NATIONAL QUALITY FORUM

+ + + + +

SURGERY STANDING COMMITTEE

+ + + + +

THURSDAY
MARCH 19, 2015

+ + + + +

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

COMMITTEE MEMBERS:
LEE FLEISHER, MD, Committee Co-Chair; Professor and Chair of Anesthesiology, University of Pennsylvania; American Society of Anesthesiologists
WILLIAM GUNNAR, MD, MPH, Committee Co-Chair, Director, National Surgery Program Office, Veterans Health Administration
ANTHONY ASHER, MD, FAANS, FACS, Carolina Neurosurgery & Spine Associates*
ROBERT CIMA, MD, MA, Professor of Surgery, Mayo Clinic
RICHARD DUTTON, MD, MBA, Executive Director, Anesthesia Quality Institute
ELISABETH EREKSON, MD, MPH, Dartmouth Hitchcock Medical Center
JOHN HANDY, MD, Thoracic Surgeon, American College of Chest Physicians
FREDERICK GROVER, MD, Professor of Cardiothoracic Surgery, University of Colorado School of Medicine
MARK JARRETT, MD, MBA, Chief Quality Officer, Associate Chief Medical Officer, North Shore-LIJ Health System*

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
BARBARA LEVY, MD, FACOG, FACS, Vice President, Health Policy, American College of Obstetricians and Gynecologists
BARRY MARKMAN, MD, Senior Medical Director, 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 Chair, Department of Pharmacy Practice, University of Nebraska Medical Center; American Society of Health-System Pharmacists
COLLETTE PITZEN, RN, BSN, CPHQ, Clinical Measure Developer, MN Community Measurement
LYNN REEDE, DNP, MBA, CRNA, Senior Director, Professional Practice, American Association of Nurse Anesthetists
GARY ROTH, DO, FACOS, FCCM, FACS, Medical Director, MHA Keystone Center
CHRISTOPHER SAIGAL, MD, MPH, Professor, UCLA
ALLAN SIPERSTEIN, MD, Chairman Endocrine Surgery, Cleveland Clinic
LARISSA TEMPLE, MD, Colorectal Service, Department of Surgery, Memorial Sloan-Kettering Cancer Center
MELISSA THOMASON, MS, PMP, Patient/Family Advisor, Vidant Health
A.J. YATES, MD, Associate Professor, University of Pittsburgh Medical Center

NQF STAFF:
MARCIA WILSON, Senior Vice President, Quality Measurement
JULIET FELDMAN, Project Manager, Stakeholder Collaboration
KAREN JOHNSON, Senior Director
ANDREW LYZENGA, Senior Project Manager
MELINDA MURPHY, Senior Director
YETUNDE ALEXANDRA OGUNGBEMI, Project Analyst

ALSO PRESENT:

RONY ADAM, MD, American Urogynecologic Society*
SVEN BERG, MD, MPH, CPE, Quality Insights of Pennsylvania
SHERYL DAVIES, MA, Stanford University*
ELIZABETH E. DRYE, MD, SM, Yale Center for Outcomes Research & Evaluation (CORE)
VIVIENNE HALPERN, MD, Society for Vascular Surgery*
COLLEEN HUGHES, American Urogynecologic Society*
BRAD JOHNSON, MD, Society for Vascular Surgery
JANE LUCAS, RN, Quality Insights of Pennsylvania
TOM MILLER, PhD, American Society of Anesthesiologists*
JAMES MOORE, MD, American Society of Anesthesiologists*
DAN MORGAN, MD, American Urogynecologic Society*
CRAIG PARZYNSKI, MD, Yale Center for Outcomes Research & Evaluation (CORE)
MATTHEW POPOVICH, PhD, American Society of Anesthesiologists
MARK PRESTON, MD, American Urogynecologic Society*
SAMANTHA J. PULLIAM, MD, American Urogynecologic Society*
PATRICK ROMANO, MD, MPH, UC Davis
CAROL STARKS, AHRQ
GARTH HARRISON UTTER, MD, MSc, UC Davis
*** present by teleconference**

TABLE OF CONTENTS

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

(8:30 a.m.)

DR. FLEISHER: Good morning. Welcome to the second meeting of the Surgery Standing Committee. I'm Lee Fleisher. I'm an anesthesiologist from the University of Pennsylvania.

Just as a reminder, when we are speaking we need to hit the button. And when we're not speaking, we need to turn it off.

