for feasibility. Its penetrance in the adult cardiac surgery arena in the United States is over 95 percent, so pretty much everybody is participating in this database, so it's highly feasible.

CO-CHAIR FLEISHER: Comments?

(No audible response.)

MS. QUINNONEZ: Voting is now open for the feasibility of Measure 0134. Option 1 is high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the feasibility of Measure 0134, 67 percent voted high; 29 percent

voted moderate; 5 percent voted low; and 0 percent insufficient.

CO-CHAIR FLEISHER: John?

MEMBER HANDY: Usability also high.

CO-CHAIR FLEISHER: We heard it's publicly reported.

MS. QUINNONEZ: Voting is now open for the usability and use of Measure 0134. Option 1 is high; option 2, moderate; option 3, low; and option 4, insufficient information.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the usability and use of Measure 0134, 67 percent voted high; 29 percent voted moderate; 5 percent voted low; and 0 percent voted for insufficient information.

MEMBER HANDY: There's really no competing measures for this one. There's a lot of related measures, but none competing.

CO-CHAIR FLEISHER: So any comments before we vote on suitability for endorsement?

(No audible response.)

CO-CHAIR FLEISHER: Okay.

MS. QUINNONEZ: Voting is now open for the overall suitability for endorsement of Measure 0134. Option 1 is yes; option 2 is no.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the overall suitability for endorsement of Measure 0134, 100 percent voted yes; 0 percent voted no.

CO-CHAIR FLEISHER: Thank you very much. We will go back to 117. So what I'd like to focus on is any change in evidence and a little bit on gap, and those would be the primary --

DR. PAONE: Sure.

CO-CHAIR FLEISHER: -- focus to get you --

DR. PAONE: Okay. Well, then I'll eliminate most of what I was going to say and suggest that there have been some recent data that suggests that use of beta blockers post-cardiac surgery has demonstrated a synergistic

and additive effect at reducing mortality when combined with statins at both a year and longer term at 11 years. And overall, patients -- it's thought that in terms of secondary prevention there's a survival advantage, not only in patients with previous heart attacks or heart failure, but also now those with normal heart function. And so this essentially includes all patients who undergo bypass surgery. The role of discharge beta blockers remains a Level -- a Class 1, Level B recommendation for patients with atrial fibrillation to prevent atrial fibrillation and the sequelae thereof.

In terms of a gap, this was considered essentially topped out in 2012 at 95½ percent. It was re-endorsed at that time. And since the gap is really -- remains the same but similar to the previous measure on IMAs, there is still a lower tail, not as much with the discharge as with the preoperative beta blockers that we'll discuss subsequently, but there still remains a little bit of a gap in use.

Added to that is there really is no burden added to the capture of this measure. It remains a component of our composite score and one of 34 NQF-endorsed measures, and we believe that removing it from active endorsement, like the other measures, could suggest that its importance has somewhat lessened. I know we talked just a minute ago about the meaning of "reserve status," that it maintains NQF endorsement, but the added description of it sort of being sunset I think is a little bit concerning in terms of what that may at least portray to those who don't understand the subtle differences.

CO-CHAIR FLEISHER: Thank you. Lynn, I think you're the primary and Kelsey is the secondary.

MEMBER REEDE: Thank you. In reference to the gap, it's still for the --

CO-CHAIR FLEISHER: How about just --

MEMBER REEDE: Okay. Well, he was --

CO-CHAIR FLEISHER: -- any comments on

evidence?

MEMBER REEDE: Evidence is fine. Good to go.

MEMBER McCARTY: So I do have a question about the evidence. The guidelines that were submitted for this measure were from 2011, and then this measure was endorsed in 2012. There haven't been any updated guidelines since then for CABG, but in 2014 the same group: ACC and AHA, published guidelines for non-cardiac surgery.

And so where I'm going with this is that in those guidelines beta blockers given on the day of surgery are flagged.

CO-CHAIR FLEISHER: I'm the Chair of that Guideline committee --

MEMBER McCARTY: I saw that. I did.

CO-CHAIR FLEISHER: -- and it's not relevant to this --

MEMBER McCARTY: It's not? Okay.

CO-CHAIR FLEISHER: -- population.

MEMBER McCARTY: So as a layperson

there's -- it perfectly makes sense why it would be so effective --

CO-CHAIR FLEISHER: It's a different --

MEMBER McCARTY: -- in CABG and dangerous outside of that. Okay.

CO-CHAIR FLEISHER: And to disclose on the task force that oversees all the guidelines, this --

MEMBER McCARTY: That was my --

CO-CHAIR FLEISHER: -- they're not in conflict.

MEMBER McCARTY: That was my question.

CO-CHAIR FLEISHER: So would anybody like to vote on evidence or -- we'll continue. Okay. Can you, Lynn, discuss gap?

MEMBER REEDE: So for gap it does appear to be topped out except that in the 10th decile it's gone from 73 percent in the 2013 to '14 data down to 15 percent -- or 50 percent in the '14 to '15 data. So there is some gap still remaining in the measure.

CO-CHAIR FLEISHER: Kelsey, any thoughts?

(No audible response.)

CO-CHAIR FLEISHER: Any comments?

(No audible response.)

MS. QUINNONEZ: Voting is now open. We are voting on Measure 0 --

CO-CHAIR FLEISHER: Collette?

MEMBER PITZEN: I can comment now or later. I just have a general comment about performance gap in general and quality improvement. And I know I'm in the minority. I have a hard time with a measure that's nationally at 98 percent, serving its purpose for moving that quality needle forward. Granted, everybody would want 100 percent, but when you're comparing and trying to get better, it seems like there's a lot of other places where we could be spending our data collection resources and efforts. So I'll try to refrain from those comments throughout because I know everyone feels differently.

CO-CHAIR FLEISHER: But it was important to -- thank you for making the comment as we --

DR. PAONE: I would just -- if I could, just very quickly respond to that. And I understand the idea of not wanting to waste resources, but capturing this measure and bringing it forth in our -- in data set really is of no particular burden. It's in the same list as three other medicines and it's on the same data set, and I don't think the computer recognizes that extra data variable as a burden either.

CO-CHAIR GUNNAR: But as chair and as a matter of process what defines a topped-out measure? If that doesn't define a topped-out measure, we'll be at a loss for defining a topped-out measure.

CO-CHAIR FLEISHER: Barbara? And then Kelsey.

MEMBER LEVY: So I think again we need to look at the STS 95 percent penetrance

differently than we would look at other types of measures. And I think a measure can be for moving the needle, but it can also be for maintaining the needle, to maintaining the quality that we have established and holding people accountable for a certain standard.

And so I think we need to think about it, Collette, maybe just a little bit differently in that I think the concern that I'm hearing is that if it is in reserve status or not endorsed, that that sends a public message that says this is less important. And there are certain things that we want people to do every single time. And I think this may be one of those things.

DR. PAONE: I would just comment that I think that's exactly what we think.

CO-CHAIR FLEISHER: Go ahead.

CO-CHAIR GUNNAR: But again, reserve status and non-endorsement are 180 degrees from one another.

CO-CHAIR GUNNAR: Lynn and Kelsey, did you have other comments on this, or no?

MEMBER McCARTY: Slightly different aspect of the gap. I was going to bring this up in validity, but I think it overlaps with gap, and that's that part of the measure allows the surgeon to indicate that beta blockers are contraindicated.

So my first question is can they just write "contraindicated," or are there a set of specific criteria that indicate that it's contraindicated?

CO-CHAIR FLEISHER: Sounds like -- that's validity, correct?

MEMBER McCARTY: So the reason I bring that up in gap is because I'm curious. If they just write "contraindicated" and doesn't have to be supported by a specific criteria, I'm curious what percentage of the registry passes the measure by listing that, because beta blockers have kind of had a long history of controversy. And so, if someone doesn't agree in giving it and can just write that down as this -- it's contraindicated -- and then I'm curious what the

reliability testing is around that. Then there may actually be a larger opportunity for improvement in giving beta blockers than this percentage indicates.

DR. PAONE: Well, as is described, there are clearly contraindications clinically to the use of beta blockers. They generally are in this analogy, which is fairly unusual, but the clinical ones are basically bradycardia or in some patients with lung disease reactive lung disease that may be an issue for some of the beta blockers. There is some data in the report that describes looking at the measure with and without the exclusions and there are some slight differences, but nothing that would suggest on a large scale that physicians are just randomly deciding they don't like beta blockers and writing "contraindicated."

Certainly in response to your first question, yes, I mean, theoretically we can just write a note in the chart that says it's contraindicated. And the measure requires a

documentation, not just that it's contraindicated, but the reason for that. And so, I don't think that this is a particularly large problem in terms of the measure itself.

CO-CHAIR FLEISHER: Cliff?

MEMBER KO: Just a short thing about the topped out and reserve. I think that we have to rework those -- the NQF should help us rework those things, because if reserve is similar to sunset, then this -- definitely not. But if this reserve is topped out at 98 percent, all the measures that we put on reserve status last time were just these numbers exactly.

So when we think about it one way versus the other, then we'll think about reserve or not. For this, if we're thinking about sunset, it probably should not be, but the numbers would suggest that it should. So we need some clarification from NQF.

CO-CHAIR FLEISHER: Great. So unfortunately we don't have -- I think, Marcia, we need to go back to the CSAC for some insights

on this, or do you have comments? It would really be helpful.

DR. WILSON: What I will mention is we hear this argument -- I shouldn't say that. We hear this discussion in every committee meeting. There are great concerns about putting measures in reserve and what message that sends. As you heard from my colleagues, reserve status does not remove endorsement. It is an endorsed measure. It can be brought back, as Elisa said, for review or if something were to change about that measure, or if there would be an indication that performance on that measure was not maintained at the level when it was put in reserve.

So I don't have an easy answer for you because this is somewhat of a philosophical divide that we see in the committees where members are reluctant to put a measure in reserve because of the message that it's sending, and yet at the same time at -- the other argument would be at 98 percent, as one of our Co-Chairs said, this measure is topped out. So it is -- you have

to make the decision yourself, but it is a philosophical divide that we see often in our committees, if that helps.

MEMBER LEVY: So it helps and it doesn't help. When we hear the consequences from the NACOR registry that measures that were put in reserve status were pulled out of PQRS and that CMS has decided --

CO-CHAIR FLEISHER: I think CMS had already signaled they were pulling it out when we put it on reserve status.

MEMBER LEVY: Okay.

CO-CHAIR FLEISHER: Because I already was aware, unless you are aware of something different.

MEMBER LEVY: No, it like happened like --

CO-CHAIR FLEISHER: Yes, because I was already talking to CMS.

MEMBER DUTTON: Yes, I don't think there's cause and effect there, more of a shared perception. But the fact that we couldn't keep

it in the QCDR meant it had a huge economic impact on our membership and that they had to go find another measure to report for PQRS. And that's the real elephant in the room here, is if reserve status means that you can't report it for your federal pay-for-performance requirements, there is a huge impact on all of the physicians, all of the eligible providers who have to report these measures. And that's one of the reasons the discussion gets so difficult.

DR. PAONE: If I could make just one last comment, and that's even at 98 percent, if that's what the average number is, it's still -- and there's many thousands of patients in the database. And so, that's still a fair number of patients that theoretically are being discharged without beta blockers, without contraindications, and at the present time the data is still fairly significant that discharge beta blockers have a survival advantage. So we would want to take every opportunity to make certain that this stays at the forefront and is an important part of our

composite and our data collection.

CO-CHAIR FLEISHER: So we have a couple comments, and we want to get through all the measures. My one advice to you, your last statement's an interesting one that isn't in the report, but when you come back it may be useful to say the gap represents X number of patients nationally. Would people agree? And that would really help in defining that it's really not topped out. It really -- opportunity for improvement I guess is a different way of saying -- and the implications for improvement.

Yes, Larry and then -- not, not Larry. We've got A.J. and, yes, Allan.

MEMBER SIPERSTEIN: Okay. So the whole reason this reserve status came about is the whole way that we define the gap and necessitating an opportunity for improvement as opposed to thinking about it in a little bit of a wider scope, saying this is an important area to maintain or to continue to monitor even though the compliance may be high. So I think we may

need to think about kind of broadening our definition of, kind of, what constitutes a gap.

CO-CHAIR FLEISHER: A.J.?

MEMBER YATES: Yes, I think there's more to it than just the number, 98 percent. When we put the antibiotic process measure on reserve, there was already literature in the health economics literature showing that it didn't improve inequality at any particular hospital, that it was something that could be done without improving quality and that it was just a checklist sort of thing that you could do.

I would argue that the difference -- the average readmission and complication rates in the joints that we talked about yesterday was 4 percent, so why is 98 percent so much different than 96 percent? I'd love to argue top those out, but I think there's a difference in terms of there being performance and quality using that measure.

The two things I would say about this is, one, does the two percent non-compliance

track with less well-performing cardiac centers? Are they lower-star centers? And the second thing is, is that looking at the trend over the last few years they actually had a worsening of utilization of the beta blockers in those people that -- in the lower 10th percentile, so they were actually going backwards.

So we're not talking about a static situation. We're talking about something that hit the wall and came back and they're retracting. And if it correlates with the quality of the center, then I -- or their overall quality, then I think it has some reason to be continued to be captured.

CO-CHAIR FLEISHER: Okay. I would like to move forward. Just remember that the CSAC did say for -- if you consider this a never or what should occur as a low-frequency event, you can consider the gap being important. So that's the counter-argument. Whether you think never missing an antibiotic is one of those events, that's how you should potentially think

about it.

Can we call for a vote?

MS. QUINNONEZ: We are now voting on Measure 0117. Voting is now open for performance gaps on Measure 0117. Option 1 is high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the performance gap of Measure 0117, 0 percent voted high; 62 percent voted moderate; 38 percent voted low; and 0 percent insufficient.

CO-CHAIR FLEISHER: Okay. We will keep going. Next, Lynn?

MEMBER REEDE: Reliability?

CO-CHAIR FLEISHER: Yes, any differences from the previous measure?

MEMBER REEDE: No. Looked at the measure score. Signal-to-noise was good. Ready to go.

CO-CHAIR FLEISHER: Okay.

MEMBER REEDE: It's not risk-adjusted.

CO-CHAIR FLEISHER: Any comments?

(No audible response.)

CO-CHAIR FLEISHER: Would anybody like to vote on this separately from our previous vote?

(No audible response.)

CO-CHAIR FLEISHER: No. Can we -- do we have the right to continue our previous -- okay. Validity? Right?

MEMBER REEDE: Is that enough?

CO-CHAIR FLEISHER: Yes.

MEMBER REEDE: You want to discuss more?

CO-CHAIR FLEISHER: Validity, any comments?

MEMBER McCARTY: I have one more question about validity. So I know this is suggested to be a companion measure to preoperative beta blockers. My question is that for whatever reason if preoperative beta blockers aren't given, and for the most part they are;

we'll get to that gap in a minute -- but is it dangerous to prescribe a beta blocker where you can't monitor the effects of what happens when that patient goes home and takes it?

CO-CHAIR FLEISHER: Essentially unintended consequences is what you're asking?

MEMBER McCARTY: I guess so.

DR. PAONE: I mean, I -- realistically there's always potential consequences any time that we prescribe a medicine to any patient at any time. I'm not sure what that has to do with the measure per se.

MEMBER McCARTY: Well, I guess I --

(Simultaneous speaking.)

DR. PAONE: We follow up. I mean, the patients are followed by their surgeons, by their cardiologists, by their internists after any cardiac procedure, and we do take their heart rate, their blood pressure, ask if there are any problems. I'm not quite sure how else to answer that. I mean, I can't --

MEMBER McCARTY: Let me rephrase the

question.

DR. PAONE: Sure.

MEMBER McCARTY: So I mean, I think that the way that this has been studied has been mostly like a beta-blocker regimen. So you start it preoperatively. You continue it at discharge. And is there any evidence to suggest that if the first half of that process is missing that if they don't start to get the drug in their system prior to surgery that there is a risk of them getting it after? So it's just a risk question more than an unintended consequence question.

DR. BADHWAR: Brief response to that. It's a good question, however, there's a dose-response to any medicine, and particularly starting post-operatively, it's well known that we are cautious in discharge beta blockers. They don't start them on the max dose if they've never seen it before. So I would just respond to that as dose response like any other medicine?

MEMBER McCARTY: Thank you.

CO-CHAIR FLEISHER: Okay. Would

people like to vote for -- would anybody like to vote?

(No audible response.)

CO-CHAIR FLEISHER: Yes? No?

(No audible response.)

CO-CHAIR FLEISHER: No? Okay. We'll carry on the previous vote for validity.

Next?

MEMBER REEDE: Feasibility. Collected the same as the previous data.

CO-CHAIR FLEISHER: Any comments?

(No audible response.)

CO-CHAIR FLEISHER: No? Would anybody like to vote separately on this measure?

(No audible response.)

CO-CHAIR FLEISHER: Okay. Next? Use and usability.

MEMBER REEDE: Usability and use. Same.

CO-CHAIR FLEISHER: Any difference?

(No audible response.)

CO-CHAIR FLEISHER: Anybody like to

vote separately?

(No audible response.)

CO-CHAIR FLEISHER: No. Okay. We're next at suitability for endorsement.

So, Desmirra, can you --

MS. QUINNONEZ: We are now voting on the overall suitability for endorsement of Measure 0117. Option 1 is yes; option 2 is no.

(Voting.)

MS. QUINNONEZ: Looking -- all votes are in and voting is now closed. For the overall suitability for endorsement of Measure 0117, 100 percent voted yes; 0 percent voted no.

CO-CHAIR FLEISHER: Okay. Moving on to 0127, preoperative. So any changes in evidence and comments on the gap?

DR. PAONE: No particular changes in evidence worth noting, but I will comment on the gap. It remains actually more significant than the previous two measures. I think the lower levels are down in the low 80s despite the high average. And so for reasons that we've discussed

for the previous two measures, but particularly with regards to an ongoing gap I think we would again strongly endorse, recommend endorsement.

As far as the -- well, I should go back and say for different -- for new information there is some controversy around this measure in the sense that recent studies looking at patients from the database have a large trial of over 500,000 patients looking at those with and without beta blockers.

Preoperatively there really was little difference in outcomes in that particular trial, and in that same journal Dr. Shahian, who is here in the audience today, had an invited commentary where he addressed many of the issues related to why those differences may have occurred.

One of the most significant differences in the -- the fact of the incidence of atrial fibrillation with and without beta blockers, that's known to occur in 25 percent of patients, carries a great deal of morbidity. And over the years, 30-plus articles have clearly

defined beta blockers as being of benefit in that population.

We have over the past iteration of the data set, data collection vehicle, added an additional data variable looking at beta blockers for greater than two weeks preoperatively in addition to the one that's before us here today, which was less than 24 hours. And the reason for adding that data point was to try to further define the true benefits of this therapy. And in addition to believing that beta blockers still confer a benefit, the addition of that data set will afford an opportunity to study this more intensely going forward. And that's just another reason that we would like to maintain focus on this measure.

CO-CHAIR FLEISHER: Kelsey, any comment?

MEMBER McCARTY: Nothing new to add on evidence or gap.

CO-CHAIR FLEISHER: Lynn, David, do you want to make any comments?

DR. SHAHIAN: No, I think --

CO-CHAIR FLEISHER: Okay. So would anybody like to vote on evidence?

(No audible response.)

CO-CHAIR FLEISHER: No. Gap, Kelsey?

MEMBER McCARTY: Just, I mentioned that it was in the 80s, the data that I had from the measure worksheet is that it was in the 80s back in 2012, but has risen to 93.5 now. So it is quite a bit higher, but I still think that shows opportunity for improvement.

CO-CHAIR FLEISHER: Okay. Any other comments?

(No audible response.)

CO-CHAIR FLEISHER: No. Why don't we vote?

MS. QUINNONEZ: We are now voting on Measure 0127. Voting is now open for the performance gap of Measure 0127. Option 1 is high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. The results read 14 percent voted high; 81 percent voted moderate; 5 percent voted low; and 0 percent voted insufficient.

CO-CHAIR FLEISHER: Sorry, Collette. Did you have any comments? I'm sorry. I missed --

MEMBER PITZEN: That's okay. My previous comments about performance gap apply here as well.

CO-CHAIR FLEISHER: And when the report comes out, remember everybody gets to see the report that goes to CSAC. Please, if there are sections that you want to emphasize something that the CSAC should consider either for this measure or in a greater sense, please make sure that that's clearly articulated in the report to reflect your concerns.

Okay. We're now up to reliability. Kelsey?

MEMBER McCARTY: Nothing new to add.

MEMBER PITZEN: I have a question. So please describe to me the numerator with the beta blocker less than 24 hours before discharge. Is that an understanding perhaps that the patient is on the beta blocker preoperatively and that's continued in that dose the day of surgery given, or what is that time frame? And when I'm looking at the actual data in the database, unlike the discharge medication this just simply has a check box for contraindicated without any reference to criteria.

DR. PAONE: So the numerator is all patients who receive beta blockers within 24 hours of surgery, and that's whether they've had it for two weeks, two months or just started the day before.

MEMBER PITZEN: Okay. So is it likely or not likely that they would be given a beta blocker the morning of surgery or for a very first dose?

DR. PAONE: Yes, that's a very good question. It's a great question. It's -- the

ongoing controversy is in that question.

So in order to satisfy the measure, it was not uncommon, and this was one of the criticisms of the study that I mentioned is that we don't know the difference between a patient -- the benefits of beta blockers in a patient who's been on it for two weeks or two months versus a patient who comes into the hospital the morning of surgery and is given a dose a beta blocker. We -- and that is not -- has not been uncommon, frankly, because it does satisfy the data set. And that's where the new data variable added will hopefully provide additional information.

The thought among many is it's probably not the dose necessarily or even the medicine but the physiologic response, and so there are other variables in terms of the specific medication given, how it's given and then whether or not the patient has responded to it. So there are some questions with this that I think really is why we'd like to maintain it as an endorsed measure and enable us to continue to

look at this issue and study it further as the number of patients increases in the database given the new added variable.

MEMBER PITZEN: So the way it -- as it is now, it could be an unintended consequence that relates to patient safety. Is that correct, or is that taking it too far? I'm just thinking of the patients perhaps that have never been on a beta blocker.

DR. PAONE: I mean, I guess theoretically there's always a potential for an unintended consequence, but that certainly hasn't been the experience.

CO-CHAIR FLEISHER: Rick?

MEMBER DUTTON: Yes, I do this about once a week, Collette. Same thing. Here's a patient not on beta blockers scheduled for surgery. It's indicated. There's a box in my EHR that says I have to decide am I going to give it or not. And different people play that different ways.

Generally, we look at the patient. If

they're hypotensive or bradycardic already, we would not give it and say contraindicated. That's what that box means. But it can be for a number of different reasons based on my assessment. And there are some anesthesiologists, honestly, who would give a drop, a homeopathic dose to satisfy the regulatory requirement, but not have any different physiologic effect, if that makes sense.

DR. PAONE: I will add just on a practical basis, there are not that many patients that come into the hospital for coronary bypass surgery not on beta blockers without a contraindication, and one of the reasons to give that dose of medicine isn't necessarily that they haven't been on it, but it's because you can't document that they've taken their last dose within 24 hours. So even though you know they're on it, you just can't find that documentation anyway. So the easiest thing to do is just give them a dose and then you write the time down in

the medical record and now you've satisfied the data set.

CO-CHAIR GUNNAR: Okay. Would anybody like to vote on reliability?

Collette, would you like us to do a separate vote on reliability? Does it matter?

MEMBER PITZEN: Sure.

CO-CHAIR FLEISHER: Sure. Why don't we do it?

MS. QUINNONEZ: Voting is now open on the reliability of Measure 0127. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the reliability of Measure 0127, 33 percent voted high; 57 percent voted moderate; 10 percent voted low; and 0 percent for insufficient.

CO-CHAIR FLEISHER: Kelsey, validity?

MEMBER McCARTY: Nothing new to add.

CO-CHAIR FLEISHER: Since there was no

discussion, I will ask if the -- my habit now will be is if there's discussion points, we'll vote. If there's no discussion, we'll ask if anyone would like to vote.

Is there any desire to vote?

(No audible response.)

CO-CHAIR FLEISHER: No. Let's move on. So now we're onto feasibility.

MEMBER McCARTY: Same as before.

CO-CHAIR FLEISHER: Collette?

MEMBER PITZEN: May I just ask a question again?

CO-CHAIR FLEISHER: Please.

MEMBER PITZEN: So when I think of feasibility, I guess I'd like to ask the developers how many of the participating institutions have really a direct pass-through from their electronic record to your registry, because when I look at the data elements that are collected in the registry, it assures my mind a little bit if there's a lot of hospitals that have this already built into their EHR without

the added abstraction that it would take to get this information.

DR. PAONE: Well, I can answer that as someone who works very closely with the data manager at our institution, and we have over the last two years instituted the use of the Epic electronic health record. At our institution the pass-through is zero at this point, but we are actively working with our IT people to institute that.

One of the issues for us actually, unrelated and just very quickly, is we've just had to change our data set. The manufacturer of the program that we use has just changed, so we're in a six-month process of just changing over the data set.

But that's a great question. And everyone that does this is looking for ways to automate this, and we are as well, but I don't know what it is, what the penetrance of automated data transfer is nationally, but we don't do it. I would guess it's very low, yes.

MEMBER PITZEN: Thank you. And the reason I ask that question again is the cost-benefit of data collection versus the output. And there is -- you guys have great measures. I'm just talking about some of the process ones that are topped out and then the future potential to build new great measures from your registry.

DR. PAONE: Yes, just very quickly I would add for the same reason as the others, there's really no added burden to checking off a box for pre-op beta blockers.

CO-CHAIR FLEISHER: Yes, David?

DR. SHAHIAN: Just in terms of automated extraction, it is an important point and we have explored multiple avenues to make that possible. Needless to say the EHR manufacturers have not been particularly interested in these efforts so far. We're continuing to work with them. But there are major differences between the unstructured data fields in EHRs and the highly structured, highly defined data fields we have in clinical

registries. That's what makes clinical registries so powerful.

And we caution people about trying to extract data from EHRs, particularly some of our more complex definitions. It's very difficult to do it even with natural language processing and all those sorts of things. So it's an interesting area. We'd like to reduce data collection burden. It isn't happening much so far. So, but good question. Thank you.

CO-CHAIR FLEISHER: Would you like -- shall we vote then, since there were questions?

(No audible response.)

CO-CHAIR FLEISHER: Go ahead.

MS. QUINNONEZ: Voting is now open for the feasibility of Measure 0127. Option 1 is high; option 2 is moderate; option 3 is low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the feasibility of Measure 0127, 57 percent voted high; 38 percent

voted moderate; 5 percent voted low; and 0 percent for insufficient.

CO-CHAIR FLEISHER: Great. A.J.?

MEMBER YATES: This is not intended just for STS, but STS routinely points out how reasonable their price is to participate and how it's underwritten and everything else, but that's the cost of being in the registry. The cost of uploading to the registry is not insignificant and can run six figures in terms of the people you have to support to do all that for you and make sure that your i's are dotted and your t's are crossed.

So I think for registry data that's being put in that's not EMR direct pass-through, which I would agree is not the sine qua non of excellence; it may be that it's not as good as we would like, it would be great if there was an estimate of what the real true cost to the hospital is for participating in the registry, not just the cost to -- the contract cost with STS. And I think that would be a reasonable

thing to expect from another developer, so I put that out as one of the things on the wish list that we can add to the conversation for down the road.

CO-CHAIR FLEISHER: Thank you.

CO-CHAIR GUNNAR: Just for perspective for the group, what's the number of data fields that have to be extracted in place per case? I think it was 178 this last time.

DR. PAONE: I honestly -- and maybe David knows the exact number. It's probably more than --

DR. SHAHIAN: About 200 --

DR. PAONE: Yes.

DR. SHAHIAN: -- or 250, something like that.

DR. PAONE: Yes, I think so. Yes.

MEMBER YATES: Which I respect immensely but -- and I think that's where registries have to go, but I think you have to recognize that there is a cost per each institution, and it's not just the 3,000 for

being contracted with the registry.

DR. SHAHIAN: And I think for valve repairs you've actually expanded that even farther, right?

CO-CHAIR FLEISHER: So -- and we recognize that it is fortunate the STS model is to get the hospital to underwrite this. So I think it's a very good point and we'll put it into the report and you'll be responsible for making sure it's well said as far as --

MEMBER YATES: Yes, sir, and I will emphasize the fact that I really think the hospital should pay for it.

(Laughter.)

CO-CHAIR FLEISHER: Thank you.

MEMBER YATES: For other registries as well.

CO-CHAIR FLEISHER: And that's a -- the only reason is I hear that, but you guys were -- you were the first, you were the best in initiating.

Use and usability? Comments?

MEMBER McCARTY: Nothing to add.

CO-CHAIR FLEISHER: Nothing? Anybody want to vote?