So the nice thing is this is -- these are now standing committees, so we have been together before. And for some of us we will be together for two to three years. We have a number of measures to go through, and the agenda, as you will see, is really structured as everything but the STS measures today, and the STS measures will be tomorrow.

But as a standing committee, we are actually -- hopefully, we will have some robust discussions about general concepts today, in addition to specific measures.

Hopefully, Bill will be here shortly. But I think at this point I can turn it over to -- Amanda, are you going to --

AMANDA: First, I just want to open the line with the operator. So, Operator, can you please open the line?

OPERATOR: You are live.

AMANDA: Great. Thank you.

Can you inform us if Dr. Asher or Dr. Jarrett are on line?

OPERATOR: Certainly. We will put a message in the chat, so that you will see it.

AMANDA: Thank you.

DR. FLEISHER: Welcome, everybody. It's good to see you all again. I will hand it over, actually, to Melinda to say a few words before we get into our disclosures of interest.

MS. MURPHY: Thank you. And I told Andrew I wanted to say hello to everyone. I have not been with this -- the entirety of this group, and was with Surgery Project some years ago. But it is very enjoyable to me to be back with this

group and back inside the NQF office.

I work from the field most of the time, so thank you very much for being here, for taking the time, for investing and being very interested in this topic on behalf of doing the right thing for the patients in this country. So welcome, and thank you for being here.

DR. FLEISHER: So with that, Marcia, I think, we can just jump right into the disclosures. As you did the last time you were here, I think we will sort of redo that process of asking you to disclose any interests, but I will let Marcia speak about that.

MS. WILSON: Again, I am Marcia Wilson. I am Senior Vice President of Quality Measurement here at National Quality Forum, and I'm going to walk us through the disclosure process this morning. I have a few comments to make.

You received a disclosure of interest form from us before you were named to this committee, and on that form we asked you a number

of questions about your professional activities. And today we'll ask you to orally disclose any information you provided that you believe is relevant to the subject matter before the committee.

When you do your oral disclosure, it is not necessary to review your entire resume. We are only interested in your disclosure of information that is directly relevant to the work that the committee will be doing. We are especially interested in grants, research or consulting, but again, only if it relates to the subject matter before the committee for the next two days.

A couple of reminders. You sit on this committee as an individual. You do not represent the interests of your employer or anyone who may have nominated you to this committee.

Secondly, we are not only interested in disclosure of activities where you were paid. For example, you may have been a volunteer on a

committee. We are looking for you to disclose those types of activities as well, but, again, only if it's relevant to the subject matter before the committee.

Just because you disclose it does not mean you have a conflict of interest. We do oral disclosures in the interest of openness and transparency.

So we will do this by going around the room. Please state your name, who you are with, and if you have anything to disclose. And then when we're finished with everyone who is present here in D.C., we will turn to the people on the phone and I'll call your name. So, sir, if we may start with you, by giving your name, who you're with, and if you have anything to disclose.

DR. GROVER: I'm Fred Grover from Denver, Colorado. I'm a cardiothoracic surgeon, and the major areas -- I have been very involved with the STS database over the years, although I am more distant from that now, so that is an area

you need to know about.

Also, I served on the American College, although -- not only at this meeting, but on the American College of Cardiology NCDR Board, so I have that.

So I suppose anything related particularly to the STS or competing measures within STS I would have to -- at least if I make a comment, say that. With the STS, I will just recuse myself.

MS. WILSON: Okay. Thank you.

DR. HANDY: John Handy. I'm a thoracic surgeon from Portland, Oregon with no disclosures.

MS. WILSON: Thank you.

MR. MARKMAN: Barry Markman. I'm a plastic surgeon. I'm also a Senior Corporate Medical Director for Aetna. And I support the SIU plans in Medicaid throughout their enterprise. That's all I can disclose.

I also have a patent on biologics, and that is pretty much what I need to disclose.

MS. WILSON: Thank you.

DR. EREKSON: Hi. Liz Erekson. I'm a gynecologic surgeon. I'm at Dartmouth, and I also am at the Dartmouth Institute, and I have no disclosures.

MS. WILSON: Thank you.

MS. MOYER: Hi. Amy Moyer. I'm the manager of value measurement at The Alliance. We are a cooperative of employers who purchase health care directly, and I have no disclosures.

MS. WILSON: Thank you.