(No audible response.)

CO-CHAIR FLEISHER: Okay. So are we up to voting for the measure?

(No audible response.)

CO-CHAIR FLEISHER: Okay. We are on time.

MS. QUINNONEZ: We are now voting for the overall suitability for endorsement of Measure 0127. Option 1 is yes; option 2, no.

(Voting.)

MS. QUINNONEZ: All votes are in; voting is now closed. For the overall suitability for endorsement of Measure 0127, 95 percent voted yes; 5 percent voted no.

CO-CHAIR FLEISHER: Okay. Now we get to actually the composite. So 3030.

DR. BADHWAR: So Composite Measure 3030 is a multi-procedural/multi-dimensional composite measure on the STS individual surgeon

composite measure for adult cardiac surgery. It includes five major operations: isolated coronary bypass grafting; isolated aortic valve replacement; AVR plus CABG; isolated mitral valve repair or replacement, or what we term as MVRR; and MVRR plus CABG, and it comprises a risk-adjusted operative mortality domain and the risk-adjusted major morbidity domain consistent with all of our previous measures. And those domains are five: major complication composite, prolonged ventilation, deep sternal wound infection, permanent stroke, renal failure and re-operations of cardiac origin such as bleeding, coronary graft occlusion, valve disfunction, et cetera.

This includes seven previous NQF-endorsed risk-adjusted mortality outcome measures and five endorsed risk-adjusted complications as noted. In the scientific aspects of this composite, a lot of effort was made to assure its reliability. In similar models it's a Markov chain simulation, 95 percent Bayesian probability interval. The evidence missing is only 0.4

percent for mortality and 0.3 percent for morbidity. And the composite consists of the reliability of each of these.

And just for frame of reference, which I won't be duplicative in the next comments, for CABG it was 0.77 reliability, isolated AVR 0.52, AVR CABG 0.5, isolated MVRR 0.58, and MVRR plus CABG is 0.5. And so this data set is three years, from July 2011 to July -- sorry, June 2014 involving 2,286 surgeons, 621,489 cases, and it is the highest reliability of all of our measures ever performed at 0.81.

And so, with this the objective is -- it's an implicit weighted score, and we propose this for our next composite measure for adult cardiac surgery.

CO-CHAIR FLEISHER: Great. Thank you. An important point; I just checked with Melinda, while they do mention the Star Ratings, we do not endorse the way they actually report it. We endorse the measure. But whether it's one star, two star and how they define the stars is not

part of our endorsement process.

DR. BADHWAR: I intentionally didn't mention that.

CO-CHAIR FLEISHER: Thank you, but it was sitting in front of me, so I had to make a comment --

DR. BADHWAR: No problem.

CO-CHAIR FLEISHER: -- for the Committee to recognize, because that would probably take another three-hour debate about how they actually defined that.

So we have -- our discussants are Karl and Barbara. Karl?

MEMBER BILIMORIA: Great. So that was a good overview and covers a lot of it. I think under the evidence category, mortality is clearly decreasing and they make a nice point about being able to focus on quality of patients who survive. And so this composite does that. The complication rates still remain pretty high with 13.7 percent prolonged ventilation. Mortality still remains 2.3 percent. So I think the

evidence is there for measuring these outcomes. As mentioned, all 12 of the components are NQF-endorsed already.

CO-CHAIR FLEISHER: So any comments, because Barbara -- can we vote on evidence?

(No audible response.)

MS. QUINNONEZ: We are now voting on the evidence for Measure 3030. Option 1 is yes; option 2 is no.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the evidence of Measure 3030, 100 percent voted yes; 0 percent voted no.

MEMBER BILIMORIA: All right. So for the performance gap I think it's pretty obvious they had 9 percent of surgeons being noted to be worse than expected and 18 percent better. Clearly there's variability here that's meaningful.

CO-CHAIR FLEISHER: Melinda is just going to comment about one thing.

MS. MURPHY: The component measures are specified for analysis at the group or practice level, and this measure is applicable at the individual physician level. So we had a fair amount of discussion internally about that because the question is the endorsement or the consideration of the component measures. And the discussion was, if it is suitable for use at a group or practice level, you could have a group or practice of two physicians and what is the difference between applying it to two versus one? So I only say that as a point of the way in which the component measure is endorsed and the way this measure will be used.

MEMBER BILIMORIA: I think it's more about the specifications that have already been well established in those prior metrics, and we can talk a little bit about some of the differences by surgeon as we move on for this composite. But as I mentioned, we have a good performance gap undoubtedly, so should we -- are there any other comments?

CO-CHAIR FLEISHER: Comments?

MEMBER LEVY: Just in general, and this is more of a clinical comment than an evidence comment, but cardiac surgery is a team sport. It's very unlikely that an outcome is related, particularly a morbidity or mortality outcome is related to an individual. And so, I would just ask the developers why report this at an individual level? Why did you decide to report this at an individual level rather than at the group level?

DR. BADHWAR: I will make a brief comment on that. That's a very insightful comment, and we agree with you that cardiac surgery is a team sport. However, currently in the United States, surgeon-level reporting using claims data is going on, and it is an unfortunate situation because that data is often not as accurate as the STS Database and the clinical registry that we have. And so we wanted to form this in the most scientifically valid way we possibly can to address that. And it's already

something that we are being asked for, and to do so in the most responsible way forward is what the objective was.

CO-CHAIR FLEISHER: Thank you. But you also report it at the group level, correct?

DR. BADHWAR: This particular measure is not. This is an individual surgeon measure.

CO-CHAIR FLEISHER: Okay. There is a debate actually at the Board when I was briefly on it of the consumerism approach of they want individual practitioner-level measures. Most of us around this table agree it's a team sport. Being an anesthesiologist, very much a team sport in many of these things. Intensivist. But that debate is beyond the purview of this Committee alone.

A.J.?

MEMBER YATES: Yes, I share the same concerns as Barbara, and in the past, STS has rejected the idea that they could do this at the individual level because of small sample size analysis and the like.

I have a question and then -- a real quick question, then I want to just make a point. What's the exclusion criteria by volume?

MEMBER LEVY: Yes, it's 100 cases over 3 years.

DR. BADHWAR: Thank you. Yes.

MEMBER YATES: Okay. So it has to be 100 over 3 years?

DR. BADHWAR: There's a minimum.

MEMBER YATES: So it's pretty low. It's pretty low in terms of 50 a year?

DR. BADHWAR: Correct.

MEMBER YATES: Or, excuse me, 30 a year.

DR. BADHWAR: Yes.

MEMBER YATES: My one question, and I would have to delve into the actual measure specifics more than time allows, but do you take the sum total of reports from a particular institution? And since you're not reporting this by institution, do you take the sum total of the results and then use that as a risk factor for

poorly performing institutions that may not have the best intensive care units or the best cardiologists or the best anesthesiologists; not that any anesthesiologist isn't perfect, Dr. Fleisher --

(Laughter.)

MEMBER YATES: -- but do you not -- do you take -- do you do -- risk-adjust for the environment that the surgeon is working in since it is a team sport? And I -- because I would think that that would be an important feedback loop to make sure that they're not analyzed in such a way that it takes the environment out of the picture.

DR. BADHWAR: It's another insightful comment. While we have site-specific data, this is tracked directly to the surgeon ID number, and there are many situations -- unlike major academic institutions, there are many situations where surgeons operate in different institutions and this -- the purpose of this measure is to track by surgeon, though your points are valid.

MEMBER YATES: Well, those effects of those institutions though -- because they have to be reporting registered institutions for those to count, those could be proportionally weighted.

DR. BADHWAR: That's valid.

CO-CHAIR GUNNAR: So just to be clear, the purpose of this, one of the purposes of this was if I'm operating at four institutions, this is a collection of all of my cases for that period of time at all of those institutions?

DR. BADHWAR: Correct.

MEMBER YATES: I'm sorry to follow up without being called on, but one of the unintended consequences of not doing a -- risk-adjusting for that is that, of those four institutions, a really good surgeon realizes that one of them is one where it's really hard to get as good a result as he can other places, yet that's an isolated location and his expertise is all they're going to get. There's a chance that he'll stop going to that institution if it's going to bring down his reporting. And I just

throw that out for consideration.

CO-CHAIR GUNNAR: So this goes back to NQF again about is the measure -- this measure is -- actually at its core is about ongoing professional practice evaluation. It's not about quality improvement at any one institution. And so philosophically is that in line with the mission and vision of the NQF?

CO-CHAIR FLEISHER: See, I would actually argue that that surgeon should stop operating at that institution that's dragging down his numbers, and that's what the public wants. The public wants us to say, if I look bad because of one institution, then I'm going to say to the institution I can't operate there anymore. You shouldn't be doing cardiac surgery unless I'm --

(Simultaneous speaking.)

MEMBER YATES: -- administration of that institution and say I'm taking my patients and going elsewhere unless you improve the quality of your ICU, your cardiac anesthesia and

so forth. So there is an opportunity to provide quality improvement.

CO-CHAIR FLEISHER: Right. Thank you. We're saying actually the same --

MEMBER YATES: Yes.

CO-CHAIR FLEISHER: -- thing. Rick and --

(Simultaneous speaking.)

MEMBER DUTTON: Does your -- as you present this as an individual result, does your statistical model include hierarchical modeling of the facility? So is the facility affecting the risk-adjustment?

DR. BADHWAR: I believe so, yes.

MEMBER DUTTON: And asked the other way, if you were reporting this at the facility level; so you're reporting to a facility their performance on this composite, would you adjust it for the surgeon?

DR. BADHWAR: So at this point that's not part of our formulation to have a surgeon-level variable, though that's an interesting

point.

MEMBER BILIMORIA: No, I mean as I think that the public wants it, surgeons want it, this -- some things drive institutional improvement, and this will drive surgeon-level improvement even if you can't be responsible --

(Simultaneous speaking.)

CO-CHAIR GUNNAR: I don't know if surgeons want it.

(Simultaneous speaking.)

MEMBER BILIMORIA: -- the entire outcome --

(Simultaneous speaking.)

CO-CHAIR GUNNAR: There's no vote on that. I mean, nobody raised their hand. I mean, one of the interesting things about public reporting -- what's the penetrance now for public reporting requests?

DR. BADHWAR: Forty-nine percent for --

(Simultaneous speaking.)

CO-CHAIR GUNNAR: Forty-nine percent.

And that's at the group or at the --

DR. BADHWAR: That's at a site participant level.

CO-CHAIR GUNNAR: -- site participant level.

DR. SHAHIAN: Can I just make one comment?

CO-CHAIR FLEISHER: Yes, David, but it would be -- go ahead. Can you come up? You're allowed. He is a Board member.

(Laughter.)

CO-CHAIR FLEISHER: So if you could come up here?

DR. SHAHIAN: Just informational, just in terms of the philosophy of developing this measure, the comments are absolutely correct that we have historically avoided surgeon-level measures for the exact reasons that were noted: sample size and the fact that it's a team sport. But the problem is that there are entities out there that are publishing grossly flawed data and we've seen two very good examples of it just in

the last year.

Several of us in the room have been involved in critiques of some of these grossly flawed measures, and we wanted to have something that was scientifically valid that if we were in a position of having to provide surgeon-level data, we wanted to have the best possible measure. That's what we've tried to do with this measure which encompasses about 80 percent of a typical cardiac surgeon's workload through these five procedures. So having five procedures, two domains, which each of them themselves have multiple measures within them, and then three years of data, that's why we have this very high reliability.

So it's not a measure that frankly we probably would have developed had not these other forces beyond our control sort of pushed towards, but we do have now a measure that we have a great deal of confidence in.

CO-CHAIR GUNNAR: So are you currently reporting the measure?

DR. SHAHIAN: No, our plan with this measure is to provide this individually to surgeons only, their own results, for 6 to 12 months and make sure that they're comfortable with it, and that there's nothing that we haven't thought of, although we've spent two years developing this. So I think it's pretty unlikely. But we're going to do this confidentially to the surgeons for about a year, and only then and probably only if we're pressed are we going to start publicly reporting it. But we want to have it ready. And I would not be at all surprised that the way things are moving that there will be public reporting of this measure within a year.

CO-CHAIR FLEISHER: So just to -- in the use and -- in the committee that looked at intended use this idea of having the end user be able to review the data was one of the major themes that came out of it. So the process you're talking about is actually this idea, if I'm thinking -- remembering correctly -- Marcia

is shaking her head -- that this would be the process that you might get to NQF endorsement-plus, that we wanted the -- so you're actually foreshadowing the idea that we want -- that once measures are out, we'd like it publicly reported. But that would be what could be the next stage of the game is people being comfortable with the data.

DR. SHAHIAN: Yes.

CO-CHAIR FLEISHER: Larry and Barry. We need to keep moving because they have to leave.

MEMBER MOSS: Just wanted to offer an alternative point of view. I agree that surgery is a team sport, but a patient goes to the operating room not with an institution, but with an individual surgeon and needs to choose an individual surgeon. And every one of us knows in this room, regardless of how uncomfortable it might make us, that there is variation in individual surgeons' talent and performance.

So I support what you're doing. I

think it's the right thing to do. We may not be able to ideally compare surgeons, but it's a step in the right direction and I support you.

DR. BADHWAR: Thank you.

CO-CHAIR FLEISHER: Thank you. And to be clear, from the consumerism standpoint that's the way they think about it. Who takes how to look at the team and whether the surgeon has to -- if they're not having a great result, look at the whole team, so to the comment you made.

Barry?

MEMBER MARKMAN: Do you an education process with poor outliers? I mean, you have the data, so what do you plan to do with it?

DR. BADHWAR: That's part of the process of feedback reports and looking at and tracking quality improvement, but that's why this is -- your point is important. We don't have an answer yet, but maybe we will in a few years.

CO-CHAIR FLEISHER: So let's vote on performance gap.

MS. QUINNONEZ: Voting is now open for

performance gaps of Measure 3030. Option 1 is high; option 2 is moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: Looking for one more vote. If you could resubmit your votes, please?

(Voting.)

CO-CHAIR FLEISHER: Okay.

MS. QUINNONEZ: Okay. Voting is now closed. For performance gaps of Measure 3030, 60 percent voted high; 40 percent voted moderate; 0 percent for low; and 0 percent for insufficient.

CO-CHAIR FLEISHER: Okay. Next?

MEMBER BILIMORIA: So we need to touch on the quality construct. Is that the next piece?

(No audible response.)

MEMBER BILIMORIA: So I think --

CO-CHAIR FLEISHER: Next is --

MEMBER BILIMORIA: Yes, for the composite.

CO-CHAIR FLEISHER: -- the composite.

MEMBER BILIMORIA: Yes, the quality constructs of the composite.

CO-CHAIR FLEISHER: Right.

MEMBER BILIMORIA: So if it was -- as mentioned, it gives a pretty comprehensive view of the individual surgeon's practice. It's 80 percent of their practice. Clearly, mortality and morbidity are important. The weighting and the approach to putting together the composite is well described, and we can talk about it below in validity, but it was vetted by an expert panel. And so, I think it's well described at this point.

Any other points that we should touch on in the composite construct?

CO-CHAIR FLEISHER: Any other comments?

(No audible response.)

CO-CHAIR FLEISHER: Okay. Let's vote.

MS. QUINNONEZ: Voting is now open for the composite quality construct of Measure 3030. Option 1 is high; option 2 is moderate; option 3,

low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: Voting is now closed. For the composite of Measure 3030, 85 percent voted high; 15 percent voted moderate; 0 percent for low; and 0 percent insufficient.

CO-CHAIR FLEISHER: Karl, do you want to continue?

MEMBER BILIMORIA: Yes, so moving on to reliability. I think the specifications are -- they're well specified. They're the same specifications from prior NQF measures. They do audits, as mentioned, on about 10 percent. And in terms of formal reliability testing, that was well covered here using multiple approaches. And the reliability for surgeons with greater than 100 cases was 0.81. And so, I think that's probably one of the best that we've seen so far. So I have no issues on reliability.

CO-CHAIR FLEISHER: Do you think this is different? It's a composite. So let's vote.

MS. QUINNONEZ: Voting is now open for

the reliability of Measure 3030. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the reliability of Measure 3030, 86 percent voted high; 14 percent voted moderate; 0 percent, low; and 0 percent insufficient.

MEMBER BILIMORIA: All right. Moving on to validity, there are clearly significant and meaningful differences in performance between providers. Again, the expert panel reviewed which measures were included under weights. Missing data is a small problem at 0.4 percent. Missing data and all the analyses around the missing data do not really change the results. It's a 0.99 correlation with and without the missing data imputed.

The one question I had here was that they mention that the SDS factors could not be examined here, but they examined them in some of

the other subsequent factors. I know this is the surgeon-level model, but I thought if you could just comment on how you addressed -- or why you didn't address the sociodemographic status.

DR. BADHWAR: So we wanted to focus strictly on the major issues of complications and mortality. And so, if I'm understanding your question correctly, I mean, these are fairly defined measures, and so we didn't delve into the items that may or not have -- while I respect your comment on socioeconomic factors influencing outcome, we've done separate analyses on the influence of that.

And in terms of the weighting of importance for this type of measure, we actually looked at different areas of the country and the impact was not huge.

MEMBER BILIMORIA: I would agree that the theoretical underpinning for -- to include SDS is not there, but I was wondering if you did examine it in this particular measure and see whether it had an influence.

DR. SHAHIAN: We did not do it in the manner we did it for example for our readmission measure where I think there is a much more plausible relationship between some of the sociodemographic factors and the outcome. We actually think that for morbidity and mortality that relationship is a little more questionable. And we've been working on this measure actually for about two-and-a-half years and actually did most of the original analytic work even before NQF considered the possibility of including SES factors. So we did not have it.

We could, although frankly dual-eligible status, for example, for our readmission measure showed essentially no impact. I think the granularity of the data that we have and most registries have for these sociodemographic factors is probably inadequate to demonstrate a difference and it's probably going to require things like geocoding, which we don't have available.

MEMBER BILIMORIA: To be clear, I

don't think that it is necessary here, but I just wanted to bring it up since it was pointed out.

CO-CHAIR FLEISHER: Barbara and then A.J.

MEMBER LEVY: I just think from a theoretical standpoint the risk-adjustment for clinical factors should correct for that.

CO-CHAIR FLEISHER: A.J.?

MEMBER YATES: Yes, unlike my concern more about an urban population yesterday, my concern here is geographic disparities in terms of isolation. And having lived in Wyoming for a while, I worry about a town like Billings or Missoula or someplace like that where there may -- those might even be too big, but there are places where people with chest pain have to see a cardiologist, have to have a cath and the cath lab has to have a cardiac surgeon nearby. And if it's a smaller hospital with very wide geographic distances between it and the next center, there is a risk of mortality or morbidity for not being able to get the patient somewhere where this sort

of service exists. And you run the risk of those isolated areas.

Maybe they score poorly. Maybe they're one of the 18 percent of the cardiac surgeons that score poorly and they realize that they can't win there and they leave. Then the cath lab shuts down and there's no cardiologist anymore and all of a sudden there's a big hole in the geography of the United States for that rural environment.

So if I was to throw out anything as being a potential risk of disparity, it would be the question of geographic isolation. And there are places where things like chest pain happen even though you live way out on a farm somewhere and it may not be possible to get to the next state in time. And I would hate to see something drive that availability. Even if it's at a slightly little level of performance, let them try and improve, but I worry about cardiac surgeons leaving that environment because of feeling chastised.

CO-CHAIR FLEISHER: Thank you. Other comments?

(No audible response.)

MS. QUINNONEZ: Voting is now open for the validity of Measure 3030. Option 1 is high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the validity of Measure 3030, 52 percent voted high; 48 percent voted moderate; 0 percent low; and 0 percent insufficient.

MEMBER BILIMORIA: In respect to the scientific acceptability for the composite, they did look at correlations between morbidity and mortality and how much drove the overall score. It was mostly driven by morbidity. That's also where the weight is, but that also makes sense given their underlying premise. Also, the weights were done empirically and also then validated by the expert panel, so I think there

is acceptability around how the weighting was done for composite.

CO-CHAIR FLEISHER: Other comments? Let's vote.

MS. QUINNONEZ: Voting is now open for the composite scientific acceptability of measure 3030. Option 1 is high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. Seventy-one percent voted high; 29 percent voted moderate; 0 percent, low; and 0 percent, insufficient.

MEMBER BILIMORIA: Moving onto feasibility. This captures most of the surgeons in the country and most of the programs. All the data are already collected and well demonstrated to be collected in a standard fashion. Question? Go ahead.

CO-CHAIR FLEISHER: Collette?

MEMBER PITZEN: I just have a general comment. We had talked about the size of the STS

Database. I was pretty sure the cardiac data elements was around 800. I just double checked again. It is 800 data elements for the cardiac surgery database. So again, as a comment in general, not picking on STS, again the cost benefit of collecting data. And I just want to say really unrelated to this measure, I am supportive of this composite.

DR. SHAHIAN: Very quick response. There are probably more than 800 elements available, but not every one of them is coded for every case and for a typical CABG or valve, much fewer than that. Many of the elements are child fields of parents, so once the parent field is coded as no, all those child fields go away.

MEMBER PITZEN: I understand that. It's still resource time to look in the record and verify that initial process that the fields don't apply.

CO-CHAIR GUNNAR: That was A.J.'s point. It requires an extractor.

CO-CHAIR FLEISHER: Interesting

question for NQF for how we put this in perspective. I don't think we've ever looked at cost as part of our mandate.

DR. SHAHIAN: Having just spent $1.2 billion at partners to implement Epic, that is not without cost either.

MEMBER DUTTON: Lee, just very quickly, the metrics that could be on there for developers to show would be very helpful. The total number of fields has been discussed, but also the throughput of one abstracter in a year. How many cases does one abstracter look at in a year or put another way, how many abstracters do I need to hire?

DR. SHAHIAN: We know that. It averages between 300 and 500 cases per year per abstracter, depending on the overall complexity of the case mix.

MEMBER PITZEN: Thank you.

MEMBER BILIMORIA: I think we should be a little careful about what we criticize. Yesterday, we were picking on administrative data

and now we're criticizing what we were calling the gold standard yesterday. This is why we like the registries.

CO-CHAIR FLEISHER: I was reminded by Melinda and Lisa and this is under feasibility is the cost and the burden and so it is something if this committee wants to see and it sounds like we do the next time, want to see that as part of our considerations. But I also heard you, Collette and Karl.

So other comments? No. Can we vote?

MS. QUINNONEZ: Voting is now open for the feasibility of measure 3030. Option 1 is high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

Looking for two more votes.

(Voting.)

All votes are in and voting is now closed. For the feasibility of measure 3030, 48 percent voted high; 48 percent voted moderate; 5 percent voted low; and 0 percent voted

insufficient.

MEMBER BILIMORIA: And then with respect to usability and use, again it's in use, most surgeons. And there hasn't been sort of a national revolt yet, so I think it's in good shape.

CO-CHAIR FLEISHER: But not publicly reported yet.

MEMBER BILIMORIA: Not publicly reported, not this particular measure, but some of the components are.

CO-CHAIR GUNNAR: When did you start giving positions, this data?

DR. BADHWAR: This will commence in --

DR. SHAHIAN: Later this year.

CO-CHAIR GUNNAR: So it has not come out to be clear.

CO-CHAIR FLEISHER: Amy.

MEMBER MOYER: So two things on this, first, thank you for the measure. Hopefully, I didn't need to disclose. We've been cheering this on from the sidelines, but not involved in

the measure.

I have a pro and a con. One of the things that I appreciate about STS is that you allow people who participate in your registry to use their data as they see fit. That's something all registries do and so that makes it easier for them to use it and for us to work with them to use in our programs.

We've already been discussing with organizations in our state how this might be publicly reported and one of the things we've heard back is that apparently there's a significant cost to them to license the data from STS to formally publicly report. I've heard this from the Wisconsin Collaborative for Healthcare Quality.

So for them, that's a usability factor that they're at least pushing back on us with. I know it's not free. I can't imagine how much it costs to develop this, so what are you going to do? But just throwing that out there.

DR. SHAHIAN: Once a program receives

its scores, they can do anything they want with it. So, for example, if they state work a mandate that to do cardiac surgery in your state you have to provide the state with your data, you have the right to do that. You can do with your data, whatever you want.

MEMBER MOYER: And I think that's awesome.

DR. SHAHIAN: Yes, okay.

CO-CHAIR FLEISHER: Allan.

MEMBER SIPERSTEIN: Specific question for the developer on the unintended consequence of surgeons being shy to operate on the very sickest of the patients. I know that you address that quite nicely for the individual components of this measure when we last reviewed them, but has that been looked at for this particular measure? Particularly, because we're now dealing with individual attribution.

DR. BADHWAR: A general comment that the concept of risk aversion, of course, is ever present and something of major consideration for

all of our measures in the entire public reporting initiative to which your question reflects, so it's an important one.

However, the robustness of the risk models really adjusts for all these, the sort of I do the highest risk operations, so I'm not going to do it because I'm being publicly scrutinized. And that's something of communication we focus on. So the risk models are very focused on adjusting for those types of high-risk patients to avoid the risk-aversion behavior.

DR. SHAHIAN: One of the things about risk aversion is that we always think of it as negative, but there are three things that can happen in a risk-averse environment. One is that you can avoid operating on patients who should be operated on. That's bad. The second thing is, however, that you can avoid those patients who are hopelessly ill. And the third thing is that there's evidence from New York State that in a public reporting environment that patients are

better matched to programs and surgeons who are able to take care of them. So you may make the surgeon of lesser ability risk adverse to take on a higher-risk patient, but that patient's probably going to get taken care at a much better place for their problem. So risk aversion -- we had to be careful about how we defined that. It sometimes has advantages actually.

CO-CHAIR FLEISHER: So what I'm hearing, since this is about to be rolled out as in their maintenance a year from now, we'd like for you to address how the plan is rolled out and the risk aversion issue as part of the one year and when it comes back for maintenance. Those should be in the database of what we want to see for re-endorsement.

Other comments before we vote on use and usability? We're defining excellent points which I think is part of our job. Is a part of our job.

MEMBER TEMPLE: I clearly missed something. Why is it coming back in a year, this

measure?

CO-CHAIR FLEISHER: All measures come back with a report from maintenance. It usually just goes to staff.

MEMBER TEMPLE: I see.

CO-CHAIR FLEISHER: No?

(Laughter.)

MS. MUNTHALI: So we have an annual update process which allows developers to come to us if there are changes to the measure. They're typically coding changes, minor changes that won't change the specifications significantly.

If those changes do change the specification significantly, that would trigger an ad hoc review and that would come back to the committee.

CO-CHAIR FLEISHER: Thank you. But that points out as we define some of the things we're concerned about, if there's a change, it will come back to the committee. If it's on track with addressing your concerns, then it could be just signed off or we might hear about

it.

Are we ready to vote?

MS. QUINNONEZ: Voting is now open for usability and use of measure 3030. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient information.

(Voting.)

All votes are in and voting is now closed. For the usability and use of measure 3030, 43 percent voted high; 52 voted moderate; 5 percent voted low; and 0 percent insufficient information.

CO-CHAIR FLEISHER: Okay, what's next? Endorsement. Okay, voting for endorsement. Any last comments? No. Let's --

MS. QUINNONEZ: Voting is now open for the overall suitability for endorsement of measure 3030. Option 1, yes; option 2, no.

(Voting.)

Okay, all votes are in and voting is now closed. For the overall suitability for endorsement of measure 3030, 95 percent voted

yes; 5 percent voted no.

CO-CHAIR FLEISHER: Thank you. We have 15 minutes to get through the last two, so 3031. Robert and John.

DR. BADHWAR: Measure 3031 --

CO-CHAIR FLEISHER: Robert is not on today, right?

MEMBER HANDY: I'm the primary discussant therefore.

CO-CHAIR FLEISHER: Okay, 3031.

DR. BADHWAR: Measure 3031 involves mitral valve repair and/or replacement as a composite score. And after CABG/AVR, AVR/CABG, isolated mitral valve operations are the most common procedures performed.

From July 2011 to June of 2014, 62,118 patients underwent mitral valve repair or replacement plus or minus tricuspid valve repair or a surgical ablation procedure or a PFO or ASD closure which is considered to be part of the mitral operation. Further discussion can happen on that if you wish.

Of those patients, we looked at a two domain composite of operative mortality and absence of major morbidity. I won't be repetitive, it's the same morbidity items of 5 as mentioned earlier.

The missing evidence of this data field was 0.5 percent for mortality and 0.4 percent for morbidity. The c-statistic was .746 for morbidity and .807 for mortality. The scaling of the measure was .74 for mortality and .26 for major morbidity.

Looking at the three-year reliability testing, we settled on one case per month or 36 cases involving 462 sites or 52,841 patients who were arriving at a reliability of .58 for the measure. That concludes my summary.

CO-CHAIR FLEISHER: John.