DR. MOSS: Hi. I'm Larry Moss. I'm a pediatric surgeon and surgeon and chief at Nationwide Children's Hospital. I sit on the Verification Review Committee for the American College of Surgeons that verifies children's surgery centers, and the Executive Steering Committee for the National Surgical Quality Improvement Program of the American College of Surgeons.

MS. WILSON: Thank you. Go ahead. If you have a red light on your mic and you're not

speaking, please turn it off.

MS. THOMASON: I am Melissa Thomason. I am a real patient, and I'm happy to be here. And I am a patient advisor from Eastern North Carolina. I have no disclosures.

MS. WILSON: Thank you.

MS. PITZEN: Collette Pitzen, Minnesota Community Measurement. I'm a measure developer. However, we have no measures in the general surgery portfolio. So I am happy to be here. Thanks.

MS. WILSON: Thank you.

MS. REEDE: Good morning. Lynn Reede. I'm a certified registered nurse anesthetist from the American Association of Nurse Anesthetists. Even hard for me to say. I'm Director of Practice. I have no disclosures.

MS. WILSON: Thank you.

DR. DUTTON: Hi. Rick Dutton. I'm an anesthesiologist and the Chief Quality Officer of the American Society of Anesthesiologists. We are a measures steward for one of the measures,

which I will be recusing myself from discussion. I also served on the Technical Expert Panel for the hospital readmission measure that is on here today.

DR. LEVY: Good morning. I'm Barbara Levy. I'm an OB/GYN. I'm Vice President for Health Policy at the American College of Obstetricians and Gynecologists. I serve on the Executive Committee, the PCPI, which has no competing measures for these today.

MS. WILSON: Thank you.

DR. SAIGAL: I'm Chris Saigal. I'm a urologist at UCLA. I sit on our -- the AUA's Quality Improvement and Patient Safety Committee, Data Committee and the EMR Committee. I'm a PI at RAND where we do work on quality using claims, and I'm the co-founder of a company called WiserCare that does shared decisionmaking software.

DR. OLSEN: I'm Keith Olsen, a pharmacist and professor at the University of Nebraska Medical Center. Disclosure -- I am an

elected member to the Board of Regents of the American College of Critical Care Medicine.

MS. McCARTY: I am Kelsey McCarty, and I'm currently in transition. Recently resigned Anesthesia Quality and Safety Manager from Massachusetts General Hospital, and incoming Director of Operations and Strategy for Boston Medical Center. I have no disclosures.

DR. ROTH: I am Gary Roth. Clinically, I'm a cardiothoracic surgeon. I'm also the Medical Director for the Michigan Health and Hospital Association. And I have no disclosures.

DR. YATES: Adolph Yates. I'm from Pittsburgh with the University of Pittsburgh Medical Center. Because we are going to be talking about gap analysis this afternoon, I have been on -- and still serve on the Technical Expert Panels for PhysicianCompare.gov with no measures for review today.

I am also on the Technical Expert Panel for the cost measure being developed for

CMS by Yale CORE for the total joint replacements.

And, finally, because it is part of the job and we look at policy issues, I am the Chairman of the Evidence-Based Medicine Committee for the American Association of Hip and Knee Surgeons.

DR. TEMPLE: I am Larissa Temple. I'm a colorectal surgeon from Memorial Sloan-Kettering in New York. I am Vice Chair of Quality, and I have no disclosures.

DR. SIPERSTEIN: Allan Siperstein. I do endocrine surgery at the Cleveland Clinic. I have no disclosures.

DR. CIMA: I'm Bob Cima, a colorectal surgeon at the Mayo Clinic, and Vice Chair of the Mayo surgical practice. No disclosures.

MS. WILSON: Okay. Thank you. And on the phone is -- oh, I'm sorry, my co-chairs. Sorry.

DR. GUNNAR: Bill Gunnar, National Director of Surgery, Department of Veterans

Affairs. I have no disclosures.

DR. FLEISHER: Lee Fleisher, Chair of Anesthesiology at the University of Pennsylvania. I'm on the Evidence-Based -- the Practice Parameter Guideline Oversight Committee for the American College of Cardiology, American Heart Association.

I'm the Technical -- the Medical Advisory Panel of the Technical Evaluation Center of Blue Cross/Blue Shield, and I was a member of the Committee on Practice and Outcome Measures for the American Society of Anesthesiologists.

That last one conflicts with the temperature measure, and I will recuse myself.

MS. WILSON: Okay. Thank you, Lee.