MEMBER HANDY: The evidence is good for this. I mean it's just got face validity with regard to this being a composite score that covers all the adverse outcomes possible.

CO-CHAIR GUNNAR: As this is a new

measure, I think we have to vote on the evidence.

MS. QUINNONEZ: We are now voting on measure 3031. Voting is now open for the evidence. Option 1, yes; option 2, no.

(Voting.)

Okay, all votes are in and voting is now closed. For the evidence of measure 3031, 100 percent voted yes; 0 percent no.

CO-CHAIR FLEISHER: Gap.

MEMBER HANDY: So for gap was the as-expected performers had a respective mortality and measure of 3.2 and 16.9 percent and the lower than expected performers were double both of those. So that's a substantial gap.

Actually, the higher than performers were about half of each of those.

CO-CHAIR FLEISHER: Comments? Let's vote.

MS. QUINNONEZ: Voting is now open for performance gap of measure 3031. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

All votes are in and voting is now closed. For performance gap of measure 3031, 50 percent voted high; 50 percent voted moderate; 0 percent for low; and 0 percent insufficient.

MEMBER HANDY: So for the composite construct, the reason for putting this measure forward is that the mortality associated with mitral valve surgery is low. And this gives you a more protean picture of the surgical experience. So I think this is a high quality construct.

MEMBER EREKSON: So I have to say I really like this measure in combining the morbidity and mortality and giving us a little bit more information. If I had a wish list of everything that STS could do for us, when they lead the way in showing us to look at these quality metrics, I would love to see more functional status six months and a year down the road in future measures. And I know that's probably something that you guys would love to

capture as well. But I really like this measure in combining the mortality and the morbidity.

DR. BADHWAR: We share your comments.

CO-CHAIR FLEISHER: Great. Are we ready to vote?

MS. QUINNONEZ: Voting is now open for the composite for measure 3031. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

Looking for three more, two more votes. Okay, all votes are in and voting is now closed. For the composite of measure 3031, 84 percent voted high; 16 percent voted moderate; 0 percent for low; and 0 percent insufficient.

CO-CHAIR FLEISHER: Yes, A.J.

MEMBER YATES: We're in middle of this, but the disparity information could use a real quick explanation because there's a lot of P values less than .001 and it looks to me like those are all correlating with being under 65. But if we could go to the disparity list, it's

the chart with the -- it's under gap -- right there. Go up.

And am I to understand that all those things are patients under 65 or does ethnicity have a significant factor?

DR. BADHWAR: So that goes into the logistic regression model as you can see there. So we adjust for those items to your point.

MEMBER YATES: Right. We're looking at the order. I don't want to drag it out, but is -- are we interpreting --

CO-CHAIR FLEISHER: Is race separate from the insurance? So the insurance status is one characteristic and then race is a separate one, correct?

DR. BADHWAR: Correct.

CO-CHAIR FLEISHER: It's not related to the above or below 65?

DR. BADHWAR: That's correct.

MEMBER YATES: So race did stand out as something that was important in this case?

CO-CHAIR FLEISHER: Yes, but this is

what we want to see so that we can correct it and be transparent about this.

DR. BADHWAR: Exactly correct.

MEMBER YATES: It went by. It was hard to read the chart.

CO-CHAIR FLEISHER: Thank you. Reliability.

MEMBER HANDY: Reliability was high for the three years' worth of data that was tested. You have to have a case cut off of 36 cases so it's about a case per month over 3 years was 0.58.

CO-CHAIR FLEISHER: Other comments? No. Shall we vote?

MS. QUINNONEZ: Voting is now open for reliability of measure 3031. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

CO-CHAIR FLEISHER: Validity.

MS. QUINNONEZ: All votes are in and voting is now closed. For the reliability of

measure 3031, 79 percent voted high; 21 percent voted moderate; 0 percent low; and 0 percent insufficient.

MEMBER HANDY: So for validity that was sort of demonstrated by looking at fairly consistent performance over the two time frames that were looked at. They were overlapping time frames, 2011 to 2014 and 2012 to 2015. So that was considered therefore valid.

CO-CHAIR FLEISHER: Comments? No.

MS. QUINNONEZ: Voting is now open for validity of measure 3031. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

All votes are in and voting is now closed. For the validity of measure 3031, 58 percent voted high; 42 percent voted moderate; 0 percent voted low; and 0 voted sufficient.

MEMBER HANDY: Nothing new to add regarding feasibility.

CO-CHAIR FLEISHER: Would anybody like

to vote on that? Is it similar to the others?

MEMBER HANDY: Probably because it's a new measure we --

CO-CHAIR FLEISHER: Okay, we'll go ahead and vote.

MS. QUINNONEZ: Voting is now open for feasibility of measure 3031. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

All votes are in and voting is now closed. For feasibility of measure 3031, 63 percent voted high; 26 percent voted moderate; 11 percent voted low; and 0 percent voted insufficient.

MEMBER HANDY: Same is true for usability.

MS. QUINNONEZ: Voting is now open for usability and use of measure 3031. Option 1 is high; option 2 is moderate; option 3 is low; and option 4, insufficient information.

(Voting.)

All votes are in and voting is now closed. For usability and use of measure 3031, 58 percent voted high; 42 percent voted moderate; 0 percent for low; and 0 percent for insufficient information.

CO-CHAIR FLEISHER: Okay, Melinda pointed out that many of these outcome measures have face validity and therefore their validity cannot be high. So the answer is that the maximum is moderate.

Is there any who are uncomfortable moving any high to moderate as a top level? So essentially all high votes will now be reflected as moderate votes where that is the top option. Comments?

CO-CHAIR GUNNAR: Congratulations to those who voted moderate.

CO-CHAIR FLEISHER: They got it right.

CO-CHAIR GUNNAR: They got it right.

CO-CHAIR FLEISHER: That gets them another six years on this committee. No.

DR. SHAHIAN: Can I just make one

comment? I don't believe that face validity has been the only type of validity demonstrated by the developers.

Construct validity, for example, the correlation of the rating categories with other measures' predictive validity. And that was mentioned. So I don't think face validity was the only measure of validity that we demonstrated.

MS. MURPHY: Okay, so we'll go back and look. What we have from the submissions, as we understand them is that that's where they sit at this moment, but we'll go back and check and if that is the case, we'll correct it with the group and reflect that.

MEMBER BILIMORIA: I would agree with that in the measures that we reviewed. It was not just case validity. There were other things mentioned in the documents.

CO-CHAIR FLEISHER: Next. Suitability for endorsement. Please vote.

MS. QUINNONEZ: Voting is now open for

overall suitability for endorsement of measure 3031. Option 1 is yes. Option 2 is no.

(Voting.)

All votes are in and voting is now closed. For overall suitability for endorsement of measure 3031, 100 percent voted yes; 0 percent voted no.

CO-CHAIR FLEISHER: Okay, we are now down to the final measure, 3032. We have three minutes within the time frame of when we're supposed to be closed. So we'll assume that that is an ex-surgical closure, so we have about 15 minutes left when we do the timing.

(Laughter.)

DR. BADHWAR: They have an hour to wake up their patients.

CO-CHAIR FLEISHER: I know that, too.

DR. BADHWAR: We have faster PAs where we are.

(Laughter.)

CO-CHAIR FLEISHER: We've been exchanging this all week. So if you could just

briefly go into that.

DR. BADHWAR: So briefly after the aforementioned operations previously discussed, mitral valve repair/replacement plus CABG is the next most common operative procedure for measure 3032. And this involves a similar data set from July 2011 to June 2014 where 26,355 cases were analyzed and for at least 10 patients over the 3-year period.

The missing evidence was .55 for mortality, .44 for morbidity, the same two domain composite exactly as the previous measure.

The scalability of the composite was .78 for mortality, .22 for morbidity with a c-statistic of .708 for morbidity and .738 for mortality, arriving at a reliability for 25 cases over three years and 341 sites or 18,924 patients at .50.

MEMBER HANDY: Me still. Nothing to add. This is the same measure, just with coronary bypass thrown in.

CO-CHAIR FLEISHER: So -- and this is

a new measure.

MEMBER HANDY: A new measure also. New composite measure.

CO-CHAIR FLEISHER: So we will need to vote.

MS. QUINNONEZ: Voting is now open for measure 3032. We're voting on evidence. Option 1 is yes. Option 2 is now.

(Voting.)

Voting is now closed. One hundred percent voted yes for the evidence of measure 3032 and 0 percent voted no.

CO-CHAIR FLEISHER: Next.

MEMBER HANDY: In the gap, this was as expected performers had a mortality and morbidity of 6.5 and 29 percent, respectively. And it was for those that were lower than expected performers. That was about double and for those that were higher than expected it was not quite half, actually. It was just somewhat approved, 4.3 versus 19.8 percent. So there is a gap.

CO-CHAIR FLEISHER: Any comments?

Ready to vote.

MS. QUINNONEZ: Voting is now open for performance gap of measure 3032. Option 1, high; option 2, moderate; option 3, low; option 4, insufficient.

(Voting.)

All votes are in and voting is now closed. For performance gap of measure 3032, 63 percent voted high; 37 percent voted moderate; 0 percent voted low; and 0 percent voted insufficient.

MEMBER HANDY: So for the quality construct the same comments apply to this measure, too, is that mortality is low and therefore a more global protean picture of combined mortality and morbidity, gives you more accurate picture of the performance. Quality construct is good.

CO-CHAIR FLEISHER: Comments? No.

MS. QUINNONEZ: Voting is now open for composite of measure 3032. Option 1, high; option 2, moderate; option 3, low; option 4,

insufficient.

(Voting.)

All votes are in. Voting is now closed. For the composite of measure 3032, 74 percent voted high; 26 percent voted moderate; 0 percent for low; and 0 percent insufficient.

MEMBER HANDY: So for reliability of the three years' worth of data for the participants that had at least 25 cases, the reliability was 0.5.

CO-CHAIR FLEISHER: Is this any different than --

MEMBER HANDY: A little smaller number with the last measure I think it was 0.57.

CO-CHAIR FLEISHER: So do you recommend voting?

MEMBER HANDY: Yes.

MS. QUINNONEZ: Voting is now open for reliability of measure 3032. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

All votes are in. Voting is now closed. For the reliability of measure 3032, 55 percent voted high; 45 percent voted moderate; 0 percent for low; and 0 percent insufficient.

MEMBER HANDY: So the same comments about validity for this measure apply for the last two measures.

CO-CHAIR FLEISHER: Further comments about our discussion with validity?

MS. MURPHY: They should go ahead and vote. Given the discussion, you should -- if you're voting as opposed to carrying the other votes forward, you should continue to vote high/moderate because there is the issue that we have to resolve.

MS. QUINNONEZ: Voting is now open for validity of measure 3032. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

CO-CHAIR GUNNAR: But this is a test.

MS. QUINNONEZ: All votes are in and voting is now closed. Sixty percent voted high;

40 percent voted moderate; 0 percent low; and 0 percent insufficient.

CO-CHAIR FLEISHER: Construct.

MEMBER HANDY: Same comments as the last two measures.

MS. QUINNONEZ: Voting is open, is now open for composite of scientific acceptability of measure 3032. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

Looking for two more votes.

(Voting.)

Voting is now closed. For the composite construct of measure 3032, 78 percent voted high; 22 percent voted moderate; 0 percent voted low; and 0 percent voted insufficient.

MEMBER HANDY: So usability is high. Sorry, I didn't really we jumped -- also high.

MS. QUINNONEZ: Voting is now open for the feasibility of measure 3032. Option 1, high; option 2, moderate; option 3, low; and option 4,

insufficient.

(Voting.)

Voting is now closed. For the feasibility of measure 3032, 63 document voted high; 26 percent voted moderate; 11 percent voted low; and 0 percent insufficient.

CO-CHAIR FLEISHER: Any comments?

MS. QUINNONEZ: Voting is now open for usability and use of measure 3032. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient information.

(Voting.)

All votes are in and voting is now closed. For the usability and use of measure 3032, 63 document voted high; 376 percent voted moderate; 0 percent voted low; and 0 percent insufficient information.

CO-CHAIR FLEISHER: Any comments before we vote on overall suitability? No, please vote.

MS. QUINNONEZ: Voting is now open for overall suitability for endorsement of measure

3032. Option 1 is yes. Option 2 is no.

CO-CHAIR FLEISHER: Cliff, you were clearly right. I misjudged the end. The patient is still asleep. So we actually beat our anticipated time and we're only off 7 minutes from being off an hour and 30 minutes.

(Voting.)

MS. QUINNONEZ: For the overall suitability for endorsement of measure 3032, 100 percent voted yes; 0 percent voted no.

CO-CHAIR FLEISHER: So we will take a break, a ten-minute break. And then we will get Bruce, you're ready, you're up next.

DR. BADHWAR: Thanks very much.

CO-CHAIR FLEISHER: Thank you. Thank you for staying. We really appreciate it.

(Whereupon, the above-entitled matter went off the record at 10:06 a.m. and resumed at 10:17 a.m.)

CO-CHAIR GUNNAR: Yes, we already had the two-minute warning, so coming back.

So Liz, you know you're on this solo?

Are we ready, Christy?

So the next measure is 0697, risk-adjusted case mix adjusted elderly surgery outcomes measure. American College of Surgeons and developers, you have three minutes to provide an overview.

DR. HALL: Thank you, Dr. Gunnar and Dr. Fletcher. Thank you, committee. I'm Bruce Hall. I'm a surgeon at Wash. U. in St. Louis and the Vice President in the BJC HealthCare System. I'm representing the American College of Surgeons today with Dr. Julia Berian next to me, who is one of our surgical scholars who has worked intently on this work.

We also have our senior scientist on the phone, Dr. Mark Cohen, and obviously, Dr. Ko is here, but is abstaining from the discussion with respect to this measure.

Dr. Fleisher and Dr. Gunner advised me to make a few brief remarks. We'll be happy to return to any category where there are more specific questions.

We're starting with the first of two measures in front of you for maintenance today, the elderly measure which has been in use since 2011. It is a risk-adjusted measure of outcomes, death or serious mortality. That's death, returned to OR unanticipated, and 12 other mortalities, rigorously defined.

And importantly, this is a measure of 30-day outcomes, strict 30-day time horizon which really differentiates us from many other quality measurement programs. This measure focuses on patients 65 and older and has a denominator of 2900 CPT codes. Notably, we have removed VTE from the prior specification of this measure which we can describe in more detail, but that was done because of the recently revealed bias in that that has been widely published, and we investigated, but did not add socio-demographic factors which I'll be happy to comment on.

The NSQIP in the private sector has been in place for 11 years or more and as you know in the VA the parallel program has been in

place probably twice that long. We've had five measures in front of the NQF and are just discussing two today.

We don't feel, under the category of importance, we don't feel there's major change to evidence. We do continue to demonstrate a variance in care. Almost 40 percent variance in care and we do continue to demonstrate ability to detect outliers.

In the category of scientific criteria, in terms of both reliability and validity, there are several aspects of how we specify our data fields and our program that yield high reliability and validity. That includes rigorous definitions, trained data collectors who are also examined, a community of data collectors continuously exchanging best practices.

We also audit nationally more than 12,000 data fields per year and have a national community which meets annually. We importantly provide feedback and revision from all national

users to our program in Chicago so that our data fields and variables are constantly scrutinized and revised over time and that includes some variables that have been in place for 11 plus years, as I mentioned.

Institutions also perform internal inter-rater reliability auditing and that is fed back to the college for continuous improvement of all of our data fields.

In terms of validity specifically, we have an outcome measure here which has inherent face validity as you have mentioned, but we also feel that we are representing an expert panel, so to speak, of more than 30,000 surgeons who have been using these data fields and this measure for years now. And particularly in the geriatric aspects of our program, we have a focused steering group and expert panel that provides sort of validity feedback, if you will, on all of the aspects of these data fields in this measure. And based on that, we've also shown widely across the program the programmatic improvement. And so

again, evidence of validity of various kinds.

The risk adjustment for this measure involves three main factors, the patient's ASA class CPT code linear risk and the patient's functional status. And so those represent both the patient's biology, the procedural endogenous risk, and then we turn those into obviously the reported hospital effect. It is a hierarchical adjustment scheme which does involve Bayesian smoothing which has more or less become standard across NQF measures nowadays.

And we are dealing with a focused population here. Obviously, 65 and older, largely Medicare covered and somewhat focused in terms of biology and presentation. As I mentioned, we investigated SDS and I've not included which I can describe in more detail and we have removed VTE.

In terms of feasibility, we have again almost 800 hospitals and 30,000 surgeons using this measure and like measures across the country now. So we know that it's feasible and useable.

And for whatever burden is attached to the way we specify data, I want to just emphasize that because we're using a strict 30-day time horizon which differentiates us from most every other program in the country, that there is a bit of a tradeoff there between burden and complexity and unique added value of our program.

We do also know the true cost of participation in our program. The subscription fee in our program now varies between $10,000 and $25,000 and you basically need some fraction of an FTE varying between one quarter and one. And if it's one, then you're going to pay that person probably $70,000. Then you are talking about a participation cost annually of something like $75,000 to $100,000. However, that covers 200 models across a variety of surgical specialties. So to say that that's the true cost of this measure in front of you would not be accurate. This measure in front of you represents probably less than a percent of everything that is obtained by participation in our program. So the

true cost is actually substantially smaller than what it might first appear to be.

With that, I'll pause and I'm happy to take further questions as the categories arise.

CO-CHAIR GUNNAR: Dr. Erekson.

MEMBER EREKSON: So I think the first thing to talk about, this is an already-approved measure coming up for renewal, so the first thing to talk about is evidence. One of the new pieces of evidence that came out since this measure was first approved was the joint statement from the American College of Surgeons and the American Geriatric Society about optimal preoperative care. And I think that only advances that there is good evidence that there are processes that we can do to perform -- to affect quality performance in this area.

The question from the NQF staffers is why wasn't this measure inclusive of patients younger than the age of 65, and then the other question by people who completed the survey before our meeting was why aren't we looking at

people over the age of -- higher age ranges.

So we're looking at patients who are 80 and over; is that going to get us more bang for our buck? I don't think there is the perfect age to set that cutoff at. I think there is good data to show that there is cognitive impact at the age of 60 that undergoing surgery could have some effects, and so I think the age of 65 is acceptable. So we could vote on new evidence, or we could leave it as is.

CO-CHAIR GUNNAR: So what's your recommendation? Should we vote on evidence, Dr. Erekson?

MEMBER EREKSON: I would say no.

CO-CHAIR GUNNAR: Okay, Dr. Dutton.

MEMBER DUTTON: Just a question for any of the smart people in the room. We tossed out DVT. I get that. But we also tossed out PE with that. Is PE better defined? It's obviously a very serious outcome for the patient. Should that have been kept in?

DR. HALL: We can comment. The PE

side of that equation is much more rare than the DVT side of that equation, so as a rare event it has an even smaller impact on our assessments. We do find that combined, there was an impact on assessments, but we feel that it's a biased impact. That was our decision for this program.

MEMBER BILIMORIA: The literature also shows that the same surveillance bias is visit for PE when examined separately, especially with new CT scanners picking up smaller and smaller sub-clinical PEs. So the same impact is seen as DVT.

CO-CHAIR GUNNAR: Moving along, does anyone feel that we need to vote on evidence? Hearing none, we'll go to gap.

MEMBER EREKSON: So currently, in the measure description, there are 460 hospitals that participate in the ACS NSQIP which is not a full penetrance across the country. But you have a good spread of risk adjusted O/E ratios between 0.59 and 1.69.

In the body of the measure, the

measure developers cite that there are 49 low outlying hospitals and 34 high-outlier hospitals that are statistically significant, representing a gap.

CO-CHAIR GUNNAR: Any additional comments? We'll vote on gap.

MS. QUINNONEZ: We are now voting on measure 0697. Voting is now open for performance gap. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in. Voting is now closed. For the performance gap of measure 0697 45 percent voted high; 55 percent voted moderate; 0 percent low; and 0 percent insufficient.

CO-CHAIR GUNNAR: Next for reliability.

MEMBER EREKSON: So for reliability testing, I think we've already been presented data showing the data element reliability of the ACS NSQIP. What the authors and the developer

did in addition is they tested the reliability of their modeling programs, and they actually published this in the peer-reviewed literature and included that article in this measure submission.

For this measure, the sample size that a hospital would need to achieve to have reliability at a threshold of 0.4 which is what they propose as moderate reliability is 180 cases, and they report that 85 percent of their hospitals meet this threshold.

CO-CHAIR GUNNAR: Any other comments? We can vote.

MS. QUINNONEZ: Okay, voting is now open for reliability of measure 0697. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: Voting is now closed. For reliability of measure 0697, 53 percent voted high; 47 percent voted moderate; 0 percent low; and 0 percent insufficient.

CO-CHAIR GUNNAR: Validity, please.

MEMBER EREKSON: So I think this is the appropriate place to talk about what the NSQIP does monitor and doesn't monitor in terms of outcomes. And when you look at outcomes in the geriatric population, some of the things that are very meaningful cannot be captured in the NSQIP. That would include post-operative delirium that has to be caught, and it's not very reliable to find post-operative delirium in chart reviews. That also includes falls outside of the hospital, which are also not incorporated in the NSQIP.

What the NSQIP does very well is they are incorporating functional status, and a lot of the things that others, not even just the NSQIP, have shown had been very predictive in these -- in geriatric surgery.

So the first thing is that while we don't have everything that we would like to have in a geriatric model of what surgery is, we have a lot. And then when you go on to the validity of the actual models, we have a c-statistic of

.75 to .77. And there is good data provided on including the SE status, not including the SE status and inclusion of the venous --

CO-CHAIR GUNNAR: Collette?

MEMBER PITZEN: Yes, I'm wondering if you could comment a little bit about the validity of the data that's collected in NSQIP versus perhaps against the medical record?

DR. HALL: I will take that question and I would like to just let Dr. Berian comment on the last remark first if that's okay.

DR. BERIAN: So I did want to comment for the record that we have a Geriatric Surgery Pilot Project that includes 26 different hospitals which are attempting to collect some of those more kind of patient-centered variables including things like delirium, mobility, functional status later down the road after an operation.

DR. HALL: In terms of the validity of the fields that we do capture, again, we have fields that have largely been in place for ten-

plus years and have been used by tens of thousands of surgeons demonstrating that improvement can occur using these fields, so I think in terms of what you often see for validity assessment is an expert panel. You may have eight experts and six of them say yes, it looks valid to me.

In fact, I think in this case, you've got 30,000 surgeons who have said over the past decade these are valid to us because we're participating, we're using them, and we're driving improvement with them. So I would say whether you want to call that only isolated face validity, or however you want to consider that, I think it's good evidence of practical validity in this case.

MEMBER PITZEN: Right. To clarify, I was talking more about the data element validity. So if you were looking for complications of wound disruption or infection, deep incisional surgical site, do you have some prior testing where you are looking at a sample of what's submitted into

the registry versus what's reflected in the patient record?

DR. HALL: Yes, I'm sorry. So in that case, we would consider that a form of reproducibility at the data element level, and as I said, we audit programs annually across the country. We shoot for between 5 and 10 percent of all programs being audited annually which means something like 10,000 to 15,000 fields get audited. And across variables, our reproducibility or our inter-rater reliability is uniformly 97 to 98 percent.

MEMBER PITZEN: That's fabulous. Probably for a future application, you could include that, and I wouldn't have any questions.

DR. HALL: Thank you. I think it's stated in one way or another in the material, but we can clarify.

MEMBER PITZEN: Okay. Thank you.

DR. HALL: Thank you.

CO-CHAIR GUNNAR: Lee.

CO-CHAIR FLEISHER: I may have missed

the reliability, but after the 30-day mortality in CABG I got lots of calls from The New York Times, ended up in The Times, related to the unintended consequence of cutting off the 30 days and therefore potentially individuals at day 31 being put on palliative care. And I don't know if that's validity or reliability. But have you looked at whether or not your 30-day outcomes are having unintended consequences?

DR. HALL: Well, that's a very insightful and also complex question. What we can tell you is that we have looked at the decay function, the time decay function of different outcomes over time. We know that no outcome is 100 percent captured at 30 days. But that the rate of capture decays dramatically over time with slightly different curves for different outcomes.

And what we have decided over the years is that the 30-day cutoff is a balance of capturing enough of those signals to generate good quality improvement versus the ongoing

burden of following patients longer and longer in outlying settings. So that's part one.

Part two is --

CO-CHAIR FLEISHER: I'm talking about inpatient, in other words, dying at day 31 in the hospital.

DR. HALL: So yes, I think part two is that we have not done a dedicated study on whether patients are dying at day 31, but there is such a study that's being -- that is coming out in the literature now. It was not done by our group. It was done by another group. So there is a very good study appearing in the literature now that indicates that there is no such bias, but that was not done by us.

CO-CHAIR GUNNAR: Actually, I published that study.

CO-CHAIR FLEISHER: Okay, good.

CO-CHAIR GUNNAR: In January.

CO-CHAIR FLEISHER: Can you send that?

CO-CHAIR GUNNAR: I can send it to you. There is no bias.

DR. HALL: I didn't want to put you on the spot.

CO-CHAIR FLEISHER: Where was it published?

CO-CHAIR GUNNAR: JAMA Surgery.

CO-CHAIR FLEISHER: Thank you. When they call us after this measure is approved --

DR. HALL: I didn't want to appear to be sucking up to Dr. Gunnar.

(Laughter.)

CO-CHAIR GUNNAR: Any other -- Karl.

MEMBER EREKSON: I just had one other comment for the developers. I didn't see because this is a composite, so we lumped a bunch of outcomes together to get to the composite score. I didn't see in this measure submission what is driving the composite.

Have you looked at, if you take out let's say -- I know you looked at VTE, but did you look at UTI? Is UTI driving the composite because it's so much more common than some of these other rare events? And can you comment on

whether or not those should be weighted in terms of looking at the outcome or if they're not weighted?

DR. HALL: Thank you, yes. So this is what we refer to as an unweighted aggregate. We don't consider this a composite. A true composite would be where you actually model each outcome, and then you aggregate the results of those models. In fact, what we do is we create an aggregate outcome. It's unweighted. Every outcome in the aggregate is treated equally.

Over the years, based on experience with hundreds of models across NSQIP, we have removed a few things such as superficial infections because superficial infections would swamp other important events and drive the composite.

And so we've reached this combination of what we refer to as death or serious morbidity based on the signals that we've seen across hundreds of models over the last ten years. They're not weighted. Each is treated equally,

so in answer to that part of your question, it's an unweighted aggregate outcome. Does that answer the question? Okay, thanks.

CO-CHAIR GUNNAR: A.J.

MEMBER YATES: Yes, I would just second what was just said and others. Ideally, the NQF's endorsed measures are patient-centric, and so you would want to have some sense of what has import to the patient. And I dare to say that if you were to do a Delphi on this and count 7 to 10 as being dire outcomes to the patient, they probably wouldn't rank an asymptomatic urinary tract infection or an incidental pickup of pneumonia on a chest x-ray as a dire outcome.

And so it really comes down to a lot of the -- if you have a hospital that is very -- most of them are backing off now, but if you have a hospital that's adamant about its fevers of unknown origin, work up on post-update, too, they're going to have a lot more little pneumonias or over-called pneumonias and urinary tract infections. It all depends on the

diagnostic intensity. So you don't have the denominator of how many tests are being ordered from one place or another. That's already been addressed. It's not weighted. It's the way it is. It's how you have your data over ten years. It won't change probably. But I would just throw out to you that it would be good to weight that with patient preference or patient weighting of that.

The one question that you might want to answer or change would be, you exclude major trauma. Is the major trauma exclusion based on a trauma score, which you know, there's good scoring for that. Or is that an arbitrary, this is major, or that's minor? Because with trauma scores that are out there, it would be easy enough to make somebody, you know, above a 5 major, and below minor.

DR. HALL: Thank you, two great insights. First on the aspects of the aggregate that you mentioned, for instance, we don't capture asymptomatic "UTIs". We don't capture

what are not true pneumonias. Our specifications are fairly rigorous. And I think that goes part way towards smoothing out that effect that you mentioned of perhaps different importance for different fields to patients. But that's not unique to us.

We've just heard about combining death and other morbidities for different cardiac procedures as well. And weighting these outcomes by patient-graded severity is something that our profession as a whole is struggling with, and it does apply to us, but it does not uniquely apply to us. All of us in the profession are struggling with that right now.

On the second aspect of your question, which was?

MEMBER YATES: The fact that there's trauma scores that exist, so are you delineating major versus minor traumas by the score?

DR. HALL: We are not using trauma score to do that. We're using a diagnosis category to do that, so if the patient has a

diagnostic code of trauma, then they're excluded. There are some minor traumas which we still allow, such as an isolated fracture or same level fall, but otherwise, any diagnostic code that indicates major or multi-system trauma gets excluded. In fact, we don't even -- it's not excluded in terms of actually being captured and then excluded. It's just they're not even eligible for accrual.