DR. FLEISHER: And I have grants from AHRQ with Jeff Silber on process measurement.

MS. WILSON: Thank you. And now I think we are ready to go to the phone. Is Anthony Asher on the phone with us? Is Mark Jarrett on the phone with us? Okay. Not at this time, so we may go back to them when they join

the meeting a little later on.

And just a few closing comments. I would like to remind you that if you believe you may have a conflict of interest at any time during the meeting, please speak up. You can speak in real time, or you can approach either of the co-chairs or any of the NQF staff.

If you believe that a fellow committee member may have a conflict of interest or is behaving in a biased manner, you may point this out during the meeting, again, by speaking up, approaching the co-chairs, or going directly to NQF staff. We don't want you to sit in silence if you believe there are any irregularities due to conflict or bias. So please do speak up.

And, at this time, are there any questions or anything we need to discuss based on the disclosures today?

AMANDA: Marcia, Dr. Sawin is on the phone as well.

MS. WILSON: Thank you. Doctor?

DR. SAWIN: Good morning. I apologize

for not being there in person. I'm Bob Sawin. I'm the Surgeon-in-Chief at Seattle Children's Hospital, Chairman of the organization at Children's Hospital Surgeon-in-Chiefs, and I have no disclosures.

MS. WILSON: Thank you very much. And I think we're done with the disclosures of interest, Andrew.

DR. LYZENGA: I think so. And maybe we should introduce ourselves as staff. I know we've got a few new folks here, faces in the room. I'm Andrew. I met you -- most of you before. I'm a Senior Project Manager here at NQF. I worked on a number of safety projects and other projects as well, and have staffed this Surgery Committee the last time around. So, again, welcome. Good to see you all again.

MS. FELDMAN: Good morning, everyone. My name is Juliet Feldman. I'm a Project Manager here. This is my first CDP project, so I'm very excited to experience today's in-person meeting. And, yes, that's it.

MS. MURPHY: I am Melinda Murphy. I'm a registered nurse. I have been with NQF for 10 years in various capacities and have worked primarily with safety-related activities.

MS. JOHNSON: Good morning. I'm Karen Johnson. I am a Senior Director here at NQF, and I am really just here to observe today.

MS. WILSON: And, again, I'm Marcia Wilson. I'm Senior Vice President of Quality Measurement and delighted to be here. Thank you.

DR. LYZENGA: All right. So we have got just a few sort of introductory items to cover. We have already introduced the committee. You've already heard Marcia say that you are acting as a proxy for the multi-stakeholder membership, but not as a representative of any particular interest group. And as we told you the last time, you'll be serving two- to three-year terms. I think we drew some names the last time here, but we won't be getting to the turnover for another year or two, I believe.

I think just to sort of reiterate what

your sort of duties as a standing committee here are, to work with NQF staff to achieve the goals of the project, evaluate the candidate measures against the measure evaluation criteria.

We will have a public comment period, as you know, and we'll get you together on a phone call after that and just go over the comments, have some -- adjudicate those comments and come up with some responses for them. And then to respond to any directions from the CSAC once we make our recommendations to them, and we'll answer any questions or anything that they have.

So, as you know, all of the members will review all of the measures, but we have assigned a few people to be discussants for each measure. And we'll ask those discussants to sort of introduce the measures after the developer speaks, and just kind of walk us through the evaluation process.

But certainly we would ask all of the committee members to jump in and add their

thoughts. It shouldn't just be the discussants discussing the measures.

We will be walking through each of the measure evaluation criteria and voting on each subcriterion. And we've got a new voting system, which Alexandra is going to walk us through a little later. We hope that that will be much better than the last time around. We know that there was some frustration with the voting buttons and everything, and we hope that this will be a little bit more efficient this time.

We will be making recommendations to the NQF membership for endorsement or not -- endorsement of these measures.

And as Lee mentioned a moment ago, we will be doing a bit of review of the sort of surgery portfolio and thinking about gaps in that portfolio, potential gaps in surgery measurement overall, where we might like to see some measurement development, if you have any ideas on potential measure concepts that could be sort of prospected for development, but we'll get more

into that conversation later today.

So, again, this -- do you want to talk through the portfolio?

MS. FELDMAN: So, very briefly, this project is to address the areas in general on specialty surgery, focusing on pre- and post-surgical care, adverse surgical outcomes, timing of prophylactic antibiotic and other related topics.