Our patient population is a focused population. It does not include transplant surgery. It does not include major or multi-system traumas.

DR. HALL: And I would just say that even isolated fractures that come in, it depends on how they arrive there, and if they come in by chopper, and they were extracted, they may have an isolated fracture when it's all said and done. But in the meanwhile, they've got the three CT scans are already done, cooked, and have to billed for, and so they get coded out potentially as a trauma.

MEMBER YATES: Yes, I don't disagree that there are potential shortcomings in the way we code in our profession absolutely. But any diagnosis of major multi-system trauma would be excluded.

DR. HALL: Ms. Whitaker.

MEMBER WHITAKER: Thank you. In the numerator, are you including other -- if a patient leaves the hospital where they had the surgery and then is readmitted to another hospital for one of those factors? Or dies at home?

DR. HALL: We -- if we learn about a death that occurred at home within the 30-day time period, it would be included. Readmissions are not part of this measure if that was your question.

MEMBER WHITAKER: I understand, but admissions for one of the morbidities.

DR. HALL: Yes.

MEMBER WHITAKER: So surgery, then followed by pneumonia or something like that.

DR. HALL: Right. So if, in fact, the outcome in question happened outside of the original hospitalization, then we rely on them either coming back to the same hospital, or we rely on communicating directly with the patient and family to determine that that occurred, yes.

MEMBER WHITAKER: Thank you.

CO-CHAIR GUNNAR: So as an aside, we welcome Dr. Whitaker who's new to the committee. We'll all get to know her over the next couple of years.

So we're back to validity. Any other comments? Amy.

MEMBER MOYER: I just want to make sure I'm completely understanding the composite. So the measure is the number of observed adverse events. So if a patient has multiple things go wrong, sometimes they start down the cascade, are each of those separate events calculated into the composite?

DR. HALL: So each of those different separated events might be reported out under

different models, but in that case it would be one case could only count as one event in the overall model. But if a patient had pneumonia and an MI, and we were modeling both of those in separate models, we would report both. Overall, they can only count once in the overall.

CO-CHAIR GUNNAR: Shall we vote on validity?

MS. QUINNONEZ: Voting is now open for validity of measure 0697. Option 1, moderate; option 2, low; and option 3, insufficient.

(Voting.)

MS. QUINNONEZ: Voting is now closed. For the validity of measure 0697, 68 percent voted moderate; 32 percent voted low; and 0 percent voted insufficient.

CO-CHAIR GUNNAR: Moving on to feasibility.

MS. JOHNSON: Excuse me. If you don't mind, I'd like to just put in a little comment here. The staff had originally assessed this as an insufficient, mainly because the information

on testing was not included in the submission. So you -- I'm assuming you have voted based on what Dr. Hall mentioned in terms of their data element testing. So what we will ask them to do is go back. We will open up their forms so that they can put that information into the submission. We want everything to be very transparent so that we understand why you guys voted the way you did.

DR. HALL: And we did provide those responses.

MS. JOHNSON: Did you go ahead and add them into -- okay, yes, we need added into the forms. Thank you very much.

CO-CHAIR GUNNAR: Feasibility.

MEMBER EREKSON: I think we've had a lot of debate over the last day or two on claims data versus registry data, and I don't have a lot to add.

CO-CHAIR GUNNAR: Any other comments? Go ahead and vote.

MS. QUINNONEZ: Voting is now open for

feasibility of measure 0697. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in. Voting is now closed. For the feasibility of measure 0697, 21 percent voted high; 79 percent voted moderate; 0 percent voted low; and 0 percent voted insufficient.

CO-CHAIR GUNNAR: Moving on to use and usability.

MEMBER EREKSON: As far as I can tell from the submission, public reporting is an aspect of this but is not required, so there's currently 131 hospitals who publicly report out of the 460 hospitals in the program.

DR. HALL: Correct, all of our programs across the country are making use of this information. And then inside of NSQIP, as an institution, you can volunteer to publicly report on Hospital Compare. So as you said, a subpopulation of hospitals are already doing

that. But all the hospitals in the program are making use of this information already.

CO-CHAIR GUNNAR: Any -- Collette.

MEMBER PITZEN: Just a quick question. Are the participants in the database provided back what happened for these patients so they can drill down further on their data?

DR. HALL: Every case in this entire -- in our entire registry program, if you will, is transparent and visible to the institution. You can reopen every single case and dive back into all details.

MEMBER PITZEN: Right, so if I had someone hit on this composite, I could drill down at my own institution and see was it an MI; was it a death, cardiac arrest, et cetera.

DR. HALL: Absolutely, yes.

MEMBER PITZEN: Thank you.

DR. HALL: The public can't do that. That's not what you're asking, right? Okay.

CO-CHAIR GUNNAR: To -- yes, to clarify, it's the extractor which is what -- it's

whoever has the data resource, they can drill down.

MEMBER PITZEN: Yes, only comment to the measures that I have grouped outcomes together from a quality improvement standpoint to the providers that are providing that care, if they can understand their combined score better, it can provide directions for them to work on.

DR. HALL: We agree. We absolutely enable that.

CO-CHAIR GUNNAR: Any other comments on usability and use? Hearing none.

MS. QUINNONEZ: Voting is now open for usability and use of measure 0697. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient information.

(Voting.)

MS. QUINNONEZ: All votes are in. Voting is now closed. For usability and use of measure 0697, 57 percent voted high; 43 percent voted moderate; 0 percent low; and 0 percent insufficient information.

CO-CHAIR GUNNAR: Overall endorsement.

MS. QUINNONEZ: Voting is now open for overall suitability for endorsement of measure 0697. Option 1, yes; option 2, no.

(Voting.)

MS. QUINNONEZ: All votes are in. Voting is now closed. For the overall suitability for endorsement of measure 0697, 100 percent voted yes; 0 percent voted no.

CO-CHAIR GUNNAR: So we'll move on to the second ACS measure which is for review which is 0706, risk-adjusted colon surgery outcome measure. This is a maintenance measure. Dr. Hall?

DR. HALL: I can be even more brief. This measure is very parallel to what we just discussed. This is a maintenance measure that's been in use since 2011. It is again a risk-adjusted measurement of outcomes; the same death or serious morbidity aggregate outcome at 30 days. It's defined by a limited set of colon CPT codes. Once again, we removed VTE from the prior

specification, and we investigated, but did not add SDS factors.

My other comments are very parallel to the prior, so I will just -- I'll stop there.

CO-CHAIR GUNNAR: Discussants are Drs. Temple and Siperstein.

Clarissa.

MEMBER TEMPLE: I don't think we need to review the evidence. We can move on.

CO-CHAIR GUNNAR: Yes, okay. Gap. Anyone else want to vote on the evidence? Okay.

MEMBER TEMPLE: So there's clearly differences in colon surgery outcomes. The authors report differences in ethnicity, as well as in socio-economic status and that there are discrepancies in the country. They did not use -- Bruce, maybe you can clarify, but I didn't actually see changes in this score between 2011 and now to sort of demonstrate an improvement and/or that disparities exist with this O/E ratio.

DR. HALL: Right. Thank you. So the odds ratio we see for this measure in the last

reporting period varies between about .66 and .14 at the 10/90 level. That's about a 40 percent up/down, if you will. And within this measure, we still do discriminate outliers even at the highest statistical level, even at the 95 percent confidence level. So if you convert that into kind of a standardized rate which everyone knows is just a construct and is not reality, but that would reflect a standardized rate across the country, that would vary from a 5 percent complication rate to a 30-plus percent complication rate. So that's the variance in performance that we're talking about.

MEMBER TEMPLE: That's very, very helpful. I just didn't see that. And so I think that's clear, there's a very high performance gap.

CO-CHAIR GUNNAR: Dr. Siperstein, anything else?

MEMBER SIPERSTEIN: I agree, there's a high gap with this measure.

CO-CHAIR GUNNAR: Any other comments? Go ahead and vote.

MS. QUINNONEZ: We are now voting on measure 0706 for performance gap. Voting is now open. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in. Voting is now closed. For performance gap of measure 0706, 94 percent voted high; 6 percent voted moderate; 0 percent for low; and 0 percent voted insufficient.

CO-CHAIR GUNNAR: All right, moving on to reliability. Dr. Temple.

MEMBER TEMPLE: So again, the numerator and denominator are clear. The only piece that perhaps you'd want to comment on is that the reliability of .04 is present when there are 99 colon cases -- 91 or 99 cases reported in a given year in NSQIP. So given the way the data is captured, sometimes less than 99 cases would be collected, unless you are doing a disease-specific data entry. So do you want to comment on that?

DR. HALL: Sure. That's a very

insightful comment. In an implementation, obviously, we would strive to have the implementation specified that X number of cases would need to be accrued. As Dr. Temple indicated, our number looks to be about 199.

It is still possible that some hospitals in the country might not do that many colectomies in a year. We've identified from the AHA database that that probably is the case, but that is a larger policy problem than the problem of internally in this metric. If some hospitals are not doing enough cases to be evaluated on quality, then that needs to be addressed in a different way.

MEMBER TEMPLE: So when you report out the O/E ratios for the various institutions, do you actually report out the confidence intervals as well?

DR. HALL: We do.

MEMBER TEMPLE: Okay.

DR. HALL: Yes.

CO-CHAIR GUNNAR: Any other comments on

reliability? Go ahead and vote.

MS. QUINNONEZ: Voting is now open for the reliability of measure 0706. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in. Voting is now closed. For the reliability of measure 0706, 52 percent voted high; 48 percent voted moderate; 0 percent voted low; and 0 percent voted insufficient.

CO-CHAIR GUNNAR: Regarding validity, the same applies as before, Karen?

MS. JOHNSON: Yes, we would ask you to -- we will open up your forms and ask you to insert the information that you mentioned earlier, if that's the same.

MEMBER TEMPLE: So my only comment is the same as well in that 10 UTIs is very different than 10 deaths, and the way this measure works is it's the same score for those two outcomes based on my understanding. And so I think that's the

standard of measurement today, but I'd like to see the standard of measurement with NSQIP different in five years.

The other piece with the validity is that the risk model is sort of not available. It's clearly proprietary, and so just to comment there's several publications with the risk model, but it's not in the public domain to look at.

DR. HALL: Thank you again for those comments. So we agree again, severity weighting of outcomes is a challenge for all of us, not unique to us, but we agree that our profession needs to make some progress on that.

Our risk elements in the model are provided. We didn't provide coefficients on them, but we do state that if this were implemented publicly that we would provide those specifications to the public if that's the intent of the measure.

Yes, Larry?

MEMBER MOSS: To the developers, patients under 18 years of age are excluded. And

I understand that there's a strong clinical rationale why you wouldn't lump them in this particular measure, but I think it's important for measure development and specifications that it be noted as an exclusion with a written rationale why. And I'd further suggest that in the committee's report, given the high morbidity of colon surgery in children, that we make a note that there's a gap there and an opportunity for measure development.

DR. HALL: Thank you for that comment. That's a NSQIP platform-wide cutoff. And we would certainly agree that that calls out for more development on the pediatric side.

MEMBER SIPERSTEIN: Larry, just to comment, there is a separate pediatric NSQIP program that addresses this.

MEMBER MOSS: No, I understand that Allan. I think in NSQIP that's well understood, but I would like to see the NQF not exclude children as standard operating procedure, but recognize when they're excluded. And if there's

a rationale, state the rationale, and if there's a need, state the need.

CO-CHAIR GUNNAR: Any other comments regarding validity? Dr. Siperstein? No. Shall we vote?

MS. QUINNONEZ: Voting is now open for validity of measure 0706. Option 1 is moderate; option 2 is low; option 3 is insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in, and voting is now closed. For the validity of measure 0706, 90 percent voted moderate; 10 percent voted low; 0 percent voted insufficient.

CO-CHAIR GUNNAR: Moving on to feasibility.

MEMBER TEMPLE: I think the comments from the previous discussion hold.

CO-CHAIR GUNNAR: Any other comments?

MEMBER MOYER: This is probably a comment that may have gone better with the last measure. Most of the business we do is in a state that has a significant proportion of critical

access hospitals, and when I hear the overall structure of this registry described, it really sounds like something that would be very hard for them to do. And while they're not doing cardiothoracic surgery, good, they probably are doing a lot of the -- a fair amount of procedures that may be captured by a database like this. I don't have a really good solution for that, but it feels like there's a chunk of hospitals where programs for which reporting those or tracking this kind of data isn't an option.

DR. HALL: Thank you. That's a very appropriate comment and I would say two things. First of all, we do have a population of critical access hospitals that participate in the program now, so we know that it is feasible, and we do reduce both the current cost and FTE requirement for them. In the future, however, in an implementation of this metric we would not require hospitals to be NSQIP participants. We would just guide them on the acquisition of these fields for this metric or the prior metric.

MEMBER MOYER: Terrific. Thank you.

CO-CHAIR GUNNAR: Any other comments regarding feasibility? Go ahead and vote.

MS. QUINNONEZ: Voting is now open for feasibility of measure 0706. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in, and voting is now closed. For the feasibility of measure 0706, 35 percent voted high; 65 percent voted moderate; 0 percent voted low; and 0 percent voted insufficient.

CO-CHAIR GUNNAR: Usability and use.

MEMBER TEMPLE: I think the comments are the same.

CO-CHAIR GUNNAR: Any other discussion? Go ahead and vote.

MS. QUINNONEZ: Voting is now open for usability and use of measure 0706. Option 1, high; option 2, moderate; option 3, low; option 4, insufficient information.

(Voting.)

MS. QUINNONEZ: All votes are in, and voting is now closed for usability and use of measure 0706. Fifty percent voted high; 50 percent moderate; 0 percent voted low; and 0 percent voted insufficient information.

CO-CHAIR GUNNAR: Voting for overall suitability.

MS. QUINNONEZ: Voting is now open for overall suitability for endorsement of measure 0706. Option 1, yes; option 2, no.

(Voting.)

MS. QUINNONEZ: All votes are in, and voting is now closed for the overall suitability for endorsement of measure 0706. 100 percent voted yes; 0 percent voted no.

DR. HALL: Thank you for your consideration.

CO-CHAIR GUNNAR: All right, next measure for discussion is 2998, infection rate in bicondylar tibial plateau. It is a new measure. Developers are present at the table. If you'll

take three minutes to provide an overview.

DR. AHN: Good morning. My name is Jaimo Ahn, and I'm an orthopedic trauma surgeon at the University of Pennsylvania. And I am here on behalf of the Orthopedic Trauma Association for a measure regarding the measurement of infection rates in bicondylar tibial plateau fractures.

I do want to emphasize that this has been a collaborative effort within one of the committees at the OTA. And Dr. Bill Obremskey, who has been one of the main shepherds of this, was not able to be here today because he's at the Metrics Consortium, which is a federally-funded meeting that he had to attend.

So with that said, this is really the OTA's foray into a national measure, so I wanted to give you a little bit more background because I think the measure itself is actually pretty simple and won't require a lot of debate or questions. I guess we'll see about that.

But in terms of the injury itself, many of you may not have been exposed to this since you

were in training. It's a bicondylar tibial plateau fracture, and very simply put, what that means is the lower end of the knee or the upper part of the tibia is shattered. And the bicondylar means that it's both the inside and the outside of the knee making this a pretty significant injury.

And as the committee was talking about measures potentially put forward to the NQF, the reason this came up was multifold. One is that it's a pretty devastating injury. Perhaps not as quite life altering as a hip fracture, but you have one of these fractures, and you're essentially non-weight bearing for three months or more. So it substantially changes your life. It also puts you at future risk of degenerative arthritis such that 20 percent or more will go on to need a knee replacement at a later time. So we thought that it was an important injury to think about.

The reason that this specific fracture pattern really came to our light is that even

though it's not terribly common, it's common enough that everyone at community hospitals will see it, and orthopedic surgeons at community hospitals will treat it, but they won't really understand their patterns because it's just rare enough that they're not going to see too many at their hospitals.

As the Level 1 trauma centers have been informally tracking this, not as part of a national initiative but just of our own accord, one of the things that we realized is that the infection rate is highly variable and really way too high. So if you were to sustain a standard, let's say an ankle fracture today and went to the operating room for a surgery, your infection rate would be maybe 1 percent, low single digits, in some places less than 1 percent. With these bicondylar tibial plateau fractures, we found that the reported rates of infection, even at highly specialized centers range somewhere from 8 to the mid-20s or higher. So we realized there was a big variance here in the discrepancy, and it's

something that we don't really know the answer to.

But the fact that the event rate is relatively high and an injury that's not too common, so it wouldn't be burdensome for the people initially reporting, we felt would make a good combination for a national measure such as this.

In terms of the actual outcome that we're trying to measure, infection, we would use the CDC criteria for deep infection. And when this involves implants, it actually goes up to a year, and so our -- the obligation would be for surgeons to track this for a year and report it, during that year period.

And also to give you just a little more background, the consequences of this infection is not just to go and wash it out. It's actually much more substantial because there's hardware involved and a fracture involved. This often means that you have to multiple stage debridements. Sometimes removal of the hardware and further reconstruction, so the burden isn't

just let's take care of the infection, you're on your way; the burden is also very high.

So we felt that because there were so many of these different elements incorporated: there's a joint; there's the bone; there's infection, that this measure as a starting point for the Orthopedic Trauma Association would also be good, not only in its simplicity, but the implications that we could learn a lot about soft tissues, how the joint is affected, the timing of surgery, and I can get more into the variables that we'd be interested in looking at in terms of further improving what we do as fracture surgeons as we go on.

So in terms of some of the nitty-gritty, the numerator/denominator is relatively simple. The numerator would be patients who sustain a deep surgical site infection according to CDC criteria. The numerator, we can discuss the appropriateness of excluding patients 18 and younger, but as for 18 and older that come in with a specific type of tibial plateau fracture, the

bicondylar tibial fracture, these are easily identified now using ICD-10 as well as CPT codes for the fracture treatment. Even in the old system, the ICD-9 would have identified it as well. And we feel that this is going to be -- the reliability and validity are going to be high.

The reliability we actually don't have a registry to look at currently, but when we did look at a larger study and applied some reliability criteria, we found that we could accurately, with both high sensitivity and specificity apply the correct criteria so that we would identify these injuries with a complication. And that it's a -- it has face validity. We understand already what the implications of infection are because that's been studied for these injuries for many, many years.

And the final thing that I think really drew us to this, before I entertain questions, is the fact that again because of the relatively high event rate and our ability to look at factors such as would a wound back potentially mitigate the

infection rate, because there was a small randomized control trial suggesting that even the way we dress the wound may affect the infection rate. Do we have to rethink antibiosis for these injuries because it's so high? Should we really stop at 24 hours of prophylaxis? Is the timing of surgery important? We know that it's important for femoral shaft fractures. Is that going to be important for these as well considering again the soft tissue envelope, the joint, and some of the implications of further treatment.

And with that, I will entertain questions.

CO-CHAIR GUNNAR: Discussants are Dr. Ko and Dr. Olsen. Dr. Ko.

DR. KO: Great. How much time do we have?

(Laughter.)

CO-CHAIR GUNNAR: As much as you'd like.

DR. KO: That's not a good sign.

CO-CHAIR GUNNAR: It's a new measure,

so evidence -- take your time with the evidence.

DR. KO: Okay, so this NQF measure 2998, the title of it is infection rate of bicondylar tibial plateau fractures, and as was just stated, this is an important clinical issue, and this is a new measure, as Dr. Gunnar said.

This is an outcome measure, and in that regard, there are likely processes that influence the outcome. There are, as was stated, there is a variation in the SSI rates. After this, ranging from 8 to 30 or even higher, 8 to 30 percent, and that's three peer-reviewed published studies. So I think that's the evidence. Is there anything further?

MEMBER DUTTON: I don't know where to fit this in, so if you want to bring it back later, that's fine. But I agree, the first thing we need to know is just the demographics, the incidence of the event across facilities. But risk adjustment is going to be absolutely critical in understanding these outcomes and particularly open/closed, transfused/not transfused, age of

patient, nutritional status.

DR. KO: I think we'll talk about the risk adjustment in the validity. At least I'll bring it up then, so maybe we can do it then.

MEMBER DUTTON: Fair enough.

DR. KO: I'm sorry, who is the other discussant?

CO-CHAIR GUNNAR: Dr. Keith Olsen.

DR. KO: Okay, sorry. Dr. Olsen, do you have anything?

CO-CHAIR GUNNAR: A.J.

MEMBER YATES: As the only orthopedic surgeon on the panel, I don't have any relationship with OTA, but I'd like to just put this in perspective as we discuss it. Orthopedics is one of the gaps that was identified by the panel last year, and the fact that one of the specialty societies has responded to that gap should be recognized.

The second thing is is that the -- we may not be doing the injury justice, but as a particular injury to gauge the quality of one

trauma center versus another or a set of orthopedic trauma surgeons, they've picked a disease that has a very high incidence and a very difficult problem to adjust to, so I think they should be applauded for that.

And the third thing to put it in perspective is is that since this is a brand new measure from a relatively young organization as compared to say the American College of Surgeons, it does have -- we do have the option of giving it a provisional approval so that they can collect the data and come back and seek a more permanent endorsement or a more standing endorsement.

So with those things said, I just want to make sure that the OTA has a chance just to put this in perspective, what's your penetrance into the trauma centers across the country? How many people are members? How many people are expected to respond?

And then the last question is -- and we will come to risk adjustment -- but you say you have no data for risk adjustment, but have you

established by consensus and literature review what risk factors you are going to collect in anticipation of establishing what may be high correlates of bad outcomes through regression analysis after you start to collect? So those three things: penetrance, likelihood of response, and have you established de novo through a consensus process which risk factors are you going to collect?

DR. KO: Can we just do the first two and then we'll talk about the risk factors in the validity?

MEMBER YATES: You can do that. I'm just saying that -- it's -- I just want to know what the homework is, just so you have a chance to present that up front.

DR. AHN: So in terms of penetrance of OTA members, essentially every Level 1 trauma center is going to be covered by members of the Orthopedic Trauma Association. In terms of centers that don't have trauma designation and how many members are OTA and what percentage would be

covered, I actually do not have that information at my fingertips.

MEMBER YATES: And participation in a registry?

DR. AHN: So participation in a registry is something that the OTA members have definitely expressed an interest in. In fact, this grew out of a task force within the -- under the Evidence Base Committee where there was a strong interest in developing a database.

And so that's something that the OTA has been working on, of having a registry. We do have a fledgling registry, but it's something that we wanted to nationalize and make something that would have more of an impact. And by working with NQF, for instance, we felt that we would increase the compliance and be able to actually have more people enter into the registry.

MEMBER YATES: And then I'll defer the risk question as requested by Dr. Ko and then these are Schatzker Vs and VIs or just Vs?

DR. AHN: These would be Schatzker Vs

and VIs. Anything that would be coded as being bicondylar.

MEMBER YATES: And for the audience, Schatzker VI means that it's not just split into both condyles, but it's also fractured below in the shaft of the tibia, so it's a disassociated bicondylar fracture.

CO-CHAIR FLEISHER: So just for the committee's purpose, there is no conditional endorsement any longer. It's either endorsed or not. And should we get that -- when we get there and should we get there, I mean it's either endorsed or not and if they need to come back either short term or long term, we'll address that. But the goal would be to focus on whether or not this deserves endorsement.

MEMBER YATES: Am I correct in saying that we used to have that?

CO-CHAIR FLEISHER: We did have a conditional endorsement, but we now only endorse or do not endorse.

MEMBER YATES: I think the handbook

needs to be updated then or maybe I missed something.

CO-CHAIR GUNNAR: So from an evidence point of view, A.J., is it your opinion that what's presented is -- that clearly links these infection outcomes to best practice and high quality processes, that if measured would drive that?

MEMBER YATES: The literature would support that, yes. The evidence supports that.

CO-CHAIR GUNNAR: That is the fundamental question that we are voting on in one minute or in 30 seconds. Any other discussion?

MEMBER GROVER: I'm going to recuse myself from the voting.

CO-CHAIR GUNNAR: Thank you. Very good.

MS. QUINNONEZ: We are now voting on measure 2998. Voting is now open for evidence. Option 1, yes; option 2, no.

(Voting.)

MS. QUINNONEZ: All votes are in and

voting is now closed. For the evidence of measure 2998, 89 percent voted yes, and 11 percent voted no.

CO-CHAIR GUNNAR: So carrying onto gap. Dr. Ko.

DR. KO: So the studies that were cited, there are three studies in the gap, in the spread in the SSI rates after this procedure. They range from 8 percent to 30 or above percent, so that's the gap.

MEMBER YATES: I would say that I'm struggling with that gap based on three retrospective single center studies with -- involved less than 500 patients total.

CO-CHAIR GUNNAR: Any other comments? Shall we vote?

(Off microphone comment.)

DR. AHN: The question was asked -- somebody was asking about the incidence of the injury. It is a relatively low occurrence event. So the best estimate is probably 1 in 100,000 per year, so -- and it constitutes, we believe, to be

approximately 1 percent of all bony fractures.

So depending on how you estimate the number of fractures in the U.S. a year, even incidental ones could be up to -- in the millions, up to ten million. So it does mean that there are thousands across the U.S., but any given center -- it's not like a hip fracture where you have many, many events, but we're measuring things for hip fracture like in-hospital mortality and one-year mortality. So I think that's very different than looking at something like an infection.

And as the OTA thought about this, that's why we migrated towards something that was common enough so that the OTA members would not feel burdened in terms of reporting it, but common enough and yet have a complication rate with an event rate that would be measurable and would be meaningful.

CO-CHAIR GUNNAR: Collette.

MEMBER PITZEN: I just wanted to make a couple comments. I applaud the Orthopedic Trauma Association again for tackling what is a

known problem, high infection rates. And just a comment that a registry doesn't have to be in place in order to collect data and have a measure.

CO-CHAIR GUNNAR: Dr. Yates.

MEMBER YATES: Yes, my one statement is is that the positive reported outcomes bias in orthopedics borders around .7.

In terms of people wanting to -- in terms of in the literature, people wanting to report what their infection rates are for a particular disease, if you're a high rate, you're not reporting it on a pretty regular basis. The fact that we have that big a spread in the literature, even though it's sparse, it's the literature and the problem is we don't have the registry and we don't have a way of collecting that through ICD-9 codes for the majority of trauma patients that lay outside of the say CMS databases.

DR. AHN: And if I may, one last thing, the study by Barei that has the lowest in the spread is from Harvard U Medical Center which in

the orthopedic trauma world is like the mecca, right? That's where they strive for perfection. They have -- they keep amazing records. That is where they're trying to do -- sort of sets the standard for the rest of the country. And when a center like that reports a rate of over 8 percent whereas a standard fracture at a standard community hospital will have an infection SSI rate of 1 percent or less, I think that also speaks to what a problem this is.

CO-CHAIR GUNNAR: Any other discussion? So we are voting on performance gap.

MS. QUINNONEZ: Voting is now open for performance gap of measure 2998. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For performance gap of measure 2998, 56 percent voted high; 39 percent voted moderate; 0 percent voted low; and 6 percent voted insufficient.

CO-CHAIR GUNNAR: Carrying forward to reliability.

DR. KO: So here are some -- I'll report the reliability issues and then maybe I have some questions for the developer.

So first of all, this is a measure that's 18 and over, and I'll save Dr. Moss the time, why is it not the children? And the inclusion is to have a fracture in an open reduction internal fixation and there's one-year follow up with three-year aggregate of the data. So that's the data.

The numerator, it was a little difficult to understand from the measure description because there seemed to be three options for numerators. So number one is a CPT code of a post-op wound, irrigation, and debridement. Number two was in the medical record an irrigation debridement and a confirmed positive culture. And number three was more like the CDC where there was a deep-space SSI in the deep soft tissue and it's purulent and all the clinical

things and constitutional symptoms of fever.

So it seemed like there were three options for wound infection for the numerator and so are they all three or is there just one that's doing that? The denominator is relatively easy, 18 and over which, Dr. Moss, and then the fracture.

The data source was a registry that is from the OTA website that is seemingly a QCDR registry, so that it's really made for the providers to participate in the Quality Payment Program with this, but it doesn't seem like this is in there already. We can talk about either here in reliability or feasibility, and that's the data source.

The testing of reliability is largely based on the two studies that have been quoted previously which are, as Dr. Olsen said, single institution retrospective studies of approximately 400 patients where it showed really good agreement sensitivity and specificity and positive predictive values. But it is in two retrospective

studies that are in the literature and not necessarily with these definitions or a registry.

So let me stop there and those are the issues of reliability.

DR. AHN: The age issue I think is a relatively simple one. If we went -- I don't think it's unreasonable to look at that population at some point, but because of the physeal closure and the timing of physeal closure depending on the age, I think that would complicate the issue more. So I think that would be a more advanced stage.

The other is some of these implants are not actually acutely indicated for patients under the age of 18, so that would also make it very hard to capture these patients.

In terms of the multi-numerator, I think that came out because there were multiple surgeons having a discussion about this. What it really comes down to is the CDC definition. And the CDC definition, the one complication here or not complication but modifier is the fact there is hardware involved, so the typical shorter term

definition doesn't apply anymore and now it applies to a year.

And of course, as orthopedic surgeons, we have to sort of discuss and say okay, we recognize that it's going to be up to a year. But the other is that when there's a deep surgical site infection that involves hardware, our standard of care as OTA members is to debride it.

And so I think that's where some of the confusion is, does it have to be debrided? Well, technically, it doesn't, but our standard of care right now is that we do debride it, so it should be captured by a debridement CPT code. But that -- as I look at this, that should not be a necessary component, but just because it's a standard of care.

The retrospective study, I'm going to perhaps use that to implore all of you to consider making this a more national prospective-type collection. That is the state of orthopedic trauma, especially in the U.S. Canadians have been better, and other countries with national

registries have been much better at collecting this data, but this is the state that we're in now and we are trying to make that move of trying to create more pre-thought out, pre-meditated prospective registries.

CO-CHAIR GUNNAR: Dr. Dutton.

MEMBER DUTTON: Well, I admit to having worked for many years at an inferior trauma center in Baltimore, but we did do a lot of washouts, routine washouts when there was associated soft tissue injury. So that wouldn't necessarily be infected -- or maybe would be, I'll ask the orthopods here -- but how would you discriminate that in the numerator, scheduled or planned washed out for an open wound versus one done specifically because there was a deep infection?

DR. AHN: It's actually -- it's nicely built into the system because I use it a lot for staged washouts. For an open fraction is a 1101X code. So that would be easily distinguishable and if the surgeon has any amount of detail in their operative record, it will give the Gustilo-

Anderson classification or some openness.

The fact that they're doing a debridement as well as a plan to return, but even based on the procedural code itself, the code that we use for an open fracture stage planned debridement is very different than if there's an infection and you're going in to treat that infection.

CO-CHAIR GUNNAR: Dr. Moss.

MEMBER MOSS: I appreciate your comments on the care of children, I just want to make two additional comments.

The growth rate issues do increase the complexity, but children do still have these operations and they are at risk for infection. There are very compelling data that the subset of children who receive surgical care in adult institutions in the United States is the trauma population. So the injured children that have these injuries are likely to be cared for by members of your organization, so I think it's particularly noteworthy to keep it under

consideration.

CO-CHAIR GUNNAR: So the other question I had, we just verified that it is 90 days for CDC follow-up. Would you modify your measure from one year to within year of release, at least recognizing that?

DR. AHN: So from what I saw that -- within one year if implant is in place and the infection appears to be related to the operation, so that's a modifier in the CDC definition. But regardless of that modification, our plan -- because the general outcome -- now not relates specifically for infection, but when we look at patient-reported outcomes, the OTA has basically decided that we would do a year and we thought that coincided well with the modification of implant in a year. And since we try striving for patient-oriented outcomes collection at a year that would be a good time frame.

CO-CHAIR GUNNAR: Dr. Ko, any other comments regarding reliability?

DR. KO: No.

CO-CHAIR GUNNAR: Dr. Olsen? All right. Ready to vote.

MS. QUINNONEZ: Voting is now open for reliability of measure 2998. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. The reliability of measure 2998, 5 percent voted high; 90 percent voted moderate; 5 percent voted low; and 0 percent voted insufficient.

CO-CHAIR GUNNAR: Moving on to validity.

DR. KO: So validity, this is an outcomes measure. The outcome, as we just heard, is going to be aligned with the CDC definition which I think most everyone uses for wound infection. So that is a valid -- and that's within the reliability, but the issue that Rick brought up and A.J. brought up is the risk adjustment.

This is a non-adjusted outcome measure and so they do cite some potential items that can be risk adjusted, but because this is not in place and they don't have those factors, they have not done the modeling to figure out if the modeling makes a difference or not, so that might be the reason why it's not risk adjusted and non-stratified so if the developer could address that.

The data source, as was mentioned, is an OTA certified QCDR which will be used and so the operative word there is will. This is, as I understand it, not in the registry yet, but probably can be put into the registry.

As far as SDS, there's a quote in there and I couldn't remember it, so I wrote it down: SDS patient factors, injury factors have not been consistently associated with differences in SSI rates so those were not included.

Again, the testing for validity are based on the two studies that have been previously discussed. And as far as threats to validity, again, no risk models, no risk adjustment. The

missing data issue, it's unknown because it's not in the registry. The meaningful differences is unknown, so reliability of distinction is unknown. And there is -- the measure has a lot of good, thoughtful planning in it, but nothing has been demonstrated to date.

CO-CHAIR GUNNAR: Dr. Yates.

MEMBER YATES: And again, it's clear that there's an intention to collect consensus agreed upon risk factors that are listed in the measure as submitted. And as noted, they can't report on the reliability or the c-statistic of those, but given the fact that it isn't collected yet. But at face value, the injury itself has the potential to -- because of its high rate of infection, the injury itself may, in fact, outweigh the risk factors. I mean it's a devastating injury and very hard to take care of in terms of avoiding complications.

You could almost do this study without any risk adjustment in the first year and move the needle in terms of the variabilities that are

brought out from having collected it for other people to learn from. And if certain centers perform or are doing something different than other centers, then there may be an opportunity for a best practice to evolve, even a non-risk adjusted.

I would just point out that the CDC criteria for hardware is a lot different than some other CDC criteria in that bacteria are harbored by prosthetics because of a number of reasons including the propensity in a low-oxygen environment for glycocalyxes and the like to form and bacteria being resistant to antibiotics. But on top of that, there are a number of these types of infections in joint replacement as well as in fractures that have hard to culture microbial etiologies. And there's a number of low grade skin bacteria. Propionibacteria is a classic that might take up to 15 days to culture if your lab even bothers to hold it that long.

So a draining sinus, purulence, and elevated CRP and sed rate are some of the reasons

why the CDC will allow that to count as an infection and would insist that it be counted as an infection so that those aren't missed just because they're culture negative.

CO-CHAIR GUNNAR: Collette.

MEMBER PITZEN: I'm just going back to the earlier discussion from Cliff about the specification of the numerator. Are we comfortable -- me not being an orthopedic person, I don't know if I from the discussion would understand okay, how would I count this deep wound infection? So I guess I would recommend a little bit of clarity around that.

DR. AHN: As I reread what we have proposed, I think there is a little bit of confusion. I think they're overlapping -- the Venn diagrams that overlap. I think as you start taking away some of the factors like one of the ones I stated which is an orthopedic surgeon as a standard of care, if somebody meets the CDC definition of a deep infection in the setting of hardware and metal, we'll take them to surgery.

So I think that's why that was put in because we were thinking of surgeons, but at the end of the day, the numerator will be a deep infection. Now whether a surgeon decides to go against what's considered a standard of care and not treat that appropriately, we can't really control. But yes, the criteria should and will be the CDC definition of deep infection.

MEMBER PITZEN: Perfect. Thank you.

CO-CHAIR GUNNAR: Cliff or Lee, you had a comment.

CO-CHAIR FLEISHER: Yes, risk adjustment still concerns me and I'm not concerned about you -- it's great that you're creating a registry and following these patients. And we're from the same institution, obviously. But what I'm concerned about is what NQF endorsement means as opposed to when a registry is ready for actually having an NQF endorsement.

So if it's not risk adjusted for the purposes for public reporting, would it be good for somebody to know what your rates are unrisk-

adjusted? Would that help anybody telling you whether that's a good hospital or a bad hospital if they get a trauma to be taken there?

Secondly, if you're talking about comparison between hospitals, would it really tell you enough without that risk adjustment? And the third -- and looking through the report, there's a whole bunch about negative pressure, vacuums, and other things. You actually talk about this treatment. I don't want NQF to be caught up in what our mission is versus any agenda with regard -- it was just interesting. I've never seen a company get so focused on acknowledging their own studies to support the measure as opposed to acknowledging how important it is to do something. It almost appeared like -- and I'll say it for the record, that if they endorse this and you don't have a good rate, non-risk adjusted, you should be using our devices.

So that's my question. Why do you think this is ready -- why do you think this -- if this went out to the public, would you be

comfortable with Penn Presbyterian's rate, be sitting out there unrisk-adjusted to drive the public to whether or not they should come to us or Rothman?

MEMBER MOYER: So related to that, if this is trauma surgery, is the public choosing? Probably, I'm guessing go where they're taken or is this a surgery where there's a delay?

CO-CHAIR FLEISHER: Well, that's why I actually said even comparing between hospitals so they can drive each other, is it good enough to risk adjust to drive improvement within a region?

DR. AHN: Our committee discussed this at length and I've had numerous conversations with Bill Obremskey about this. And we do realize that as it stands, not being able to risk stratify is a weakness, but we believe part of -- we also had a number of other measures that we were concerning.

At least part of the decision to move forward with something like this is because, for better or for worse, the OTA has not been able to

create registries like the American College of Surgeons, that because the event rate is high, we believe that in the early time period that we'll be able to collect the data to show the variation and variability to be able to risk stratify.

But without being able to collect that data, and without putting this on the national agenda, orthopedic surgeons are not -- they haven't been -- let's say the easiest -- my colleagues haven't been the easiest to work with in terms of driving those types of measures. So I think this was our way of saying we can make this an important thing for the OTA.

The OTA Board has certainly signed on to it and to make this an important agenda where by collecting data, then we can risk stratify. But I do agree with you that without that risk stratification, then it becomes very difficult to interpret and that's something that orthopedic traumatologists have been saying all the time. Without having that data, we say we take care of the most mangled and sick patients and fractures.

We can't be held accountable in the same way as another center that doesn't have this complexity.

So I think that is something that is foremost in our minds and that we hope by having that data we'll be able to create that risk stratification.

CO-CHAIR GUNNAR: We'll go Fred, Karl, Chris, and then back to Cliff.

MEMBER GROVER: I mean I think it's a reasonable strategy to collect data and have it not risk adjusted until you have enough data to do the model for risk adjustment. That's certainly how we started the STS and you get a few thousand patients or whatever it takes to develop your risk model and then you go with that.

I guess the only question I would have is now do you actually have people that are signed up to do the database yet or not?

DR. AHN: Yes, our incipient OTA database is something that people do contribute to on a periodic active basis.

MEMBER GROVER: How many members are

signed up?

DR. AHN: That I don't know.

MEMBER GROVER: One question I would have and I think this is a very reasonable thing. The question is to me when does it merit NQF approval? Do we want to see some evidence of what the participation is and the consistency and reliability of the data that you are collecting? And I think at some point we probably have got to touch on that.

CO-CHAIR GUNNAR: Karl.

MEMBER BILIMORIA: Yes, I guess I'd have -- sort of the end of Dr. Grover's comment is the same as mine. At what point are we ready for NQF endorsement? It just seems like it's premature. It's NQF supported, maybe. We all like the idea. It's a good measure, but should it get the stamp based on where it is now? And as the newcomer, I would just pose that as a question to the group.

CO-CHAIR GUNNAR: So just historically, there was an avenue for having that --

MS. MUNTHALI: Yes. We no longer have that. We have it for eMeasures that are not fully spec'd. They don't have testing information. You're going to hear from our colleague, Jason Goldwater, who will talk about trial use, I think, for this project. For other projects, claim-based or otherwise, no, there is no path. It would have to be fulled spec'd.

CO-CHAIR GUNNAR: Chris.

MEMBER SAIGAL: Just a follow up on Dr. Fleisher's comment. Is there some kind of secondary gain by a company that's being promoted in this measure? The developer didn't really address that. Is there some involvement of the company or -- I just want to clarify what he said.

DR. AHN: To tell you perfectly honestly, when Bill and I saw that, we thought it was a little surprising too. We didn't really understand why a -- why Smith & Nephew would come out in that fashion. I can tell you that I use Smith & Nephew products, but I have no relationship with them. And certainly between

Bill and myself, we have no relationship with Smith & Nephew wound care.

Maybe they saw it as an opportunity to have some spotlight, but the -- and because it is an infection and involves wounds, but yes, we do not have any particular leaning toward the utilization of a vac or any particular implant or device.

DR. KO: So I didn't want to make editorialized comments in the beginning to start things off, but if I'm the last speaker for validity, I will now. So -- oh, if I'm not the last speaker, should I wait? Yes, wait? Okay.

CO-CHAIR GUNNAR: College, then Amy, and then Cliff.

MEMBER PITZEN: This is Collette. I just want to share my own experience. We have two statewide orthopedic measures in Minnesota implemented across all of the practices using a direct data submission process from the practices, so not a registry per se. But through the last eight years that I've been working with NQF, I

have had the fortune of bringing forward a measure that was almost done.

For example, we have a spine surgery functional status measure and we were at the point where we had picked variables for risk adjustment, but we had not enough data yet to run that. And like an ad hoc process was evoked when we had that information, so I don't know if there's any avenue. If risk adjustment is the only thing that's holding us back from a validity standpoint, I guess I'm asking staff direction. Thank you.

MS. MUNTHALI: If you feel the measure is fully specified and there is this aspect of validity that you have some concerns with, we could possibly ask the developer to come back during the annual update process that I mentioned before. That would be a significant change to the measure. We would bring it back in front of the committee to evaluate, to see if they've met that condition.

But that would only mean that the rest of the criterion have passed and you feel that

they've met all of the conditions for endorsement.

CO-CHAIR GUNNAR: Amy.

MEMBER MOYER: From validity perspectives, it sounded like today individuals are submitting data to this registry kind of on a voluntary -- it sounds like a little informal basis. If you move to using a measure based on that data, I think you would need to be able to demonstrate that individuals who are being measured are necessarily submitting all of the data and submitting all of their cases to the registry, and can you talk a little bit about how you plan to account for that?

DR. AHN: I'm not sure there's a great external way of enforcing that. I think we could certainly do data biopsies and see how much of the data we feel have been submitted. But with many -- similar to many other measures and things that the OTA has tried to do, I think it's going to be very difficult to enforce that at an institution level.

CO-CHAIR GUNNAR: Any other comments

except for Dr. Ko? Cliff.

DR. KO: So I tried to as objectively talk about the validity and a big thing is the risk adjustment and it's not just the risk adjustment wasn't there and we just said it doesn't need it. We just don't know. And we don't know because there's no data and we don't know there's data because there hasn't been any data collected except for these prior two studies.

Looking at just the application overall, I think that nobody denies that this is an important clinical topic. It is. And the numbers would specify that it is. And SSI is a very important thing and so that's absolutely true. This is almost like a first draft of a term paper where you need to write seven drafts to have your final -- maybe not seven, maybe three. But I think these are important topics and the validity is important and it's going down the road, but we can't really decide on validity because we don't have enough data.

And this is how it's kind of -- how

validity and feasibility is kind of linked because we can't get validity, the data for validity, until we address the feasibility of doing it through a QCDR and all this stuff that was brought up. It's voluntary and do people even do the QCDR and it's self-reported and all that stuff about feasibility.

So I think that this a great try and it's a very worthy topic and that's why I think A.J. said give it this status of keep going because this is worthwhile. But it's probably not -- it doesn't have -- we can't really address the validity in a meaningful way because we don't have enough data.

CO-CHAIR GUNNAR: Any other comments? We will carry on.

MS. QUINNONEZ: Voting is now open for the validity of measure 2998. Option 1, moderate; option 2, low; option 3, insufficient.

(Voting.)

MS. QUINNONEZ: Voting is now closed. For the validity of measure 2998, 0 percent voted

moderate; 16 percent voted low; and 84 percent voted insufficient.

CO-CHAIR GUNNAR: So does that close? Okay. So the measure does not pass, but I think the message is to continue and collect the data and work towards risk adjustment. Any other comments anyone else would like to add?

CO-CHAIR FLEISHER: I'm just curious, Cliff, can the college at all provide some guidance?

DR. KO: Yes, I know that -- Fred and I were talking about STS and the college would be very happy to help because this is a very worthwhile topic and so absolutely.

DR. AHN: Thank you for the feedback.

MEMBER MOYER: I'd just like to say in addition to being clinically important, I was really impressed with the organization of the application and the materials and the presentation. I think you did a terrific job for a first time. Thank you.

CO-CHAIR WARREN: So we have a didactic

session next regarding eMeasures. Will someone introduce -- okay, great. Welcome.

MR. GOLDWATER: So good afternoon, everyone. My name is Jason Goldwater. I'm a senior director here at NQF and I have a multiple number of responsibilities --- oh hi, Melinda -- one of which is to oversee our eMeasure review and acceptance cycle before these measures get to this committee or any other committee.

I don't think it's a surprise to say that the dawn of eMeasures is upon us. The widespread implementation of electronic health records, of data registries that has been growing over the last several years has really led to a greater increase in the number of measures being submitted with electronic specifications.

I have been around this field long enough -- 20-plus years, which I rarely admit to people -- but long enough to know when CMS back in the days when it was called HCFA was really trying to move out of the chart abstraction into the electronic measure form. And did not succeed as

much as they had hoped to largely because there was such low EHR adoption.

Now we flash forward into 2016, and roughly 80 percent of hospitals and 75 percent of physician offices have electronic health records and that continues to rise. So clearly, there is a greater emphasis, not just from CMS, but from the quality improvement community to learn how to develop electronic measure specifications and to measure clinical quality electronically with the technology that is available to them.

So what I'm going to do in the next few minutes is not get into an elongated history of the electronic measurement world because I'm sure as much as all of you would appreciate hearing about that, recognizing that I am the speaker standing between you and lunch, I'm certainly not going to do that.

But what I do want to talk about is sort of how we look at eMeasures, and in particular, the eMeasures that you all will be looking at and considering this afternoon. So

next slide.

Reviewing electronic clinical quality measures, this particular project, surgery project, includes the evaluation of five eMeasures that are being considered for trial use. Now what do we mean by that?

When eMeasures are submitted to NQF for review and moving onto a committee, they can come in through one of four ways. The first way is a de novo measure, obviously a brand new measure that has been electronically specified.

Another one is what we would call a respecified measure. So a claims-based measure or a chart-abstracted measure that an organization, usually sponsored by CMS, is transitioning from chart abstraction into electronic measurement and mapping the specifications accordingly.

The third one is what we call legacy measure which is a measure that's already used in a national program. It's also chart abstracted and they are moving that as well into an electronic specification.

And the fourth one is what we are doing now. Given as I've already said that there has been significant rise in the amount of eMeasures that are being submitted because there is such rapid proliferation of electronic health record adoption in hospitals and physician offices, there is still an issue at times when measures are being created, brand new measures. And that is that the data that is available to test the measure is not always easily accessible because, as some of you may know that are familiar with this technology, if you've seen one EHR system, you have seen one EHR system.

Even if it's the same vendor, there are very different EHR implementations. And one of those issues is that the way data is structured or the way the data is captured in an EHR, or even in a registry, can vary. Some of it is unstructured elements. Some of it is structured elements. Some of it conforms to a national standard. Some of it does not.

And so when someone is creating a brand

new measure that they want to use to fill what is an obvious gap, particularly in surgery, they may find it difficult when they get to the testing part because they don't have the data necessary to test the measure.

So do we stop the process then and there and go well, that's too bad. We would like to really put the kibosh on innovation and not let you use your creativity to come up with something that is needed. Obviously, the answer to that is no. So we came up with the Trial Use Program.

The Trial Use Program is a path to endorsement for new, innovative, electronically specified measures that cannot at this time -- emphasis cannot at this time -- fully satisfy NQF testing requirements for endorsement, but they can be implemented in the real world. So they can be implemented in hospitals.

They can be implemented in physician offices. In this case, it would just be hospitals, I guess. So they can be implemented, but they don't or are not able to get enough data

for testing to adequately fulfill our criteria.

So we do not want to hinder that measure because that data is not available. So trial use is an ability to put the measure in the field to be used and in essence, they are collecting data while the measure is being used.

Approval for trial use is not -- and I will reemphasize is not, and Elisa will quiz you at the end of the day -- it is not NQF endorsement. It is approval to continue to test the measure. You are approving the measure to go into the Trial Use Program, not for NQF endorsement. It will not receive a number. It will not be put into a national program. It will not be considered endorsed. It will be considered a trial use measure which means you're giving approval for the measure to go into the field for further testing.

The developer can then choose the sites to put the measure in the field and they have a three-year window to bring back the testing for endorsement. So in other words, the measure

collects data while it's in the field. Once the developer feels they have enough data because the measure has been used, that it adequately satisfies the testing criteria for NQF. They pull the measure out of the Trial Use Program. They evaluate the testing results from the data that's been collected and then they resubmit the measure for all of you to consider for endorsement at that time.

Next slide. So approval for trial use. This committee will consider the full NQF criteria when reviewing these measures for approval, just like you have been doing for the last day and a half, but you'll only vote on selected criteria due to limited testing data. You'll be looking at evidence and performance gap, importance to measure and report, and those are both voting criteria for approval to trial use.

You'll vote on one portion of scientific acceptability that the measure specifications are consistent with evidence. This is what we call a must pass criteria. This is

only what you will consider for trial use. Feasibility and usability and use should also be considered to determine if a measure should receive approval for trial use. However, please keep in mind what you're going to be looking in terms of results is not necessarily real live testing data.

What they have done is used a simulated test data set to determine if the logic calculates correctly and the measure is produced in the right metric. And what they are using is the Bonnie Tool that is developed and maintained by MITRE.

Brief aside, I always get asked this question and this joke never fails. Everybody generally asks me what does Bonnie mean? I have absolutely no idea. It's not an acronym for anything. If it is, nobody has told me. As a former developer myself, we are very fond of naming things after our pets or our children. I would suspect that Bonnie is one of those two.

Bonnie is where you create a synthetic test deck of patients that would normally be used

in the measure itself and they test to the measure against that synthetic test to see if the logic is calculating correctly. Now that being said, it doesn't mean that they create a test deck of patients that are always going to pass. They have to be able to show you that the measure will use the inclusion and exclusion criteria correctly so that you know if the measure gets put into the field it will calculate as it should.

Next slide.

So there are five measures that are going to be considered for trial use. And these are the five that are in front of you. So I'm going to hang around for the first discussion in case there's any questions that need to be answered. But what I do want to emphasize are three things. One is you're not reviewing for endorsement. Again, you are not reviewing for endorsement. You are reviewing for acceptance into the program.

Number two, the criteria that you are looking at is somewhat more limited than what you

have been using to date. And number three, if the program, if the measure actually gets accepted into the program, it will be placed into the field where it will be tested, and once that data is collected they will pull out, analyze the testing results, and determine if the measure should come back before this committee to be considered for full endorsement if the measure passes and there is always the possibility that the measure may not.

With that in mind, do you have any questions?

CO-CHAIR FLEISHER: Larissa.

MEMBER TEMPLE: So I remember a couple of years ago, we looked at some measures that had gone through some sort of different process and we were -- at the time, we were told not to evaluate the measures for their importance in the gap and evidence. If these measures, today, if we look at the importance, the evidence, and the gaps, will we reevaluate when it comes back or would that be considered completed today?

MR. GOLDWATER: That would be completed, yes. What you're looking at when the measure comes back are the results of the test --

MEMBER TEMPLE: So just reliability and validity then?

MR. GOLDWATER: That's correct.

MEMBER TEMPLE: Thank you.

MEMBER SAIGAL: Just curious, that ortho measure, couldn't that fall under this kind of pilot testing? I mean why is this eMeasure only? Is that a way to basically facilitate the identification of measures before they're really fully vetted, but they get the approval or the endorsement of this group as being important and then kind of like --

MR. GOLDWATER: I think your point is well made. At this point, the program is only for eMeasures. And the reason for that is because of the great emphasis that CMS is putting on, transitioning out of chart abstraction into eMeasures themselves. That has been a directive they've given to us and that is something that

certainly the quality measurement community is trying to embrace. One of the difficulties, however, is the lack of available testing data which is why we've restricted this just to eMeasures only.

CO-CHAIR FLEISHER: So maybe you can, Jason, define what the program is which is I think the difference. In other words, trial use, a lot of things are. How is the program going to evaluate how eMeasures work which changes the equation?

MR. GOLDWATER: So the Trial Use Program, there's a difference between the Trial Use Program and what we had in the past which was time limited endorsement which is no longer.

Trial use in the eMeasure program is you're evaluating the measure initially on the criteria that I went over, importance to measure. You're looking at feasibility only through a very narrow lens of synthetic test results. But the program itself is allowing the measure to get put into the field because it is filling at least if

the evidence is presented strongly enough, a necessary gap in care in which a measure is clearly needed. The difficulty with putting this measure into endorsement is that they are unable to get enough testing data in the time that they had during the call for measures. So they decided then to go this route which then allows them to submit the measure that it can be evaluated on some criteria minus the testing. It goes into the field itself. Data is collected and essentially the measure is tested in a real-life setting. And then it is brought back out once enough data is collected to determine whether it can be moved on for further consideration for endorsement.

CO-CHAIR FLEISHER: So getting back to the problem of why the last one failed to extend your question was they didn't have data on risk adjustment was probably the biggest question.

MR. GOLDWATER: Right.

CO-CHAIR FLEISHER: If we find that the measure should be risk adjusted, let's make a supposition, would that be -- and it's an

eMeasure, and they don't have a model. Should it be approved or should it not be approved? Because there's two different testing, testing of the elements, obtaining it from an electronic record, versus having built the correct model to allow --

MR. GOLDWATER: It's possible if they want to resubmit for consideration for that program, but generally, when measures come in to us it's indicated by the developer that they want to be considered for trial use because of the testing limitations. So The Joint Commission submitted these five measures because in the process of development were unable to find enough testing data to sufficiently fill the criteria.

The measure that is before these was not asked to be considered for trial use. So we did not look at it in that terms. If they want to do that, that is a potential consideration.

CO-CHAIR FLEISHER: See --

MR. GOLDWATER: Was it an eMeasure? I thought it was an eMeasure.

CO-CHAIR FLEISHER: No, it was not --

MR. GOLDWATER: If it's not an eMeasure, then we don't look at it in terms of trial use.

CO-CHAIR FLEISHER: But the second question is if an eMeasure had a reason to be risk adjusted, and they had no model built for risk adjustment, should we approve that for trial use? In other words, they may not have the data yet, but suppose it was in a -- from a different data source. I mean that's --

MR. GOLDWATER: You could consider that for approval for trial use with the notation that the data that they collect has to then be -- they have to adequately risk adjust the measure once the data has been collected.

CO-CHAIR FLEISHER: I think that would concern me and perhaps others because that's not testing. That's a threat to validity that we would have concerns about. I would have concerns about. So I'll let others -- Collette, Amy, and then Fred. Do you have a comment you want to --

MEMBER GROVER: My question is is there

a reason that they can't just do this without coming to NQF first and then come back after they've done it showing that they're able to collect the data and all the feasibility that you're talking about?

MR. GOLDWATER: There's the possibility, but there's also the possibility that they may be contracted to develop these measures and as a result of that do not have enough available testing data to complete the process which is why they're opting for the Trial Use Program.

And I'm not aware of the situation. I haven't inquired about that. But -- yes.

MS. WATT: I'm Ann Watt from The Joint Commission. And this is our team you'll be hearing from. We are not -- it's a very good question. We've asked ourselves the same question, why are we doing this? The answer is because we have a great respect for the NQF process and this is new for all of us.

These are not contracted measures.

These are measures that The Joint Commission has chosen to develop because we feel that they are important measures. We are very interested in this process. We also feel that approval for trial use gives us an opportunity, some leverage with potential testing sites which nobody is beating down our door these days to say yes, we'd love to test your measures. And so we're hoping that this would provide that incentive, too.

CO-CHAIR FLEISHER: Collette.

MEMBER PITZEN: Can staff put up the slide of the criteria that we are going to be talking about today? And if gap is one of them, I have a hard time just making a decision about that with no data. So I just want to understand better.

CO-CHAIR FLEISHER: Can you actually describe any measures that have gone through this trial of how other committees have looked at this?

MR. GOLDWATER: So there are a few measures that have been passed and moved on to trial use. One of the most recent ones was a

hemolysis measure which I'm just blanking on the exact specifications. But there was a strong evidence base as to why there is a current gap, why that process is not being used and the outcomes it's leading to if the measure was not implemented.

Unfortunately, there was not a lot of data to collect around hemolysis that would provide enough to go through for endorsements, so the measure was brought up for potential inclusion in trial use and was passed because there was enough evidence indicating a clear measurement gap indicating that quality performance would be improved if the measure was implemented, a strong evidence base about its importance to measure and report as well.

There was a small scientific acceptability in that the developer created the specifications that were consistent with the evidence and actually had had those conversations with CDC prior to submitting the measure submission form and they did do a fairly

comprehensive synthetic test deck in Bonnie to look at feasibility. So yes, it was done. It has been done.