Currently, this is one of NQF's biggest portfolios of measures related to surgeries. It's one of the biggest areas of measurement. There are over 100 NQF-endorsed measures related to surgery, and 69 of those are assigned to this committee.

I'm not going to go into detail into this, because we have a -- as Andrew said, we have a discussion later today that will focus on reviewing the portfolio in further detail.

There are 24 measures that we will be reviewing over today and tomorrow, so this slide just lists those out for you in detail.

And just to speak to the activities and time -- timeline after this meeting. On March 27th, that's next Friday, we have a post-meeting webinar scheduled from 2:00 to 4:00 p.m. This will be where, if anything, that we don't have the opportunity to address over the next two days, this will be the opportunity for us to shore things up. If we are able to be very efficient over the next two days, we won't need to have this meeting.

After this meeting, and after the webinar, NQF staff will be writing the draft report. This will be posted for NQF member and public comment from April 17th to May 18th. There will be a standing committee called to review -- a call with the Steering Committee to review and respond to the comments.

The draft report will be posted to the NQF website for NQF member vote. Then, the CSAC will review. It will go to the Board for endorsement, and then the appeals process, if necessary, so -- and then, to conclude the

project, it will be finished by the end of September.

So I'll turn it back to Andrew.

DR. LYZENGA: So, yes, just a few ground rules. We will ask that you have reviewed the measures beforehand. We hope that you have. And we do ask that you base your evaluation -- try to base the discussion as much as possible in the measure evaluation criteria. We have -- those criteria are pretty carefully crafted and designed to get to the important issues here, and we try to keep the discussion focused around those criteria and not go too much outside of that, if possible.

We do ask that you attend the meetings and try to remain engaged without distractions, if possible. Keep your comments concise and focused, to the extent you can. And, you know, allow others to speak. You know, try to foster a meaningful participation and discussion. And indicate agreement if you need to, but, you know, we'll try not to repeat ourselves too much, just

in the interest of time and efficiency.

So the process for discussing each measure, we will start out by asking our measure developers for their respective measures to come up here to the table. We've got a couple of places here at the front. And they will just take a couple of minutes and introduce their measure, say a few words about them, and then we'll ask our lead discussants to get us started with the committee discussion by just giving us a quick summary of the pre-meeting evaluation comments that were provided by your colleagues on the committee, if there are any, emphasizing any particular areas of concern or differences of opinion, if those have emerged, and to just sort of walk us through each of the criteria, at which point we will vote on each.

The developers will be here and will be able to respond to your questions, as needed.

Also, just to, again, go over the criteria again very briefly. We do endorse -- NQF endorses measures for accountability

applications. That means public reporting, payment and accreditation, as well as quality improvement purposes. So when NQF endorses a measure, that implies that it is suitable for accountability purposes, including public reporting and payment programs.

We do have a standard -- set of standardized evaluation criteria. These -- you know, the quality measurement enterprise is evolving, and the criteria evolved over time in response to that. One change I think since the last time we had you here is that we have removed the high priority subcriterion of importance. We weren't finding that that was providing a lot of value.

Pretty much every measure is a high priority or can be construed that way in some sense, and that was kind of how the votes were turning out. So we decided to sort of skip over that, and we won't be voting on that particular subcriterion this time, although the developers have provided some information along those lines,

just for your information.

Here is the endorsement criteria, again, just to remind you the importance to measure and report. The goal there is to measure those aspects with greatest potential of driving improvements. Scientific acceptability, we want to make sure that these measures are making valid conclusions about quality, that you can collect the data reliably and consistently, and that there is -- you know, the results of the measure will lead to appropriate interpretations about quality.

Usability and use -- the goal here is to look at how the measure is used, how usable it is for accountability and improvement purposes. In terms of feasibility, the burden on providers in terms of data collection, the goal is ideally to cause as little burden as possible. If a measure is not feasible, to consider alternative approaches.

So we do have these voting tools here. It's a new system, and I'll ask Alexandra to come

up here and talk us through it for a moment.

MS. OGUNGBEMI: Good morning, everyone. My name is Alexandra Ogungbemi, and I am the project analyst on this surgery project, this phase.

For those committee members in the room, you all have a remote control. When you are voting, you will point towards me. And I'm over here by the windows, on the east side of the building. And that's when you will make your selection to vote.