MEMBER MOYER: So to follow up on that, I think it would be useful to relook at performance gap after the measure has been tested because part of it is is there gap in practice. But part of what we're also looking at, I believe, as we're evaluating the measure is the measure's ability as specified to identify and shed light on that gap. So looking at it in the wild as it's going to be used I think would be helpful for us when we're actually endorsing it.

MR. GOLDWATER: Again, I do want to emphasize this, again, you're not considering endorsement. So this is not going to be an endorsed measure.

CO-CHAIR FLEISHER: No, I think what's been said is when it comes back.

MR. GOLDWATER: When it comes -- okay.

CO-CHAIR FLEISHER: Can that be defined as part of the CDP process? Because the concern

is, at least what I'm hearing from the committee is if this comes back we want to vote on certain other areas or we do not feel that the CDP process has been fulfilled. So can we make that as a requirement, at least if it comes back to this committee?

MS. MUNTHALI: A suggestion.

MR. GOLDWATER: A suggestion.

CO-CHAIR FLEISHER: Well, then we can vote on -- we still vote on up or down if they come back in for suitability for endorsement?

MS. MUNTHALI: Yes, we will do an overall --

CO-CHAIR FLEISHER: We will do an overall.

MS. MUNTHALI: An overall.

CO-CHAIR FLEISHER: So people still have the -- so what's been clarified is at the end, we can vote, even if we don't get to vote for gap, we've only made a suggestion. We still have the overall endorsement question.

MR. GOLDWATER: Right. We can still

endorse it. Not today, when it comes back from trial.

CO-CHAIR FLEISHER: Who is next? Liz, Barry, Cliff.

MEMBER EREKSON: So throughout the last day and a half, we've been talking about how we're supposed to separate how things that we endorse then get applied. And what I'm struggling to wrap my brain around is how does approval for trial use then get applied in the next three years as these developers are using their data? What are the ramifications of that? So I'm just trying to kind of put it in context because we know what NQF endorsement means and how that gets applied.

MR. GOLDWATER: I mean what you're looking at is -- the only thing when it comes to trial use, the only thing that is incomplete is the testing component of this because there's not enough data to adequately do the reliability and validity testing and they've made that case when the application came in.

The other information that is on the

measure submission form is articulated about why there is a gap, the evidence for the measure, its importance to measure and report, what the metric will produce, how the metric will improve quality.

So they don't have the testing data to go through the criteria that we've established. But the other criteria that is essential, they've already addressed to that extent. So if the measure is getting implemented, if you vote for approval for trail use, and the measure gets put into the field, the measure at that moment then is filling the gap that they've articulated and they are then collecting data to determine whether the measure is reliable and valid and feasible.

That being said, there's always the case that it may come back and it may not be. That is the possibility. We are not saying that every measure that goes through the Trial Use Program is going to come back clean. It's just not -- we can't possibly foresee that and neither can the developer. But as I think Ann was articulating, it provides a basis to collect data

that they otherwise would not be able to get. It potentially makes it easier for them to implement the measure in sites. And then it's possible to evaluate whether the metric is driving potentially systemic changes in quality that are needed that they've articulated that are not there.

CO-CHAIR FLEISHER: So I spoke to other staff. The way I would frame this is we have clearly articulated from several people that you would like to vote on additional areas when this comes back. Since we are not the ones who can approve that change in the current process, that will be taken into consideration by NQF and we'll get back. Today, we should vote on this process.

MR. GOLDWATER: Correct.

CO-CHAIR FLEISHER: But the concerns of the committee were clearly heard today. It just can't be answered today. Fair? So that's very clear. And we do vote again during that process, we do get to vote a final approval or non-approval.

Barry?

MEMBER MARKMAN: Yes, I think it's important that -- I mean this is the future. This is coming into play, so I commend it, the forum to -- in doing this. But it still has to meet the importance and we can't stipulate, but we can make suggestions that when they do come back in three years, like risk adjustment, that they should include that variable in the data set.

MR. GOLDWATER: Yes, you can make those suggestions. I wouldn't stop you from doing that.

CO-CHAIR FLEISHER: That will be clearly in the report to CSAC that the Board sees the concerns.

Yes, Barbee.

MEMBER WHITAKER: So we're not supposed to address usability and use, but some of the things that are questioned there including unintended consequences are not covered elsewhere. And I think those are important for determining the validity or the value of the measure overall. So can we include those in our discussion prior to that?

MR. GOLDWATER: Yes, you can include those in the discussion, yes.

CO-CHAIR FLEISHER: So perhaps that would be for when this comes back if we state those concerns, they can be in the report so that should the measure come back from The Joint Commission, they should address these concerns and that would be great to have in the report.

Larissa, then Collette.

MEMBER TEMPLE: So not to jump out of turn, but there are some sort of scientific issues with some of the measures that affect the numerator and denominator which would theoretically be within the discussion about reliability. And I'm wondering where that discussion will come up if we're not discussing the testing characteristics because you'd hate them to do something for three years and for us to say the numerator and denominator were sort of incorrect.

MR. GOLDWATER: Right, I understand. So that's sort of where the Bonnie testing to some

extent comes into play because if the only available test -- I mean what we ask the developers to do is to use Bonnie as a baseline. And if they are able to to the best of their extent get a test data extract that actually have allowed patients to use that as well.

MEMBER TEMPLE: So I'm actually talking about the inclusion and exclusion criteria more than the Bonnie testing. That's my question.

MR. GOLDWATER: Are you asking whether -- I'm not sure what you're asking.

MEMBER TEMPLE: So if there's discussion or concerns about the developers' definitions for the numerator and denominator and their choice of inclusion and exclusion, we usually talk about that in the reliability section of our discussions.

If we have concerns about the inclusion and exclusion criteria, where do we put that in the context of this trial?

MR. GOLDWATER: So they had to use Bonnie as the test data for the

inclusion/exclusion criteria. So they will discuss it at that point in time. If you've got an issue with the inclusion/exclusion or the numerator/denominator section, that's where it should be discussed. When they do the specifications --

MS. MUNTHALI: It's in the measure specs, so even though you're not talking about reliability and validity testing, you will talk about it in specifications.

CO-CHAIR FLEISHER: Collette again, and then Barbara.

MEMBER PITZEN: Did importance to measure get added back into the criteria for eMeasure review?

MR. GOLDWATER: Are you talking about after the Bonnie -- after the approval for trial use testing is completed to look at that?

MEMBER PITZEN: No, the high priority importance to measure as a criteria, is that no longer --

MR. GOLDWATER: You're going to look at

importance to measure.

CO-CHAIR FLEISHER: Barbara.

MEMBER LEVY: So I still have a little bit of problem with the unintended consequences. We could potentially look at some of that when we look at the importance and the evidence. We could bring it up in evidence, but I think it is important for us to look at that before we say go ahead and put these measures out there.

Similarly, when we look at feasibility and look at cost of some of these measures potentially, and it's not just the cost to collect the measure, but it's also the cost of the intervention. So I think there's some things that we do need to discuss that don't fit exactly into the evidence and importance, but that actually need to come up in the discussion before we send a measure out for trial use.

I don't think it's just the testing piece that's missing.

CO-CHAIR FLEISHER: So -- did you have a comment?

So how would you, either Lisa or Jason or Melinda, we'll clearly go through the traditional, but then what should we do before we vote?

MR. GOLDWATER: I don't think there's any issue with discussing unintended consequences if you feel that there are going to -- if that's potentially going to exist as a result of the implementation of this measure. Clearly, if the measure gets implemented in trial use, we don't -- there's nobody that wants an unintended consequence, but those do need to be articulated to determine whether or not that's significant enough to delay or to reject the approval into the program.

CO-CHAIR FLEISHER: So perhaps if we continue the current way we do it, but not vote in those areas --

MR. GOLDWATER: Right.

CO-CHAIR FLEISHER: So reliability, validity, and usability, we will actually structure our discussion that way, if that's okay

with everyone. But it's more to give The Joint Commission our concerns when they bring it back. There will not be a vote on -- is it those three areas we do not vote on?

MR. GOLDWATER: That's right, because you won't have enough testing data to do so.

CO-CHAIR FLEISHER: Great. Barbara.

MEMBER LEVY: So I agree that we don't have enough data to vote on those things, but it might influence my decision to vote up or down to release something for trial use.

CO-CHAIR FLEISHER: I didn't say we shouldn't have a full discussion and response. All I'm saying is we don't per NQF rules vote on them.

MR. GOLDWATER: I want to go to the last bullet point which is you do need to consider feasibility and usability of use. They are going to have granted synthetic data to indicate feasibility and whether you believe that can be extrapolated into the field once it's put into the program. So those are -- I mean if you look at

their Bonnie test results and think to yourself, well, this just doesn't seem feasible to me. And even with the way they structured this, this is just simply not going to have any impact because the data is not going to be available, the structure data is simply not there, and we don't think that the results of the Bonnie test could be extrapolated into a real-life setting, that's the discussion you need to have. I mean that's why they have to do that testing.

CO-CHAIR FLEISHER: Shall we try one? I think that's the best way. We have both Helen back and we have Jason here.

MR. GOLDWATER: No, I'm going, bye-bye.

CO-CHAIR FLEISHER: No, you're not. You had a lot of definitive comments. So if we can ask our colleagues from The Joint Commission to come up. One of the options is we all get the food and -- good idea?

Okay, we will get food and in ten minutes we will start.

(Whereupon, the above-entitled matter

went off the record at 12:21 p.m. and resumed at 12:34 p.m.)

CO-CHAIR FLEISHER: I essentially see all of the standing committee back. So we will go through these measures.

These are the new guidelines of how we're going to actually have the discussion. Importance to measure and report, we will vote on evidence of the importance to measure and report and performance gap. On the scientific acceptability, we will only vote on the measure specifications are consistent with the evidence. We can discuss the other issues in that context what the concerns of the committee are. Feasibility, usability and use, and then overall suitability.

So we'll start with, and we'll see how this goes, the preoperative 3016, preoperative anemia screening, and if you could introduce yourselves and give us a short --

MS. DOMZALSKI: Good afternoon, everyone. My name is Kathy Domzalski from The

Joint Commission. I'm the co-lead for the project. The other co-lead is Michelle Dardis, who is an informaticist. Also from The Joint Commission, we have Ann Watt and JohnMarc Alban who are project directors. And we're pleased to have with us Dr. Jonathan Waters who is the chairperson of our advisory panel.

Thank you for the opportunity to share with you these five Joint Commission patient blood management measures. Blood transfusion is the most commonly performed procedure in hospitals having increased 126 percent between 1997 and 2010. While it can be a life-saving procedure, it does pose hazards such as risk of infection, increased length of stay, decreased function at discharge, and other associated complications. It uses a resource that must be conserved for the most urgent need and it has been unfortunately assessed in one study to have a rate of inappropriateness as high as 100 percent of the time when the hemoglobin is over 8.0.

We have developed and assessed these

five evidence-based measures to assist hospitals in identifying their opportunities to improve blood management efforts. While there are those efforts currently underway in some U.S. hospitals, certainly it needs to spread to all U.S. hospitals. And we look forward to your approval for trial use.

MS. DARDIS: And an overview of all five measures, I briefly wanted to address the testing that was performed on these measures thus far.

Kathy did introduce me. My name is Michelle. I am a nurse informaticist working for The Joint Commission on eCQMs. And I was responsible for much of the testing effort for this project.

In addition to a clinical advisory panel for the measures, we also had a Technical Advisory Committee and that was made up of hospital clinical analysts, informaticists, and blood project leads from hospitals who could advise on feasibility of the measures. And we

also performed site visits with five hospitals where we watched through their blood work flows to assess how to structure the measures appropriately for data capture with the EHR. So we'll be sharing the results of those visits as well.

MS. DOMZALSKI: Measure 3016 concerns preoperative anemia screening. And it assesses the proportion of elective surgical patients who had a timely preoperative anemia assessment. The denominator is selected elective surgical patients and the numerator is the number of them who had an anemia assessment between 14 to 45 days prior to the said scheduled surgical procedure.

Hospitals currently, by and large, do anemia testing, but it is done so close to the procedure that any assessment of the cause of anemia or correction is virtually impossible to effect.

The clinical intent is to ensure an adequate preop time frame to assess and correct anemia since uncorrected preop anemia really is the biggest predictor of a perioperative

transfusion.

Thank you.

CO-CHAIR FLEISHER: Thank you. Okay. Amy, are you the lead discussant member.

MEMBER MOYER: So I think I was initially partially thrown by the title of the measure which was preoperative anemia screening which to me suggested what we were looking to address was a lack of screening of patients prior. But it almost sounds like from your description of it it's not that screening isn't happening, it's that it's not happening in a timely manner. And that wasn't clear to me until actually right now.

So in terms of the evidence, there was evidence supported that anemia screening is important to perform in certain procedures and certain populations. In looking through the code set, it seems consistent with the evidence supplied, which was at a high quality.

I don't know that there was evidence that the specific time frame given, the 14 to 45 days, is set forth in the evidence, but it sounds

like that is the practicality of resolving the anemia prior to the surgery and that's the reasoning behind that. Those were my thoughts on the science.

CO-CHAIR FLEISHER: Barbee, comments.

MEMBER WHITAKER: I don't really have anything to add to that.

CO-CHAIR FLEISHER: Rick.

MEMBER DUTTON: I think we all agree that unnecessary transfusion is bad and these are a series of sort of stacked measures to get at that problem one of which is you have to assess the patient in a timely-enough fashion that you can produce an intervention and so Amy, the time limit is picked to how long does it take for oral IN therapy, for example, to work in a patient that it's going to work in.

My concern with this is the unintended consequence. There's also very good evidence that there's a huge amount of unnecessary preoperative testing in anesthesia. Lee has written some of that literature. And I am concerned that the

simplest response to a measure like this would be well, let's just test everybody whether they need it or not because we can't tell.

So I asked the developers already this question. For me, this measure is going to very strongly depend on exactly what surgeries we're talking about which is going to be a moving target over time. A knee replacement got transfused five years ago. It doesn't today, for example.

And so having a SNOMED code for these are the operations we're including and I understand that's 2,000 very granular SNOMED codes. I get that, but we do need to hear what operations are we talking about here and how are you going to put that out to the public in this measure?

CO-CHAIR FLEISHER: So let's frame that in the context of evidence as opposed to specifications in that the evidence -- and I did write some of the evidence -- that there's only selected operations. But the way you, in some ways, the way you specified it it says all

elective, your denominators are all elective surgery.

MS. DOMZALSKI: It's all selected elective surgical patients. That list are value set because of the complexity of ICD-10 and SNOMED coding contains approximately 2,000 procedures. However, they are all major procedures. We're not talking about any old hernia repairs here. We're talking about abdominal aortic aneurysms, those procedures likely to require blood replacement.

CO-CHAIR FLEISHER: I have a question, but I will defer to the phone and then we'll go Barbee, we'll just go with this side. I'll amend that. On the phone?

OPERATOR: To make a comment, please press star 1.

CO-CHAIR FLEISHER: No, this is not open for comment. Is there anybody on the standing committee --

OPERATOR: There are no committee members on the phone at this time.

MEMBER WHITAKER: So I'm confused about

when to talk about the actual characteristics of the measure. Is that now or is that at --

CO-CHAIR FLEISHER: This is just about evidence. There is evidence supporting obtaining a --

MEMBER WHITAKER: -- a timely.

CO-CHAIR FLEISHER: A timely hemoglobin. Larissa.

MEMBER TEMPLE: So I guess what I'm having troubles with is the supposition that all patients preoperatively can be -- can go to the operating room with normal hemoglobins and particularly I think of the oncologic patients. I don't think any patient undergoing ovarian debulking will be anything but anemic. The patients with pelvic radiation, colon cancer surgeries, so I guess I'm having issues because I don't know of any evidence that supports improving preoperative hemoglobins in that kind of cohort. And certainly, we wouldn't want to be advocating Neupogen which is actually already black labeled for cancer patients. So that's why I'm having

real problems with the evidence.

MS. DOMZALSKI: The measure only requires an assessment to be made. It does not require the correction to be made.

CO-CHAIR FLEISHER: I'll follow that up so that you for Larissa's patients, should they delay surgery to be able to meet the 14- to 35-day window?

MS. DOMZALSKI: I don't see why that would need to occur.

CO-CHAIR FLEISHER: I will tell you our surgeons book things for within two days of seeing a patient.

MS. DOMZALSKI: That has been a problem across hospitals that we visited. The surgeons want to operate quite quickly and this would require a lengthier period of preoperative time in order for an assessment to occur.

MEMBER TEMPLE: But that then implies that I need to check the hemoglobin in these patients and if it's low, I need to do something about it before I take them to the operating room.

That's how I'm hearing this measure. And that's what I'm hearing the rationale behind it. And I completely support optimizing patients preoperatively, but there are some patients that you just cannot and it's a fairly big proportion of patients.

MS. DOMZALSKI: I think that rationale would apply to the next measure in which we look at what the actual hemoglobin was prior to surgery. But this measure only looks at the assessment, that it's made in time.

CO-CHAIR FLEISHER: A.J. and then Allan.

MEMBER MOYER: Can I just jump in really quickly to help inform this?

CO-CHAIR FLEISHER: Yes.

MEMBER MOYER: They did actually provide the full code set of the specifications. It took me a while to find it, so if you go on the SharePoint site under this measure, the last spreadsheet listed has the SNOMED and the ICD-10 codes and everything for these code sets that

they're mentioning in the specification. It does include joint replacement to the point that was made earlier.

MEMBER YATES: It does not?

MEMBER MOYER: It does.

MEMBER YATES: It does include. Right. On a facetious basis, I'm surprised your surgeons wait two days, but a non-facetious -- I'm just trying to tie the evidence to the measure. And in the evidence they refer to hemoglobin as the test and in the numerator in the measure is anemia testing which is a very broad term. I mean I'm not going to -- just to take it to hyperbole, I'm not going to run sickle cell anemia tests or look for macrocytic anemia or something like that with Vitamin B12, etcetera, and serum iron levels.

Does the literature imply that you're asking for an H&H sometime between 45 and 14 days?

MS. DOMZALSKI: Correct. That's correct.

MEMBER YATES: Then I would argue that as opposed to anemia screening, it would be much

better to say that hemoglobin screening would be much more specific and much easier to understand.

MS. DOMZALSKI: Thank you.

MS. DARDIS: And the actual specification and the numerator looks just for hemoglobin screening.

MEMBER YATES: I know, but the way it's phrased, anemia screening means a lot to -- it means something a lot different to a hematologist than it does to me.

MEMBER SIPERSTEIN: So my question is kind of understanding the rationale of putting together the numerator and the denominator. I mean I understand the value of doing less transfusions. I understand the value of doing more anemia screening, but if you're following this as a ratio over time, you may be in a system where you have not cut down your transfusion rate at all. You have simply quote gamed the system by doing a bunch more screening. And that really isn't -- you really haven't solved the problem yet, even though your ratio has improved. So that

really is kind of my main problem in that the number that you're going to follow really isn't focusing on the key problem that is reducing transfusions.

MS. DOMZALSKI: Well, according to what our findings were, the current rate of anemia detection is so low that it's not possible for that anemia to be corrected before surgery at all. So that by increasing the timely preoperative anemia assessment and correcting the anemia, you're therefore reducing the chances of a perioperative transfusion.

CO-CHAIR FLEISHER: Fred.

MEMBER GROVER: I'll pass.

CO-CHAIR FLEISHER: Pass. Barb.

MEMBER LEVY: Yes, I do think -- first of all, I have an issue with the timing. What if it's 36 days or 38 days? I had patients who were coming in from out of town, other places. You're going to have outpatient data. You have no idea when you go and survey a hospital what I have in my office from a month ago or two months ago. For

my patients in GYN surgery, I've got a problem because I've got an open spigot. And so anything I do -- the surgery is going to stop the outflow. Assessing her for anemia isn't really going to help me. I know she's anemic. It's a question of what am I going to stop it? What am I going to do to fix it and how can I optimize it?

So I have a problem with the specification for the timing. I have a problem in the sense that there's an awful lot of outpatient or office or distant evaluation being done and how are we going to capture that because why would I repeat it if it were done by somebody else in an outlying area and I'm being referred the patient? So there are a whole bunch of issues I think with this particular measure that have some difficulties that I'm having problems with.

MEMBER PITZEN: Hi, I have a technical question. I don't know where the right place to ask it, but in your eMeasure then, were you planning on accessing lab values from the practices when you were looking at that 45 to 14-

day window?

MS. DARDIS: Yes, and thank you for bringing that up to build on what Barbara raised. When we went out to the hospitals, it was really important to us to assess the feasibility of collecting that data from the practices because, of course that's the automatic next question, right, how likely are we to get that data?

We saw facilities where it was more likely to collect the data where they had integrated EHRs between the practice setting and the inpatient that was not everywhere, of course. There's different EHRs between practice and inpatient. You have referring providers who aren't in your system. We saw that many hospitals have data that could be captured in the time frame, but it comes to the hospital as a document rather than a structure and then coded data we can capture in an eCQM.

We did score the feasibility of collecting this data in this time frame and shared that in our submission. Overall, the hospitals we

visited which were demographically diverse, financially diverse, some were academic, some were not, felt that there were ways to address the lack of data, whether it would be abstraction, improving their interfaces, other ways of getting the data into the inpatient EHR and this data element was one of the key reasons we felt approval for trial use would give us the opportunity to further flesh out the issues with bringing in that practice data.

CO-CHAIR FLEISHER: Sal?

MEMBER SCALI: I just wanted to echo just the timing element. It is, for someone like me to, you know, in AAA as your index cases you sort of describe for aortic abdominal aneurysm and lower extremity bypass is a big part of most vascular surgeons' practices, for example.

So, you know, the timing -- that it doesn't make a lot of sense with the supposition that it will give you an opportunity to potentially correct because, frankly, most anemia, at least in the vascular patient populations, is

anemia of chronic disease.

It is true you can pick up iron deficiency anemia for these elected patients but then the unintended consequence, which I guess we'll get to in a moment, you get this number that you're supposed to react to.

So then you're doing a stool guaiac. Now you're going to have to get additional lab tests and a full iron study panel, probably an LFT panel.

So now you're going to do an additional work-up for the overwhelming majority of patients that are -- it's going to make little impact in terms of what the outcome to the surgery would be for elective AAA patients unless the patient is profoundly anemic.

So doing that screening at that time point, 14 days pre-op or more, has little impact and also operationalizing it as was already described patients who get found, say, for a AAA they get an ultrasound that's scheduled by their provider, perhaps for screening, and then that may

be in some, you know, out of window time period when they've had some routine lab work that was done with their annual checkup.

They get referred to a surgeon and then they get booked. And so you're not going to put the breaks on that. It's just not -- that behavior will not change and it doesn't make sense to change it because I have -- you know, the anemia prevalence in vascular patients, for example, is extremely high, like 30 percent or more depending on because the anemia chronic illness, the renal insufficiency patients that we see, et cetera.

And so I would say at least from a vascular surgeon's perspective it doesn't make any sense to screen -- have that designation of the timing.

I do agree for major elective surgery with risk of bleeding that having a hemoglobin pre-op is critical. But the timing piece for me just doesn't make a lot of sense. I don't see, at least in our population, that there's any

literature to support that.

CO-CHAIR FLEISHER: Thank you.

Are there other comments? I'm going to allow -- because Aryeh asked to speak too.

DR. WATERS: Aryeh does, yes. Of course.

CO-CHAIR FLEISHER: But sure, please.

DR. WATERS: I just wanted to address that 14 to 45 day window. The 14 days, as was previously mentioned, was a time that was chosen so that we would have time to fix it.

The 45 days was an interval that we chose in association with the length of time that a type and cross match was good for and then we also figured that something that was longer than that was probably not clinically relevant. So that was some of the rationale for those particular time limits.

I just wanted to mention a case that I took care of when I was at the Cleveland Clinic many years ago where we had a patient that came for a total hip replacement and we did do

preoperative anemia screening for the patient and she ended up having an occult anemia.

She also had a colorectal cancer that was probably needing to be addressed more than she needed her hip replaced. And so that was kind of the initiation of this measure as something that was important.

CO-CHAIR FLEISHER: Thank you. I want to be careful about anecdotes to be honest. What I would -- so we will be voting on evidence. The one question I would ask the Joint Commission, it was pointed out that you have a blood conservation center of excellence program because it mentions -- certification.

Can you mention how these measures fit into that? Because you started with the statement that this was to eventually go into other programs. But it appears from your documentation that there is something sooner in mind.

MS. DOMZALSKI: The Blood Management Certification program is a new program for the Joint Commission and a joint -- a number of other

certification programs such as in perinatal care or acute heart failure and CHF management. And it does denote a center of excellence, as you've indicated.

This program just became effective in the spring and currently, unlike other certification programs, there is no requirement in the blood program to have hospitals monitor and report on the results of measures related to their services.

However, should these measures be approved for trial use we would then have those measures available in the program and require hospitals to report on a certain number of them in order to maintain their certification.

CO-CHAIR FLEISHER: Thank you.

Any other comments about evidence? Can we vote?

MEMBER KO: May I just ask a quick question?

CO-CHAIR FLEISHER: Sure.

MEMBER KO: The last thing you

mentioned -- so if these do not pass you're not going to do that?

MS. DOMZALSKI: Pardon me?

MEMBER KO: If these measures don't pass you are not going to require it of the certified centers?

MS. DOMZALSKI: That determination has not been made yet.

MEMBER KO: But if they do pass you will require it?

MS. DOMZALSKI: Yes.

MEMBER KO: Got it. Thank you.

CO-CHAIR FLEISHER: Okay. Evidence?

MS. QUINNONEZ: We are now voting on Measure 3016. Voting is now open for evidence. Option one, high, option two, moderate, option three, low, and option four, insufficient.

All votes are in and voting is now closed. For the evidence of Measure 3016, zero percent voted high, 14 percent voted moderate, 48 percent voted low and 38 percent voted insufficient.

CO-CHAIR FLEISHER: Next steps. Aryeh, if you're on the -- if you can please -- a brief comment, that would be great. Thank you.

DR. SHANDER: So, you know, you said no anecdotes. But the discussion I'm hearing is mostly anecdotal, because the fact that there's no evidence I don't think is the correct way to put this.

There is sufficient evidence on non-cardiac patients, now looking at more than a million encounters, of patients where anemia is found to be an independent risk factor for, again, outcomes of surgical patients, whether they're vascular or not.

In the cardiac world this has been established already for quite some time that anemia is an independent risk factor in terms of both complications, as well as the major one is transfusion, which is, again, another independent risk factor.

To state that, you know, that this won't have any difference or that it's more work,

anemia is a condition. Some of us will say anemia is actually a disease, and we screen for other things which have less of a risk factor.

Anemia is an epidemic. There's more than 2 billion people across the globe that suffer from this condition and the majority of these are not being screened at all.

So the argument that we're screening for hemoglobin and not for anemia may be rational in terms of the actual testing that we're doing, but we are screening for low hemoglobin. And then trying to, again, those patients, we selected out, I think, in this particular measure, we're selecting patients where we think we can reduce the incidence of transfusion and therefore improve their outcome.

So we're looking at patients who have significant blood loss. And those patients, there is data, both from Europe and now from the U.S., that mitigating this risk, or adjusting their risk in terms of treating their anemia, which requires some time prior to surgery. There's no question

about that.

Rick's comment about oral iron, we know that oral iron, actually there's data now that's been published in Hematology, as in the journal, showing that oral iron in the first two to three weeks actually increases hepcidin level, and that blocks the absorption of oral iron.

So therefore, again, in these situations, these patients, just like chemotherapy and antibiotics, get sent to an infusion center and do get their iron replenished.

Now, whether it's cancer, whether it's renal failure, where there's just iron deficiency, the therapy itself, with iron and erythropoietin, has been demonstrated to, again, correct anemia in these patients.

The data that's completely -- not completely missing at this point, because there is some European data that's been published -- is does that improve their outcome? There's one thing that is unequivocal, and that is reduces another independent risk factor, which is

transfusion.

And therefore we come back and we're doing two things -- number one, we're looking at patients who are anemic and trying to rectify that as a risk. And, two, there has been significant data -- again, I think that both European, as well as now Australian, as well as the Pacific Rim, as well as U.S. -- that it reduces the other risk, which is transfusion.

So I just want to make the comment that there is sufficient data even though the vote went the different way. There are a lot of anecdotes in terms of what to do with patients who have cancer, and I will tell you that even those patients can benefit from their anemia to be treated. And there is now -- the EU is looking at it, and there's a new publication that is right now in review looking at the treatment of anemia in patients who are going for surgical for -- who have cancer and are going of surgery. That is, cancer surgery.

So I just want to make those comments.

So I think it's an important measure that we get this on the books, because everybody else around the world is starting to look at it in a different -- in a different eye, and we're still lagging in terms of implementation of any kind of anemia assessment in hospitalized patients in the U.S.

CO-CHAIR FLEISHER: Thank you, Aryeh.

DR. SHANDER: Thank you for the opportunity.

CO-CHAIR FLEISHER: Thank you very much. I will take a motion. Would anybody like -- Allan, do you want to comment first?

MEMBER SIPERSTEIN: Yes, I just want to make a comment back to the developers, and just make it very clear, from my perspective, you know, your enumerator is a very laudable goal, and your denominator independently is a very laudable goal. Expressing it as a ratio has no clinical meaning to me. So I think you should, you know, go back and rethink how you want to, you know, affect your improvement.

CO-CHAIR FLEISHER: Fred?

MEMBER GROVER: I think virtually everybody -- and you and Anastasia can correct me if I'm wrong -- that has a major operative procedure gets a preoperative hemoglobin. And a lot of this is the question of the timing, and to have everybody -- a lot of us practice in areas where patients come from 500 miles -- you can get these things checked. But that isn't practical. And patients like to -- a lot of time when a decision is made for a major procedure, for anxiety reasons they want to get it over with with the family.

So what we do now, if we have somebody that we're really concerned about, we pick that up and we only have to delay their operation, we aren't putting everybody else through that inconvenience of waiting a certain period of time or an extra clinic visit.

So, that's one of my issues. We are getting preoperative hemoglobin. It's more a question of timing and then breaking up the efficiency of what the patients want. Our

patients usually want to be treated rather quickly.

CO-CHAIR FLEISHER: So, does anyone want to reopen and revote on this after hearing the comments from Dr. Shander, who is an expert in blood transfusion, anesthesiologist at Englewood Hospital, or the comments here? If anybody makes a motion, we will reopen.

Hearing none, we will go forward.

MS. DOMZALSKI: Measure two, the preoperative hemoglobin level is Measure 3017. The measure is designed to allow transfusion or blood use review committees to identify patients undergoing elective surgery with suboptimal uncorrected hemoglobin levels that may have led to a peri-op transfusion.

It assesses via stratification pre-op hemoglobin levels immediately prior to the selected elective surgical procedure. Thank you.

CO-CHAIR FLEISHER: Thank you. Larissa? Or Barbee, you want to go first?

MEMBER WHITAKER: Okay. So this

measure is designed to identify patients who could have benefitted from pre-surgical treatment to enhance their iron stores and to reverse anemia.

Identified in the measure are the numbers of patients who are anemic, had hemoglobins lower than 12 grams per deciliter prior to elective surgery, of the elective surgical patients receiving a transfusion during or within five days after transfusion.

Obviously, unnecessary blood transfusions are undesirable and perioperative optimization of anemia is preferred. But the data is not clear on the true cut point of 12, and there's some other areas of performance gap, if we get to that.

Larissa, do you want to pick up from there?

MEMBER TEMPLE: I'll just make the comment that, you know, I look forward to seeing that paper that's under review published about showing how oncologic patients can be optimized with iron preoperatively.

But I think it's a fairly large cohort of patients who present with anemia that's either from chronic disease and/or from oncological cause that is quite challenging to improve in the period of time where the disease isn't progressing. And, clearly, there is certainly some concerns about anything but iron in those patients.

So I think that the issue, to me, is the evidence gap and demonstrating, again, that all patients can be optimized to a hemoglobin of 12 preoperatively. And I think it's a fairly big -- it's a substantial subgroup that cannot be optimized, and that's -- maybe if they wanted to rethink the new way or show us the new paper that's about to be published.

MS. DOMZALSKI: The aim of the measure is not to say that all patients should be optimized. The aim of the measure is to profile for blood use review committees and clinicians who can review those who have not been optimized to see if that was possible.

Over time, there should be a reduction

in the number of preoperative anemic patients. The idea is for hospitals to formulate educational interventions for practitioners who have not optimized a patient who is optimizable, so that the reduction in preoperative anemia will occur.

MEMBER TEMPLE: Will there be -- I would imagine, though, at some point these numbers will be compared between hospitals, and obviously the different case mixes, whether it's an orthopedic hospital where you would think optimization is possible versus other types of facilities where there's a more blended mix.

So I worry that -- initially, it's just a percentage, but I worry that that's not the ultimate goal if we're looking at quality, because we want to be not just intra-institutional but also between.

CO-CHAIR FLEISHER: Barbara?

MEMBER LEVY: So, help me understand why a blood utilization committee would need this measure when they have access to the chart, they have access to these data already? So how is this

helping the blood utilization committee?

MS. DOMZALSKI: You're saying that they've asked for the data already?

MEMBER LEVY: I'm saying, when I sit on a blood utilization committee I've got the chart and I'm looking clinically at the scenario to assess why did we transfuse this patient, what was going on with this patient?

So I'm trying to understand why I need a measure that holds everyone to a standard that perhaps a subset of patients actually need -- and I may not know that. So, just help me understand how the measure helps the blood utilization committee.

MS. DOMZALSKI: I guess most simply, not all blood transfusion committees operate the same, and while that may be the procedure in your facility that's certainly not the procedure in a number of facilities.

So we would like to present for these committees who are not looking at charts and who are not taking clinical conditions into

consideration, a profile, if you will, of what patients are going to surgery with what hemoglobins and they should be looking at those prior to going with suboptimal hemoglobins, to see, A, are these people optimizable; B, did they really have a preoperative assessment; and C, what educational interventions -- is it an individual, is it a department, is it a group of individuals, what educational interventions will help in raising those preoperative hemoglobin levels? Because some hospitals just look at the number of units they transfuse and call it a day or a month.

MEMBER LEVY: So, I appreciate that. I guess I'm still seeing an unintended consequence when you create a number that says 12, hemoglobin of 12, that that sets a -- you know, if a hospital is not sophisticated enough to look at who they're transfusing, then to set a number that says "hemoglobin of 12," my concern is that that's setting a benchmark that says that that is an optimal level. And I'm not sure we have the evidence to support that.

You know, chronic anemia is different than acute anemia. We have different thresholds for transfusion. The threshold for transfusion is as important as the preoperative hemoglobin, in my experience, to determine whether a transfusion occurs or doesn't.

CO-CHAIR FLEISHER: Okay. Other comments, because Dr. Shander -- we're going to consider him your developer, or part of the development team. Is that an accurate consideration?

DR. WATERS: Part of the team.

CO-CHAIR FLEISHER: So, Operator, if you can open it up, and Aryeh, if you will have a brief comment.

DR. SHANDER: Is it open? Is it open?

CO-CHAIR FLEISHER: It is.

OPERATOR: Your line is open.

DR. SHANDER: Oh, okay. Yes, it feels muzzled. But anyway, I just wanted to make a couple of points. In terms of the discussion that just went on for this particular measure, which,

you know, again, I really don't know what I can add to the discussions going on, other than, again, there was some mention about optimizing patients preoperatively and that some patients are just going to have anemia and that's all we could do about that.

Well, I think we need to relook at this because there is very good evidence that you don't have to have all of the amount of time that you're looking for to -- that is, all the 14 days to 49 days -- even a few days prior to, especially oncologic patients where you want to minimize transfusion, because, again, there has been risks associated with that in terms of that particular surgical population.

So even earlier intervention, if you can, and there now -- as I said already, there is data to suggest that even short intervals of therapy prior to surgery raise hemoglobin and reduce the chance of the patient being transfused.

The review on 3017 I don't know what else I can add to the discussion back and forth on

that. Thank you.

CO-CHAIR FLEISHER: Thank you.

MEMBER WHITAKER: I think another one of these measures gets to the point about trying to optimize them, you know, the optimization issue.

But I do think that this issue is actually better managed through guidelines and standards than measures, because, as Barbara said, if you're not in a sophisticated blood utilization committee and you go in and you say, oh, well, we've been transfusing too much at 9 or 10 and just apply a blanket requirement to change, I think you need to look at the clinical profile of each patient.

So guidelines and standards can describe and recommend how to implement a blood utilization review committee that would have a better impact, I believe.

MEMBER SIPERSTEIN: So, a question for the developer. Explain to me the clinical meaning of your result -- I mean, the numerator and the

denominator statement. So, explain to me how the data would be presented. So, in a report, then you would have four different ratios, one for each of the hemoglobin ranges, correct?

MS. DARDIS: So you would see an overall rate. So you would first see an overall rate, what is the numerator rate of patients who received a transfusion.

Then you would see the strata and the strata would break out which cases were in a range of eight to nine, nine to ten, and so on. And so this would allow a committee to say, okay, here is the patient populations that fell into this range and this range and let's dig into the specific cases.

MEMBER SIPERSTEIN: I mean, I understand going through the granular data but I don't understand the clinical significance of the ratio, particularly, you know, as the numerator is the number of patients and the denominator is that subset that's transfused. So this ratio is going to be greater than one.

MS. DARDIS: And, again, the intent of the measure was for the blood utilization or the blood management committee, not the ratio or rate. The value is really which patients fell into which strata and where do we want to focus our efforts based on what we know our challenges are?

MEMBER SIPERSTEIN: Right. Again, I understand that's a laudable goal, but the way the measure is written, the key number that's coming out, this ratio, like looking at mortality, I understand what a mortality rate is.

When I look at this number, it doesn't have any -- it doesn't have cleaning meaning or significance. I understand how the bits and pieces would be used, but as the measure is written, again, I'd like you to try to explain or justify that how -- why you're expressing the data that way.

CO-CHAIR FLEISHER: Rick.

MEMBER DUTTON: So I'm a big fan of internal quality improvement, understand how this can do. But NQF measures are intended for public

reporting and there's going to be a real strong reluctance for anybody to report something that makes them look bad when it's not their fault.

So if I'm reporting this measure, I transfused a hundred patients last month and 20 of them had a low hematocrit before the surgery. So an opportunity for improvement, perhaps. But that ratio is completely meaningless without an understanding of the risk adjustment of that: how many of them were oncologic patients, how many of them had renal failure, how many of them were just post-trauma, whatever.

And I'm not going to want to see that number publicly reported unless I am certain that that assessment is a fair comparison of me to other facilities.

Similarly, it's not useful for your internal benchmark unless it can be a comparison of apples to apples across multiple centers. So what are your plans for the risk adjustment of this model or the presentation of this data?

MS. DARDIS: So, currently, the measure

does exclude a number of populations, including traumatic injury, obstetric procedures and diagnoses, sickle cell disease and patients on ECMO, because those were identified as populations that had unique needs and that would obscure the numbers we received in the strata.

I don't know if you want to build on that, Dr. Waters.

DR. WATERS: Well, I think the measure was primarily developed to raise awareness of anemia as something that we need to manage.

In my particular facility, from a women's hospital where we do a large number of hysterectomies, we take care of a large OB population. And in those particular segments of the population you have widespread iron deficiency anemia, whereas if you're working in a different kind of facility you might see a different patient population with a different spread of hemoglobin values. So it gives you an opportunity or an awareness of where you stand relative to different hospitals.

CO-CHAIR FLEISHER: Barbee, do you still have comments?

MEMBER WHITAKER: The exclusions, it seemed that there -- sorry, I can't talk about this? Okay. Oh, he brought it up so -- or she brought it up. Yes. So I was just going to say that if you were going to include trauma you ought to include other emergency surgeries where there could be a lot of blood loss, for ruptured aneurysm or things like that.

CO-CHAIR FLEISHER: So, Allan, did you still have -- no? Collette.

MEMBER PITZEN: I just have a technical suggestion based on what Allan was saying, and being that it's an electronic measure that you're pulling from a large population, perhaps your denominator would be patients with selected surgical procedures and taking into the feedback that the surgeons have given. But the denominator being surgical procedures and the numerator being those that received transfusion. And then, of course, you can stratify by pre-op hemoglobin.

Just a suggestion.

CO-CHAIR FLEISHER: Thank you. Any other comments before we vote on evidence?

MS. QUINNONEZ: We are now voting on Measure 3017. Voting is now open for evidence. Option one, high, option two, moderate, option three, low and option four, insufficient.

All votes are in and voting is now closed. For the evidence of Measure 3017, zero percent voted high, 14 percent voted moderate, 57 percent voted low and 29 percent voted insufficient.

CO-CHAIR FLEISHER: Thank you. So we will not be going forward on this measure.

3019, Preoperative Blood Type Testing and Antibody Screening.

MS. DOMZALSKI: The intent of this measure is to assess the proportion of, again, selected elective surgical adult patients who had a type and cross match or a type and screen completed prior to the procedure -- completed prior to the procedure.

The denominator is adult elective surgicals. The numerator is the number who had a type and cross or type and screen anytime within the 45 days prior to the procedure to the start of the procedure, and that type and screen or type and cross needs to be completed. That's what we're measuring here.

There is some evidence that there is occasional non-completion of those procedures prior to performance of the procedure.

CO-CHAIR FLEISHER: Barbee, and then Barry.

MEMBER WHITAKER: Okay. So there is some evidence, although the evidence is fairly low, with the gap. It was one percent of the time that they weren't completed. And I question whether the -- okay, so the measure follows from clinical guidelines cited by the developer, that in order to effectively utilize resources, pre-transfusion testing should be completed. However, the desired outcome is that the patient get an appropriate unit of blood if transfusion is

required.

And I would like to ask whether you considered using a numerator of elective surgery patients receiving un-cross-matched blood, rather than the possible unintended consequence of having a lot of type and cross and type and screens that were never required.

CO-CHAIR FLEISHER: Do you want to make comments, Barry?

MEMBER MARKMAN: I get the eMeasures. I think they're great, as I said before, and I think this is your barrier. You know, in terms of conservation, the numerator, those who had a type and cross match, and the denominator are selected elective surgical procedures.

I did not see the list of surgical procedures, but from what I hear they're major surgical procedures. Okay.

It was one percent. However, there are other studies that say it's as high as 7 percent. You have the longest measure I've ever seen.

However, one of the comments that you

make in this that stood out in my mind is that -- is by the College of American Pathologists. It's 108 public and private participating institutions. And of the type and screens for these major elective procedures, 64 -- or 65 percent were collected prior to the day of surgery and then there were a number, 23 percent after the start of surgery. And of those samples that were sent off late or close to late, 79 percent were considered clinically significant positive in their antibody screens.

So here we have major surgery, the type and screen is collected either the day of the surgery or during the surgery, and I think, you know, that the evidence is strong that there's a quality issue with that, to go into major surgery and send off the type and screen. And that's from the American College of Pathologists. I think the evidence is there that this is a quality measure.

CO-CHAIR FLEISHER: Okay. Well, I know Aryeh wants to make comments, but anybody --

DR. SHANDER: No, I just want to

support the argument. That argument is absolutely the reason for this measure, and you don't want somebody who can have a type and cross done where they're matched for a unit of blood receive un-cross-matched blood. That's for elective cases or any cases where you're planning significant blood loss.

You want to make sure that the type and cross is performed, or type and screen is performed, and concluded prior to surgery, because if there are antibodies and you need more time, surgery should not start.

That's the spirit behind this particular, and I think that what you quoted from the American College of Pathologists is clearly the data that will support this being universal.

CO-CHAIR FLEISHER: Any comments? Barbee?

MEMBER WHITAKER: I'd just like to consider it as an outcome measure rather than a process measure, because ultimately the goal is to assess your patient carefully, determine whether

they will need a type and cross or a type and screen and then do it in the appropriate time.

So a little tweaking to both the numerator and the denominator would make this a much better measure.

MEMBER MARKMAN: In the 45 days. I mean, we'll go to suggestions, because that's what we were instructed to do. Forty-five days sometimes is hard to, you know -- Yes, 45 days.

CO-CHAIR FLEISHER: Other comments from around the room from the standing committee? Barbara?

MEMBER LEVY: Yes, I just think -- I really like reformatting this as an outcome measure because the unintended consequence is driving type and screen or type and cross match in a large proportion of patients who don't need it at all. And so we really want to capture those patients who clearly did need it and it was done inappropriately. That's really a good outcome measure.

But the unintended consequence of

hospitals deciding that they're going to have to go back to doing type and screen or type and cross match for a large proportion of patients who really don't need them, that's expensive. It utilizes resources that need to be used in some other way, and it's really, I think, a problem.

DR. WATERS: There is something called a MSBOS, or maximum surgical blood ordering schedule, that basically does what you suggest, which is to focus your resources on not unnecessarily cross-matching or screening patients. And I think the measure's surgical procedures that we selected fall within generally accepted procedures where there is greater than 25 percent of blood being needed.

We looked at this particular measure at the University of Pittsburgh and found that we had 11 patients in the course of one year that got to the operating room without appropriate preoperative testing. And these particular patients ended up having antibodies that made it difficult to get blood to them. And in two of the

11 patients, they dropped down below three grams per deciliter of hemoglobin before appropriate blood was available for them.

So it's the severity of this particular measure: what we're trying to prevent with this measure is fairly significant.

CO-CHAIR FLEISHER: A.J., you had a comment now?

MEMBER YATES: I just had a question. Good to see you. We're both from the same institution. The MSBOS is something that came about in the last two years, last year, thereabout.

DR. WATERS: That's actually not correct. It's been around since 1971.

MEMBER YATES: But the current one that we're -- the updated one --

DR. WATERS: Yes, we have a data-driven MSBOS now, and previously it was an opinion guided. But we and several other facilities in the United States have moved towards a data-driven MBOS.

MEMBER YATES: And our data driven one, I would actually argue that for a total knee replacement I don't even need a type and screen.

DR. WATERS: Yes, absolutely.

MEMBER YATES: And so my question is -- two questions. Where is the list of which surgeries? Which list are you using, MSBOS, the most data driven one, the most recent?

And the second question is, the literature talking about the gap in this -- I know we're not to gap yet -- but the literature on the gap is limited to around 2004 in terms of time. So is there any current literature or data showing a gap using the new MSBOS?

MS. DOMZALSKI: I'll address the last question first. Hospitals are not particularly fond of publicizing their errors and their difficulties and their failures. For example, their need to transfuse un-cross-matched blood because they didn't complete a procedure. So, when doing the literature review we found all the literature we possibly could.

As for the MSBOS, the list of procedures that are covered are covered in a value set here, and it resembles the selective elective surgical procedure list, except that it is minus certain maxillofacial procedures, certain gynecological procedures, and others that are not likely to require blood replacement.

MEMBER YATES: And just out of ignorance, which section is the value set listed? Is it in one of the additional sets?

MS. DARDIS: It's a part of the specifications. I'm not sure where NQF packaged it in the package you received.

MEMBER YATES: Okay. Well, I'll have to search for it. Thanks.

CO-CHAIR FLEISHER: Thank you. Rick?

MEMBER DUTTON: Amy found the list. Knee replacement is on it.

CO-CHAIR FLEISHER: Thank you. So --

MEMBER LEVY: So is LAVH, and the current data show an average blood loss of laparoscopically-assisted vaginal hysterectomy at

less than 100 cc's.

CO-CHAIR FLEISHER: So the issue is for evidence, unless anybody else has other comments, that we vote on evidence as currently constructed. So in the specs that we were given. And if we can call it.

MS. QUINNONEZ: We are now voting on Measure 2019. Voting is now open for evidence. Option number two is moderate. Option number three, low. Option number four, insufficient. Option number two, moderate. Option number three, low. Option number four, insufficient.

All votes are in and voting is now closed. For the evidence of Measure 3019, 24 percent voted moderate, 48 percent voted low and 29 percent voted insufficient.

CO-CHAIR FLEISHER: Okay. I was just clarified, given this didn't meet on evidence, although several people had suggestions for The Joint Commission on other ways to assess this, we will be moving forward to the next.

3020, Initial Transfusion Threshold.

MS. DOMZALSKI: Once again, this measure has been constructed for transfusion committees to have a sense and see how their blood transfusion administration is occurring. And as you are probably all familiar with, a restrictive transfusion strategy is supported rather than transfusing patients at a higher level of hemoglobin or hematocrit.

And so this measure gives those transfusion committees a profile of blood usage in their institution by showing them the hemoglobin values of each patient immediately prior to the first unit of a transfusion.

There are exclusions in the measure, such things as emergency room patients, et cetera. The denominator is all hospitalized adult patients who get a red blood cell transfusion, and we're looking at the first unit only.

The numerator is stratified by hemoglobin values in one gram per dL increments in a range from an aggregate of less than seven through an aggregate of 10 or greater.

CO-CHAIR FLEISHER: Okay. Barbee?

MEMBER WHITAKER: Okay. This measure is to monitor proportions of patients transfused at initial hemoglobin levels from less than seven to greater than ten. The evidence, while moderate, is sufficiently strong to introduce a program of monitoring with the intent of having more transfusions occur at the lower, more restrictive end of the spectrum than at the higher, liberal end.

The blood utilization committee or the transfusion committee will have a metric from which to investigate transfusion practice. I think there's always a question of what you do with the metric. We have evidence through a survey that we conducted at AABB that indicates that many hospitals, most hospitals, have policies for transfusing between seven and eight and eight and nine for different circumstances, but that their actual transfusion values don't always meet those policies to which they intend to adhere.

So this measure would allow measurement

of the adherence to that policy -- to their internal policies, but I would say that for an investigational purpose, a self-monitoring purpose.

CO-CHAIR FLEISHER: Great. And Lynn, you're the second. Any additional comments?

MEMBER REEDE: Just one small comment. As far as evidence from STS and also from the Society of Cardiovascular -- or Anesthesiologist guidelines, they recommended, and the developer noted this in their rationale, that also underlying cardiac disease post-operative status in clinical conditions might be considered because there is not good evidence to support keeping those patients at a seven hemoglobin.

So just that the exclusion did not include, like in other ones we've had contraindications, kind of a general decision process for the clinician.

CO-CHAIR FLEISHER: Thanks. Christopher?

MEMBER SAIGAL: Yes. I think that the

rationale behind using these measures for developing evidence was that they were using the EMR and there was specified fields and three types of fields and it was more complicated to develop evidence.

This sounds like it's really a research project to understand what the prevalence of transfusion rates are in the country or different levels.

So as a quality measure, I don't see how this would be implemented because it sounds like there's different standards locally. There's no agreement.

So I don't think this really fits the idea of a quality measure. It fits the idea of a research or survey project.

CO-CHAIR FLEISHER: Rick and then Liz.

MEMBER DUTTON: I've worked in a number of hospitals that routinely look at all transfusions to see what the hemoglobin level was when we gave the blood for purposes of quality improvement, and I've looked at a lot of

exceptions as to why that might have happened.

I am very concerned about the unintended consequence in this measure of delaying transfusion in a patient who needs it while you wait for a hemoglobin value. If there's a hole in the aorta, I don't care what the hemoglobin is. The patient should be transfused.

MEMBER EREKSON: So I think this is also --

DR. SHANDER: I think that's excluded already.

MEMBER EREKSON: In the exclusion criteria, one of the things that I notice is pregnancy is not an exclusion, although major trauma is. And when you're having a patient who is having a massive bleeding, hemorrhage postpartum, that I would equate to some of the trauma patients, and so that might be considered.

CO-CHAIR FLEISHER: So that --

MS. DOMZALSKI: I'm sorry. In the world of e-specifications it's very difficult to find in a chart the items that would exclude or

necessarily explain why a patient had received a transfusion outside the accepted parameters.

For example, it's generally agreed in the guidelines that heart failure or heart difficulties are reason for transfusing the patient at a higher hemoglobin level. But in an electronic record, you really cannot find that. It will be in the problem list, but it will be undated. So you don't know if that heart problem preceded the transfusion or followed the transfusion.

In addition, there's an imprecise definition of what heart failure is. Is that heart failure 20 years ago or is it heart failure yesterday? There is no precise definition.

Similarly, a patient with a hole in the aorta may be described as having active bleeding. But if you ask a number of clinicians what is active bleeding, the interpretation of that is going to vary.

Is it 4 ABD's in an hour or is it a Hemovac full inside of 30 minutes? What is active

bleeding?

So because of the imprecise definitions, we did want transfusion committees to look at the clinical record and make a clinical determination rather than to subject the outcome of a measure or a publicly reported item to what is in an electronic record that may be imprecise.

CO-CHAIR FLEISHER: So, just as a reminder, the evidence criteria is that a process or intermediate outcome is based on a systematic review in grading the body of empirical evidence where the specific focus of the evidence matches what is being measured.

There's a lot of concerns I had about specification that I'm hearing from the committee, which would be -- we could address after the evidence, correct? Sal?

MEMBER SCALI: I guess just a clarification. So, as a quality measure, one of the points that was raised about sort of how there are different thresholds, so how would this number, if it ever were to be reported, because if

one center uses seven, another uses eight, is it just a relative percent of where you are with your threshold at your center? Or is there some agreed upon threshold that the centers are being held to in terms of how you're determining the numerator and denominator?

I guess that was my first question in terms of clarification. Two, clarification of the measure, is it for the entire episode of care or only inter-op transfusion?

If it includes the entire episode of care, you've got your open AAA repair and then on post-op day four somebody pulled the trigger and said give them a blood transfusion. But, oh wait, the patient's having an MI and the hemoglobin was eight and cardiology said to go ahead and transfuse, but the institutional trigger is seven. So how does that sort of get factored?

Three, for patients who undergo contemporary aortic surgery actually there's compelling evidence to suggest that restrictive transfusion leads to higher rates of spinal cord

ischemia. And so actually liberal transfusion is now becoming the standard, many of the times, for patients who undergo open thoracical abdominal aneurysm repair and if you're trying to do preemptive protocols to prevent spinal ischemia risk.

Lastly, Jehovah's Witness, I didn't see that that was an exclusion for trigger. So if you would address those.

MS. DOMZALSKI: Spinal surgery patients --

MEMBER SCALI: Patients at risk of spinal cord ischemia after open aortic surgery.

DR. SHANDER: No good data for that for transfusion, by the way.

MEMBER SCALI: I disagree.

DR. SHANDER: You may, but you can quote the data then, because I think that making a statement like that for preemptive transfusion, that would change, I think, the whole transfusion world, in a sense, knowing that that was effective in prevention.

I think that the data comes from retrospective data in terms of prone patients and --

CO-CHAIR FLEISHER: Let's -- let's -- let's focus on the question at hand, which is evidence or whether or not the measure does what its specification is.

MEMBER GROVER: Well, I don't think you're collecting enough data that we would consider -- I don't know whether this is under evidence or not -- but that we would consider it important, or reliable data that we'd consider important.

We have, personally, a cut-off of less than seven hemoglobin in our cardiac surgery patients, but we factor in the age. We factor in if they have COPD. We factor in if it's an incomplete revascularization, all of those things where a hemoglobin, a higher hemoglobin might alleviate the patient.

We also factor in whether the patient is so tired they can't get up and walk and they

may be at risk for a pulmonary embolism if we don't give them a transfusion.

There are several things that come in to -- which you, you know, when you receive the patient at bedside, that affect your judgment. It's not gaming and it's trying -- I mean, we try to really conserve and we've driven down our blood utilization. But you're at risk in this of not capturing those types of things and penalizing or misidentifying people that are abusing the blood transfusions.

MS. DOMZALSKI: And that's why we wanted the clinician to look at the record and factor in all of those types of things that you mentioned, rather than to make a cut-off judgment.

And to answer your previous question, yes, this does cover the entire episode of care. But, again, it's the first unit of any transfusion. It's only one unit per person.

And your first question -- I'm sorry -- could you repeat that again?

MEMBER BILIMORIA: So we've been

looking at -- we've done this with about 45 hospitals and implemented a blood use measure.

It's a registry measure, and then based on the NSQIP data you can tell what -- sort of whether they have CHF and whatnot and so you have some validated definitions to combine. It's a registry measure. It's extremely intensive, but we find a lot of the inappropriate blood use is beyond the first transfusion as well and you can set different thresholds but you can give people a lot of information by setting one.

I was particularly interested in an eMeasure for this because the extraction burden is extremely high. It's the one that gets that most complaints. So maybe there's some way to put these two together in some fashion.

MS. DARDIS: I would like to thank you for bringing that up. I think that one of the unique advantages of an eMeasure for this kind of thing and the reason we've structured this as a tool for performance improvement rather than a metric that is for comparison is the fact that

you're using that real-time clinical data and you can use the same data that you're using to present measures to the board to provide the clinician with decision support.

And so that maybe on a facility level you're working with specific conditions and you have specific thresholds for specific conditions you can use common data elements for measurement and for your specific modifications for your facility and so I think it's an opportunity with eMeasures that we need to explore further.

The other is the reduction of burden in the data collection that you're using the EHR data and we don't have 200 data points in this measure to capture every single condition but we are able to provide some information real-time to the facility.

CO-CHAIR FLEISHER: Fred, did you have another comment? No? Okay. So we are voting on evidence.

DR. WATERS: Can I say one more thing?

CO-CHAIR FLEISHER: Sure.

DR. WATERS: This measure -- part of the intent was to get people to give red cell transfusions based on quantitative data rather than guessing, which is kind of the current standard is that guesswork -- if somebody has a hemoglobin and it doesn't really matter what that particular value is, dependent upon the circumstance, but I mean it should be a quantitative decision rather than guesswork.

CO-CHAIR FLEISHER: Thank you.