Once we get a solid number of votes during each voting slide, upon the discretion of our chairs, we will move on to the next criteria or the next vote. I will actually also act as the proxy member for those who are not in the room. So Dr. Sawin, Dr. Asher, and Dr. Jarrett, when they do join, I will act as the voting member for them. And if you have any questions, please let me know.

DR. LYZENGA: Thanks, Alexandra. So, yes, just let us know if you have any questions.

We'll probably -- can we do a test run maybe on the first one? I don't know. We'll see how it goes. Hopefully, again, it will be a better process than the last time around.

So, yes, we just wanted to note also that we have a few measures that we identified as related to measures that are being considered today. And prior to the meeting, we notified the developers if a related or competing measure had been identified. We asked them to consider how they could work together with the other developer of the measure to harmonize or otherwise align those measures in terms of definitions and other elements.

The committee is invited to ask the developers about those harmonization opportunities when the measure is being discussed. And we've got, I think, a bit of time after the discussion of each measure to -- each measure in which that harmonization discussion is relevant to talk over those little -- issues a little bit after we have discussed the measure.

For the gaps discussion -- and, again, we'll go over this a little bit more later, we are thinking that we would like to assign each of you a topic area based on your expertise and area of focus. We will certainly -- we would appreciate your feedback on what you would like to -- sort of what topic area you would like to cover in terms of identifying gaps. But we've sort of made some tentative assignments, and I think we may distribute that to you today.

We will ask, moving forward, that you -- we will have some exercises that we will go through to try to identify potential measure concepts that might identify any gaps in measurement in your given topic area. Again, we will walk through this a little bit later at 4:00 today. But we just wanted to kind of mention that up front.

And now I think we can jump into the evaluation process. We are going to start out with a measure from AHRQ, so we would ask the developers to come up to the table at this point.

DR. FLEISHER: Okay. I'd like -- before we start that discussion, I'd like two things. I think Marcia needs to have Cliff introduced and then, secondarily, I'd like to have a brief discussion about types of evidence.

MS. WILSON: Okay. Thank you, Lee.

Dr. Cliff Ko has joined the meeting here in D.C., and, Dr. Ko, we have been doing oral disclosures of any professional activities that may be relevant to the subject matter before this committee. So I would ask that you give us your name, your organization and if you have anything to disclose, please.

DR. KO: Hi. Good morning. My name is Clifford Ko. I work at -- I'm a professor of surgery at UCLA, and I work at the American College of Surgeons and run their Division of Quality.

MS. WILSON: Thank you. And I will also take a moment -- has Dr. Jarrett joined us on the phone to do an oral disclosure? Or Dr. Asher? Thank you.

DR. FLEISHER: One of the things we thought would be useful this morning is we do have measures that are administrative, we have measures that are clinical, and we have registry measures. And maybe Marcia can give us a brief overview of how NQF looks at this. I can comment -- no?

Just from the perspective of CSAC, the question in the current space, things are in flux. We are not at the point where we can have e-measures. And although we acknowledge that ideally robust clinical data to inform measurement would be the ideal once we get to e-measures, and some are becoming e-measures, that they may provide complementary information.

So it is from a hierarchical standpoint, and I don't mean that from an analysis standpoint. But they are all useful, and that -- if we feel they truly should be harmonized, we can have that discussion. But if we feel that an administrative measure, such as something developed by AHRQ or Yale CORE,

complements a registry measure, for example, by STS, then both can be endorsed.

So I would like to open that up for any discussion, because we thought that should actually be addressed at the beginning. If anyone has any thoughts on these issues. And I think as we go around, it is probably easiest -- tradition at NQF is just to put your name plaque up, and, therefore, we can call --

MS. McCARTY: Can you just clarify what an e-measure is?

DR. FLEISHER: An electronic measure, just something that would come from an electronic health record.

MS. McCARTY: Oh, I see. Okay.

DR. FLEISHER: There are separate groups where --

DR. DUTTON: I can help with that. We are grappling with this quite a bit in our registry now, because we are -- we have a high penetration of electronic records, and we are right in the evolution. So, for instance,

Kelsey, we can measure post-operative nausea and vomiting by asking the PACU nurse to check a box. Did this patient have nausea and vomiting? Or we can get the same or similar information by looking at the pharmacy records to see if an antiemetic was dispensed.

You get different answers from those two. They are not perfectly aligned. But they are the same spirit and the same concept.

And I agree. I actually was putting my thing up to agree with Lee that both can coexist for now.

DR. FLEISHER: Right. That may change over the course of our time here.

Any other comments? Thoughts on this topic? Okay. Bill, did you want to say anything?

DR. GUNNAR: This is Bill Gunnar. I have no additional comments to make before we get started. Let's get into it.

DR. LYZENGA: If we could just have our measure developers introduce themselves

briefly. Patrick?

MS. STARKS: Hi. My name is Carol Starks, Agency for Healthcare Research and Quality. And I'm the task lead for the quality indicators that -- the inpatient quality indicators and the prevention of quality indicators.

We have a -- our current contract is with Stanford University, and Patrick Romano is an important part of that from UC Davis.

DR. ROMANO: So my name is Patrick Romano. I think I've met many of you before. I am a practicing general internist based at UC Davis School of Medicine, Sacramento, California. And we search as subcontractors to Stanford on the enhancement of the AHRQ quality indicators.

DR. FLEISHER: So if you'd like to just very briefly introduce the first measure here, which I believe is -- we're talking about hip fracture mortality first. So it's Measure 354.

DR. ROMANO: Thank you. So, yes, so

this measure is one of a family of measures that most of you are probably at least somewhat familiar with. So these are the AHRQ quality indicators, which are based on administrative data that are collected by state health data organizations and compiled by the Agency for Healthcare Research and Quality, made available to researchers and others.

They are also used extensively by state health departments, by regional coalitions, researchers, and others for a variety of purposes.

The particular measures that -- the first measure that we're talking about is part of the module called the inpatient quality indicators. This module focuses on hospital outcome measures and structural measures of hospital care that are related to outcomes.

The IQI for hip fracture mortality, specifically, is one of a subset of these IQIs that focus on inpatient mortality for patients who undergo certain common procedures in acute

care hospitals. So the focus of these measures is on inpatient mortality. The reason for that is that many of our users only have access to inpatient data. Ideally, they might like to have data on post-discharge follow-up of patients, but a hospital's ability to collect that information obviously is limited.

So these indicators were designed to focus on inpatient mortality and to include some fairly sophisticated risk adjustment to account for variation in severity of illness across hospitals. There are a number of these mortality IQIs that I think are in the domain of this committee. I think the only one that is under review today is the one focusing on hip fracture mortality. But I'll stop there.

DR. FLEISHER: Thanks, Dr. Romano.

And I think we have a couple of discussants on this. Dr. Ko and Dr. Yates are discussants for this measure. So can you kind of get us started with the discussion?

DR. KO: Sure. Good morning, again.

Can I -- I'm trying to use your template for doing this, so -- but keep -- it was a little complex, so can you tell me -- I know we have to go a certain way and then put it up for a vote. So can you help me with the first area until we -- to present until we vote?

DR. LYZENGA: Sure. So, I mean, we will start out with -- evidence is the first subcriterion for the importance to measure and report criterion. And this is an outcome measure, so I think we discussed the last time around that for an outcome measure we do not require the same sort of volume or type of evidence as we do for a process measure.

For a process measure, we ask that the evidence be based on a guideline, a systematic review of the evidence. We ask for some information on the quality, quantity and consistency of that evidence supporting that particular process of care.

For an outcome measure, really what we are asking for is that there is a plausible

rationale connecting at least one health care structure, process or intervention to the outcome, that there is some sort of rationale for a linkage there, that providers can influence the outcome in question.

So, again, we are not looking for a sort of systematic review of the evidence there. Just a sort of justification of sorts, a rationale for that connection between processes and the outcome in question. So that is sort of the question at hand here, at least first.

DR. KO: Perfect. So, again, this is Measure 0354, and it's entitled the Hip Fracture Mortality Rate, IQI 19. It's -- the steward is AHRQ, and you just met them. The description of the measure -- it's an in-hospital death per 1,000 hospital discharges with hip fractures, the principal diagnosis for patients 65 and older.

The rationale is that providers can adopt processes of care -- of best performers, and consumers can select the best-performing providers in order to reduce the overall

mortality rate.

Again, the data source is administrative. The level of analysis is facility.

In terms of evidence, this is an outcomes measure. And in terms of processes of care that can influence the outcome, the number one process is probably time to surgery, when somebody has a hip fracture, when they are brought to the operating room. In working with the orthopods, the AAOS, and other of the orthopedic societies, they all agree that this is an important issue, although some have -- there is still a little variability as agreement in terms of time of surgery, but most believe that time to surgery is the process that can influence that outcome.