MS. QUINNONEZ: We are now voting on measure 3020. Voting is now open for evidence. Option one, high, option two, moderate, option three, low, option four, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the evidence of measure 3020, 5 percent voted high, 58 percent voted moderate, 26 percent voted low and 11 percent voted insufficient.

CO-CHAIR FLEISHER: Okay. We can move on with the discussion. So it passed on evidence.

Performance gap -- do we discuss that? Okay. Barbee?

MEMBER WHITAKER: Okay. So there's not a lot of data for a performance gap but based on our own experience with the AABB there is a gap with what people say they're going to do and what they do. So I think that it's definitely a measure that is worthy of monitoring.

CO-CHAIR FLEISHER: Lynn, any further -- Barry?

MEMBER MARKMAN: Yes. There's also a bigger gap in terms of using an eMeasure and it's not -- it's more in the macro sense of how you capture this data in an eMeasure and subsequently disperse it in a coordinated care center.

So, I mean, we could look at the specific measure and say, you know, there's a gap or not, but there's a gap with this technology. There really is a true gap with this technology and it's in a bigger sense and it's for, you know, for quality care of patients. I'm just going to put that out as a -- as a general comment.

CO-CHAIR FLEISHER: Amy?

MEMBER MOYER: I'm struggling a little with the concept of gap as it relates to this measure. I think what we'll find when the results come back are there differences in transfusion thresholds and where people are being transfused? It sure sounds like it from around the room but I think the bigger question is is there in appropriate variation and that are, you know, relevant to what.

Just saying, okay, there's variance there -- that's one thing. But what's the gap and what are we looking for?

CO-CHAIR FLEISHER: Comments?

DR. WATERS: I'd like to say that the gap is huge. At the University of Pittsburgh we've looked at a number of different surgical procedures. Primarily or at least orthopedic total joint replacements has been an area of focus and for primary total hip replacement for arthritis we saw a variability from one surgeon who transfused 100 percent of his patients to a

different surgeon that transfused 1 percent.

So if you're using gap as a measure of -- as a word for variability, the variability is huge. Fortunately, we've made huge strides in narrowing that gap where most of our surgeons are now down under 5 percent for transfusing total hips, but there is a big opportunity here.

MS. DOMZALSKI: When we place these measures out for public comment, we had about 140 hospitals respond and one of the questions was is there a gap between what this measure asked you to do and what you're currently doing.

Well, more than two-thirds of them reported, yes, there is a difference between what you're expecting and what we're doing and as we were going around to test these at various hospitals every hospital could put their finger exactly on where the problem was in terms of people being transfused over a hemoglobin of nine.

In one hospital it was an OB/Gyn who did it. In another, it was a particular orthopedic surgeon. So yes, there's a gap. It's

apparent at every hospital that we were at.

CO-CHAIR FLEISHER: A.J.?

MEMBER YATES: My understanding is that the tool is -- or the measure is made to help the blood bank do its job better, correct?

MS. DOMZALSKI: Not necessarily the blood bank but the blood transfusion --

MEMBER YATES: Well, the committee. The blood --

MS. DOMZALSKI: Yes.

MEMBER YATES: I stand corrected. The blood transfusion committee. At least where we practice, and again, the University of Pittsburgh -- the act of ordering the blood requires the insertion of the hemoglobin and hematocrit and the indication, that being ongoing bleeding, et cetera, such that that automatically goes to the blood bank.

Is that not a common scenario or is -- are we -- are we talking about much less sophisticated electronic medical records elsewhere?

MS. DARDIS: You are special. Yes. We looked up the literature that exists for clinical decision support and order sets for blood use. There are a few medical sites -- UPMC, University of Iowa where those order sets exist and you have to enter a condition.

There is not standardization around that. We did not see that everywhere we visited just among five hospitals I think one had that built into their ordering. There's a lot of fear in EHR implementation as an -- I guess this is kind of an aside but there's a lot of fear about what you put in front of the physician as a requirement in ordering and I think this is an area where we heard a lot of that.

And so we actually had some excitement in seeing this measure as a way to bring those things into their practice because they've heard about them but they haven't implemented them themselves.

MS. DOMZALSKI: At one of our test hospitals they do have CPOE and they do have

decision support tools and it was described to us that every time they have a new groups of residents their blood utilization goes up because the resident is bypassing all those decision support tools by giving in indication of other for the transfusion.

And so it takes a lot of intervening with each of the residents to stop that other as being an option for transfusion.

CO-CHAIR FLEISHER: Okay. Barbee, do you want to make another comment?

MEMBER WHITAKER: I'd just like to comment that there's no -- there's no category for hemoglobin not measured before first unit transfused and it might be valuable to kind of get at the work-up not being complete prior to -- I mean, not just the type and screen but the H&H, anything that would not be complete prior to the first transfusion or maybe that doesn't happen but it seems like there could be circumstances where you could push the practice to be making sure that there is a hemoglobin on everyone before they do

get a transfusion.

MS. DOMZALSKI: Thank you.

CO-CHAIR FLEISHER: Great. So why don't we vote on gap?

MS. QUINNONEZ: Voting is now open for a performance gap of measure 3020. Option one is high, option two, moderate, option three, low and option four, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For performance gap of measure 3020, 10 percent voted high, 65 percent voted moderate, 0 percent voted low and 25 percent voted insufficient.

CO-CHAIR FLEISHER: Great. So we will continue. So next what do we have -- what else are we -- okay. Now it's actually the specifications -- a discussion.

MEMBER WHITAKER: So like the numerator and the denominator?

CO-CHAIR FLEISHER: Correct, and the exclusions and the comment about, you know --

MEMBER WHITAKER: Okay.

CO-CHAIR FLEISHER: -- some of the comments I heard about, you know, interoperatively, massive bleeding and how the specifications --

MEMBER WHITAKER: Okay. So thank you. Regarding the numerator, I would suggest adding the additional not captured categorization and then --

CO-CHAIR FLEISHER: So let me just get some clarification, either Melinda or Elisa. When we're voting, we're voting as it is but -- and it's a -- and it's a must pass criteria. However, if it doesn't pass and the Joint Commission is willing to make changes that can occur in -- during the call or no?

MS. MUNTHALI: It depends on how significant those changes are and how quickly the Joint Commission can turn those around. So we'd have to get an agreement from you today to say that you can make those changes by the post-comment call.

CO-CHAIR FLEISHER: We'll probably vote as is or as is with modification might be something we can discuss if it doesn't pass. But we'll first -- but please, as is and things that are -- they should do or must do from your perspective.

MEMBER WHITAKER: Okay. I'd like to comment that patients under the age of 18 can benefit from hemoglobin optimization so would like to see it extended to pediatric patients. I don't know if that's the standard of care -- for the quality measures here but obviously that would be optimal.

And then regarding exclusions, other emergent surgeries which may have high blood loss, if there's not time to get a hemoglobin in advance might be an exclusion and then the emergency department -- I question whether they should be given an opt out categorically and whether they could also wait or, you know, benefit from paying attention to hemoglobin levels before they transfuse.

MS. DARDIS: I'm sorry. Can I ask for a clarification? Would you like to see emergency department excluded from the measure or included?

MEMBER WHITAKER: Included.

MS. DARDIS: Okay. Thank you.

CO-CHAIR GUNNAR: Fred?

MEMBER GROVER: I'm concerned under your specifications for who's included just as a hemoglobin again and it doesn't included hemorrhagic shock, bleeding, current active bleeding -- some of those things we talked about earlier which are obvious indications for a transfusion regardless of the hemoglobin.

CO-CHAIR GUNNAR: Barry? Any other comments? Amy?

MEMBER MOYER: I apologize for timeliness. I'm just going to ask instead of trying to look this up. So a denominator exclusion is patients whose first unit of whole blood was given while in an emergency department. If these are elective surgeries shouldn't admission source emergency department perhaps be

excluded entirely?

MS. DARDIS: This measure is for all patients, not elective surgicals.

CO-CHAIR FLEISHER: Lynn? Did Lynn already go? Nothing? Anybody else have any comments on the specifications?

MS. QUINNONEZ: Voting is now open for eMeasure specifications for measure 3020. Option one, high, option two, moderate, option three, low and option four, insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the eMeasure approval for trial use for measure specifications of measure 3020, 5 percent voted high, 37 percent voted moderate, 47 percent voted low and 11 percent voted insufficient.

CO-CHAIR FLEISHER: So this is consensus not reached, which means we will continue to evaluate the measure. One of the questions is is the Joint Commission interested in working with some of the things they heard today

to potentially change it during the phone call.

MS. DARDIS: I think from a time frame perspective, yes. I think it's unclear -- and Ann, jump in -- but I think it's unclear which changes that were suggested are the changes that are most of import to the group. I think any changes we make we want to take back to our advisory panel.

CO-CHAIR FLEISHER: Sure, Barbee, do you have a -- or maybe in the next couple days we can get some consensus on -- can you work with staff to give some significant recommendations of what is most concerning?

MS. WATT: I was going to say this is Ann, you know that. I'm sort of confused actually about what the ask is.

It was our understanding as you're looking at the HQMF specifications you were actually looking at the construct of the eMeasure as opposed -- the technical construct as opposed to the content of the measure itself.

And so as I'm listening to the

conversation I'm understanding that you are asking us to change the clinical content of the specifications irrespective of the HQMF specification. Is that correct?

CO-CHAIR FLEISHER: Can you be specific? Can someone articulate? Because I missed the --

MS. SKIPPER: I felt like I heard a committee member ask for the measure to be extended to the pediatric population and then also --

CO-CHAIR FLEISHER: That's not a critical issue. That's a typical ask of this committee.

MS. WATT: Yes, I understand that.

CO-CHAIR FLEISHER: So I think I'll take chair prerogative in saying if that doesn't happen that's -- we'd like it to be. Anything else that was asked, Barbee?

CO-CHAIR GUNNAR: I think the -- Fred's -- his --

MEMBER GROVER: Yes, I mean, my concern

was that there's more than the hemoglobin that's a trigger to transfusion and at least on some of the major ones, such as active bleeding hemorrhagic shock, if you're just looking at hemoglobin you're going to miss a fair amount of things. Even in the operating room, you can have things that you aren't expecting on elective cases that require them. So I think --

CO-CHAIR GUNNAR: How will the -- how will the threshold reflect the clinical condition?

MEMBER GROVER: Yes.

CO-CHAIR GUNNAR: And is it reflective of current variable practice across health care facilities?

MEMBER GROVER: Right. There is variation in what your hemoglobin level is but if people are bleeding significantly and they're in shock there's very little variation in that, I would think.

MS. DOMZALSKI: The number in the threshold by itself is insufficient to determine if a transfusion could have been avoided. It does

need to be evaluated by a clinician in relationship to the clinical signs and symptoms, which is why this is being done for the transfusion or blood usage committee to review.

As before, it's very difficult to specify what active bleeding is. Every clinician has a different concept of what that is and that's why we'd like the committee to look at this when it turns out that a transfusion was given above a threshold whether it be seven in your facility or eight or nine. And so that's why we'd like the committee to look at it. We recognize that that would be perhaps understandable.

DR. WATERS: I've got to say that measuring a hemoglobin in the operating room takes 10 microliters in about 20 seconds. So I don't know how it would interfere with the process of getting a patient blood.

MEMBER GROVER: Well, if you wait for a hemoglobin change in somebody who's bleeding rapidly, they could be dead.

CO-CHAIR GUNNAR: Exactly. Hemoglobin

will be normal when you're providing active transfusions to an actively bleeding patient.

CO-CHAIR FLEISHER: Liz?

MEMBER EREKSON: So I had said this before but just to remind, in pregnant patients undergoing postpartum hemorrhage they may be considered, like, traumatic injury and you may want to exclude them or consider them in a different measure.

CO-CHAIR FLEISHER: So -- Allan?

MEMBER SIPERSTEIN: Yes. So again, the ratio as expressed, as I've said before, doesn't make clinical sense, but let me play devil's advocate for a minute.

You bring all your patients to anemia screening. Nobody ever goes to the operating room with a crit, you know, less than -- or a hemoglobin less than 10. Well, if you look at your measure a couple years from now, all of your transfusions will have been in patients who are over 10 because they've had, has been nicely expressed, unexpected bleeding in the operating

room. And so it's not really clear to me how you're going to use this number over time to monitor quality improvement.

Again, as I've said, each of the numerator and denominator are laudable goals but the end result as expressed does not make clinical sense to me.

DR. SHANDER: Well, from your lips to God's ear, as they say, but because I don't think we're there yet. But I think that if that is the consequence of anemia management is that transfusions are -- you know, we're going to have to relook at this measure at some point.

I think the idea -- and I think was already mentioned -- is this is basically just to trigger a review. If there is justification for a transfusion at high level of hemoglobin, meaning that the patient is hemorrhaging, I don't think there's going to be an argument and that's not the intent. The intent is really to have a measure for somebody to start in terms of a threshold for the transfusion.

Once you've identified that hemoglobin is not the indication for a transfusion, as we all talked about was hemorrhage, then I don't think that there is an issue and to some extent I think that's -- there is some comments on the exclusion on that, if I recall.

CO-CHAIR FLEISHER: So let me throw out something to the Joint Commission on what I'm hearing. I actually think if you stratified this by an intraoperative transfusion versus a post-operative transfusion, you might get some interest -- it might have some more validity or some other ways to get to some of the concerns.

I don't know whether others think about that but that may be a way to -- because what I'm hearing from the committee, and I actually -- I don't want to have to worry about getting a hemoglobin level and I'll say 10, that'll actually it would be the hemoglobin level -- it'd be the pre-op hemoglobin level and that will make it very different to compare to, you know, a transfusion.

But if I said intraoperatively, there is a preoperative hemoglobin level and that's why we see some screwy data that doesn't make sense. Post-operatively, I think, most of the time we usually have time to get a hemoglobin level unless they're acutely bleeding. So I don't know if that's something.

MS. DARDIS: I understand the concern of the intra-op setting. One thing I'd like to raise is that this is an all -- again, this is an all patients measure. It's part of the set that we brought to the surgical committee. This specific measure is for all hospitalized patients, not just surgical patients.

So the strata might have to look at something different, but the intent was to improve blood management for all patients with this measure, not just the surgical. So I'm wondering if another option -- I'm not the clinical expert on this -- would be to exclude the intra-op transfusion.

CO-CHAIR FLEISHER: Is that of -- or

stratified, which would be the other way to do it. Barbee, did you have another comment?

MEMBER WHITAKER: My other comment was just the value of having an eMeasure for this to establish the infrastructure to be able to monitor and report internally.

I think that eMeasure process would be very valuable for everyone even if -- even if it's not -- if we could keep it from being something that you're -- something that you're looking at but not something that you're held to.

CO-CHAIR FLEISHER: And thank you. Comment? Rick?

MEMBER DUTTON: I think some of the reaction of the committee is to the very arbitrary nature of this: transfusion judged on a hemoglobin level. The measures that have done better here, just in the last two days, have been ones that are measuring physician or practitioner judgment.

So if you could get at the underlying physiology of this and, say, create an outcome measure where, of the transfusions given, what

percent were indicated because of bleeding or ischemia or a clinical indication for the transfusion.

Do I know how to e-specify that? No. But that's where we would like to see a good quality measure appear.

MS. DARDIS: Understood, that is not a eCQM.

MEMBER MOSS: I just want to ask if you could help talk us through how this would sort of play out in the real world. So you do these thresholds.

A certain number of patients meet them. That will trigger a local review. Each case will be looked at on an individual basis with a clinical decision whether it was appropriate or not. Then we come back next year and do these thresholds again.

We don't know whether they should go up or down or stay the same. How will be know if this is working?

DR. WATERS: Well, getting back to the

University of Pittsburgh, what we've done for our anesthesiology department is we've made this measure for -- it's a value measure for our anesthesiologists and how they get paid.

What we found from hospital to hospital is that one hospital will transfuse 95 percent of their patients in the operating room with a hemoglobin less than eight, whereas another hospital in our system will transfuse their patients only 45 percent of the time with a hemoglobin less than eight. So there's huge cultural variation from institution to institution, or rather, hospital to hospital.

But the opportunity here is to decrease that variability from hospital to hospital and I think there's a large opportunity there.

CO-CHAIR FLEISHER: Okay. So we're worried about losing our quorum. So Barry, specifically --

MEMBER MARKMAN: Yes, just one question. Now, this is for a trial, right?

CO-CHAIR FLEISHER: Yes.

MEMBER MARKMAN: And then they'll come back -- I mean, can they tweak or, you know --

CO-CHAIR FLEISHER: So the answer is I would hope that a few people over the next couple days can send to Melinda if they have specific questions for the Joint Commission to consider in how they specify, correct?

And if the Joint Commission is interested in addressing them then potentially we could go from this gray area to get it back into where it would be -- would pass the criteria. Am I accurately -- at the time of the discussion, yes.

MS. WATT: I'm sorry, and I think that maybe, Jason, this is a question for you. It was our understanding that this category had to do with the appropriateness of the technical specification of the measures, not the clinical content of the specifications but the way that they're represented for collection by an electronic record.

What we are talking about has nothing

to do with that, and so I'm a little confused as to what -- we would be happy to look at making modifications to the clinical specifications but I'm not sure how that relates to this criterion for the HQMF specifications and we're hoping, Jason, that you can clarify.

MR. GOLDWATER: So Ann, you're correct that this part of the feasibility really is talking about being the electronic construction the HQM on the format of the measure, the sources the data collection could be implemented as not referencing the clinical specifications. She's correct.

MS. MURPHY: But the piece of the measure, Jason -- if you'd help us just a minute -- the piece of the measure that they're talking about and that they are -- they voted on are the specifications of the measure.

MR. GOLDWATER: That was the path that they just voted on. Correct.

CO-CHAIR FLEISHER: Right.

MS. MURPHY: Yes, that's what we're

talking about.

MR. GOLDWATER: Those were the clinical specifications, right.

MS. MURPHY: Yes.

CO-CHAIR FLEISHER: Yes, we voted correctly from your perspective if we had concerns about the clinical specifications or --

MR. GOLDWATER: That's correct. Right. This -- this section, what we were just at, feasibility is talking about the electronic development of the measure. That's correct.

MS. MURPHY: And the first part of the feasibility are the specifications. That's the one piece that we understood that the committee was to vote on were those specifications of the measure that in fact are listed in the feasibility section.

MR. GOLDWATER: That this -- the electronic specifications of the measure lead to the metric that matches the purpose and intent. That's correct. The specifications that were voted on before are clinical in nature.

MS. WATT: You're talking about the evidence and the gap that we voted -- that were voted on previously -- the content of the measure.

MR. GOLDWATER: That's correct. That's what you just voted on recently.

CO-CHAIR FLEISHER: And then specifications.

MR. GOLDWATER: That's passed, right. The electronics that you were just on in the feasibility part in specifications, is the construction of the measure electronically, will it lead to the metric that matches the intent of the measure?

CO-CHAIR FLEISHER: So the concerns of many of the committee is whether the clinical specifications were valid to -- it's almost validity if we were in a different measure. Where do we put that in this construct?

Because what we were told is when you come back, we don't vote on that again. So I think what you're hearing is some confabulation.

MR. GOLDWATER: So, I mean, I think if

there's concerns about the clinical nature -- clinical specifications of the measure then -- I don't know, that's kind of a -- you have to sort of balance that, I guess.

If you've got concerns about the clinical specs and you're not sure about whether or not it's conducive or not to, I guess, improving overall quality based on the objective and intent of the measure, then when you get to this feasibility part all you're really looking at there is, you know, electronically speaking was it constructed in such a way that it matches the overall intent?

So if you've got a problem with the intent, subsequently, there may be an issue with the way the measure is designed.

CO-CHAIR FLEISHER: I'm not sure if that helped you.

MS. WATT: Thanks. We'll look forward to hearing what Melinda shares with us.

MS. MURPHY: So I think -- and Christy, will help -- if they send the comments to surgery

--

MS. SKIPPER: Yes. Surgery@qualityforum.org. Any comments or input and provided to the Joint Commission on this measure.

CO-CHAIR FLEISHER: Let me be specific. What we want is comments, not to change the whole measure. If there are tweaks to the way inclusion, exclusion -- and please correct me if I'm overstating my interpretation.

MS. SKIPPER: It's clinical content.

CO-CHAIR FLEISHER: Clinical content that they could take back to their TEP to see whether or not they agree with us. That might change how we vote on the November call. That's what we're looking for over the next three to four days.

MR. GOLDWATER: And then whatever changes -- if there are recommended changes or changes that are made, then subsequently, the electronic specifications may also be changed as well, potentially. Right.

CO-CHAIR FLEISHER: Fred?

MEMBER GROVER: I'm just a -- I think we've got it backwards here because what we're trying to do -- at least my thought was when we got into this -- was to streamline this for you so we could give you this advice up front because if you go with what you've got now, I'll tell you, when this comes back this doesn't meet, you know, what we understand are the important variables.

There's a big disconnect there. You'd be wasting your time and if you get it -- if we tell you now what we think is important, then when you go to the electronic thing you've hopefully got it as close together so we don't come back a year from now and say why did you do this and start all over again.

CO-CHAIR FLEISHER: Thank you. Okay. We should continue. Feasibility.

MEMBER WHITAKER: I don't have an comments on feasibility.

CO-CHAIR FLEISHER: Okay.

MS. QUINNONEZ: Voting is now open for

the feasibility of measure 3020. Option one, high, option two, moderate, option three, low and option four, insufficient.

(Pause.)

MS. QUINNONEZ: Looking for one more vote. If you could resubmit your votes please and point them this way.

(Voting.)

MS. QUINNONEZ: Voting is now closed. Feasibility for measure 3020 reads 18 percent voted high, 35 percent voted moderate, 35 percent voted low and 12 percent voted insufficient. Consensus not reached.

Are there any further comments on use and usability? We are now voting for use -- usability and use of measure 3020. Option one, high, option two, moderate, option three, low and option four, insufficient information.

(Voting.)

MS. QUINNONEZ: All votes are in and voting is now closed. For the usability and use of measure 3020, 0 percent voted high, 29 percent

voted moderate, 35 percent voted low and 35 percent voted for insufficient information.

CO-CHAIR FLEISHER: Not a must pass. So next. So okay, so vote on that. So we are done with this measure? Right, and we can't vote on that, given -- right.

So since consensus is not reached we will not vote on this. We will be voting, going through this again on the November call and hopefully we'll be able to get those comments in to make it -- that the committee is interested in going forward.

So the last measure is 3021, blood usage.

MS. DOMZALSKI: Measure 3021 looks at blood usage in general in selected elective surgical patients. It has a denominator of patients who had the timely pre-op assessment that is between 14 and 45 days prior to the procedure.

And by identifying which of these patients had pre-op transfusions, which is the numerator, there is an opportunity again for the

blood management review committee by review of the case to identify any other blood conservation methods that should have been undertaken in order to avoid that transfusion.

Whether those methods are preoperative anemia correction, use of a Cell Saver or other technologies or education in general to what restrictive transfusion strategies encompass that would be a matter of determination for each hospital's committee to decide.

Again, it may confirm that no additional education efforts are needed. But it was for that particular month or that particular time frame a number of patients whose transfusions were unavoidable.

But we do want the transfusion committee to look at those patients who had transfusions with an eye towards a more restrictive strategy.

CO-CHAIR FLEISHER: Okay. Larissa? Barbee, either one?

MEMBER WHITAKER: Thank you. So this

measure is intended to assess the effectiveness of the preoperative anemia screening by identifying those patients who had been appropriately screened but still required a perioperative blood transfusion. So not as -- the effectiveness of actual -- the actual screening but the treatment after the screening.

I had a little trouble with this because it seems like there's still a big gap in the denominator whether there would be sufficient 14 to 45-day preoperative anemia screening because we had the whole discussion about the appropriateness of that.

So that's really my comment on this particular measure. I do have some comments on numerator and that I think nonautologous shouldn't be mentioned at all.

CO-CHAIR FLEISHER: So I'm going to actually ask staff to comment. Since we didn't pass that first one, it makes it difficult to look at the issue of in those who got transfused was it appropriate. So what's the best way to address

that? It's a question, Barbee, for you too. Yes, Barry?

MEMBER MARKMAN: Can I just take that variable out and still proceed with the data collection?

CO-CHAIR FLEISHER: So I guess that's a question to the Joint Commission because I think what I'm -- one question is without that being passed, how do people feel about the rest of the measure? It's up to you whether you want us to vote with or without that question.

MEMBER WHITAKER: You take that out.

MEMBER MARKMAN: It doesn't mean -- I mean, in the measure itself, I mean, I don't know. There's certain variables that you're looking at. I mean, this is more of a data collection type of measure.

CO-CHAIR FLEISHER: Other comments?

(Pause.)

CO-CHAIR FLEISHER: So I'm told that we don't have a quorum. So we -- for this measure, unfortunately, so we will not be voting. No, this

is not Chicago. So why don't we --

MEMBER MARKMAN: Yes. Yes, can we call them?

CO-CHAIR FLEISHER: How many do we need? Call Rick? I don't know if I have Rick's cell.

MS. DOMZALSKI: I have his.

CO-CHAIR FLEISHER: You have it? Can you tell him to dial in? He's not on the phone. I don't have Barbara's -- I don't think I have Barbara's cell. Other comments regarding evidence?

DR. SHANDER: Lee, I can't -- I can't vote. I would love to but I can't.

CO-CHAIR FLEISHER: Any comments from the committee on this measure?

MEMBER SIPERSTEIN: I mean, this discussion parallels exactly 3016 and so we've already been through this. There's the rub.

CO-CHAIR FLEISHER: So your suggestion, Allan, is --

MEMBER SIPERSTEIN: Well, if there's

not a quorum, we'd have to follow this with a phone conversation but I think a good chunk of the discussion it would just be, you know, repetitive to redo the discussion at this point. If we need to vote, we can vote on the phone conference.

CO-CHAIR FLEISHER: Yes, we can and we can do the evidence review. We can do the evidence on this, or just actually if you give it to -- if Christy can take his vote over the -- the hand. Okay. We have a quorum. Can you call the vote?

MS. QUINNONEZ: I can. We are now voting --

CO-CHAIR FLEISHER: Wait one second.

(Pause.)

CO-CHAIR FLEISHER: So one important thing is they actually don't say the fact -- it's not fully linked, this measure, to 3016. They use the term -- use the term timely preoperative screening, nut it's -- is it defined in the specifications as 3016? It does say that in the specifications? Okay. So why don't we call the

vote?

MS. QUINNONEZ: We are now voting on measure 3021. Voting is now open for evidence. Option number one is -- actually there is no option number one.

Option number two is moderate. Option number three is low and option number four is insufficient. Option two, moderate. Option three, low, and option four insufficient.

(Voting.)

MS. QUINNONEZ: All votes are in. Voting is now closed. For the evidence of measure 3021, 25 percent voted moderate, 44 percent voted low and 31 percent voted insufficient.

CO-CHAIR FLEISHER: So this measure does not go forward. It would be great from my perspective to get feedback on how people think about this trial measure and then maybe, Jason, we can continue the discussion with some input from the --

MR. GOLDWATER: Always welcome.

CO-CHAIR FLEISHER: Can we -- can we

open the phones for public comment?

OPERATOR: Thank you. At this time, if you'd like to make a comment please press star one. We'll pause for just a moment.

(Pause.)

OPERATOR: And there are no public comments at this time.

MS. SKIPPER: Thank you, everyone. We have successfully made a recommendation on almost all of the measures but I'm glad that we were able to get through this last one.

I just wanted to share that we do have a post-meeting call next Thursday, August 25th, from 2:00 to 4:00 p.m., and then the next call that we would have as a group is our post-draft comment call. So once we do put the report together and put it out for comment, we'll bring you all back to respond to the comments and vote on any measures where consensus is not reached.

Finally, I want to say that for our new committee members, to Karl, Barbee and -- I'm drawing a blank -- Sal, yes, sorry about that --

we will be asking you to pull a number from our magic bowl down here to select your committee term.

New committee members may serve -- well, you'll pick a number and either you'll be assigned a two-year term or a three-year term and after your first term, you may elect to serve a second term. And for all remaining and continuing committee members, I'll be in touch with information on your next term.

So right now I'm going to go around and have our new committee members select their term and state that for the record.

CO-CHAIR FLEISHER: Before we all leave, I'd like to say a thank you to the staff. Really appreciate it, Christy and the team.

(Applause.)

CO-CHAIR FLEISHER: For Melinda also for leading us and appreciate Elisa, please send our thanks to Marcia and Helen for coming in and helping us out when we got into quandaries.

MEMBER SCALI: For the record it's two

years -- Sal Scali.

MEMBER BILIMORIA: Karl Bilimoria -- two years.

MEMBER WHITAKER: Barbee Whitaker -- three years.

(Whereupon, the above-entitled matter went off the record at 2:34 p.m.)

E

P

S

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Surgery Phase 3
Standing Committee

Before: NQF

Date: 08-17-16

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