There are a number of other issues in terms of influencing the outcome in terms of rescuing from a complication with PE and MI as the big complications occurring post-op or peri-op during a hip fracture case. And so there

are other processes as well.

DR. YATES: I am the co-discussant. This is Adolph Yates speaking. If I'm not mistaken, however, there is nothing in the measure that captures time to OR. Am I correct in that?

This is data that is extracted from an administrative -- from administrative data sets or codes. And my one comment is, is that it is a fairly blunt measurement. Without a doubt, there is a process of intervention, which is somebody comes to the hospital with a hip fracture, and ideally they leave with the ability to heal and eventually walk. And death would be an outcome.

So I -- there is no question that there is a process here. The problem is, is that there is a black box in this process, and the black box is, is that the numerator and denominator are only defined by the level of the fracture, i.e. whether or not it was attributed to being or assigned to being, for instance, a femoral neck fracture versus a subtrochanteric

fracture. And the other parameter is age.

And so you really only have a couple different variables with which to both assess risk and both to assess the nature of the baseline population. So my concerns or my questions with this is, yes, it is an outcomes measure. But when you look at page 2, the percentile of the distribution of outcomes of deaths, in the fifth percentile, in the 25th percentile are zero.

So it seems to me that there is -- it is a strange distribution of the curve here. This measure includes pediatric hospitals, it includes orthopedic-specific or specialty hospitals, which don't have emergency rooms frequently. It includes a whole slew of hospitals that may never see a hip fracture. So that's the one question I have.

And the other question I have --

DR. FLEISHER: Can I --

DR. YATES: Yes.

DR. FLEISHER: -- just hold because

those -- Marcia, if I'm not mistaken, and Melinda, that is not the evidence criteria, correct? You are moving on to --

DR. YATES: Discussion.

DR. FLEISHER: -- how it's defined. Do we want to go and --

DR. YATES: I'm sorry. I moved to discussion.

DR. LYZENGA: Yes. Typically, we like to try to kind of keep the discussion focused around the criteria that we are, you know, speaking about at the moment. It's okay to, you know, give you --

DR. YATES: Well, what I'm getting at is the numerator and denominator. And the numerator and denominator are -- I have questions about that.

DR. LYZENGA: I think that is probably in scientific acceptability. I would say that the specifications go under validity and reliability.

DR. YATES: And so that's at the very

beginning, I believe.

DR. LYZENGA: Importance. So evidence is at the very beginning.

DR. YATES: Right. And I agree it's outcomes.

DR. LYZENGA: Okay. Great.

DR. FLEISHER: As we go through this -- and this is great, because we will -- we have to get back into that theme. I think we should vote on evidence, and then we can quickly get to those very important points.

DR. GUNNAR: Just to go back and, Cliff and Adolph, can you just sort of sum up your advice or your perspective on evidence?

DR. KO: As far as the opportunity for mortality itself to be a metric for hip fracture, the evidence shows that there are processes that are linked, although, you know, with what Dr. Yates said, that should be understood in that context. But it's mortality itself that -- the process of time to surgery is thought to be an important process where the outcome may be

mutable.

DR. YATES: And I agree that there is something to be measured here and that it has evidence. I would just say that -- and this is why I slipped into that slope of discussion. There is no time of outcome measured in this measure, just for clarification. That's why I started talking.

DR. FLEISHER: No. This is great. Just want to --

DR. GUNNAR: And just one other -- just to point out this is in-hospital mortality as opposed to a defined period of time, which we will discuss tomorrow in some depth?

DR. YATES: Correct. In-house mortality is the discharge to diagnosis.

DR. EREKSON: So, and I don't know if this falls into reliability or if it falls into evidence. But one big glaring thing on this measure is patient preference, and I don't know if that's in this discussion, which is evidence, or if that should be an exclusion of this

measure. That if a patient is 102 and falls, do they choose not to have surgery? And does that affect your mortality?

MS. JOHNSON: I think that would be discussed in two different places, probably in the specifications, and then also under validity when you think about exclusions to the measure. So you have two opportunities to talk about that.

DR. LYZENGA: So if there is -- go ahead, Patrick.

DR. ROMANO: I just wanted to point out that -- so the form asks developers to focus on at least one process measure that demonstrates this link. There are of course others. There is 25 years of literature on preventing thrombosis after hip surgery using both mechanical and pharmacologic measures. There is also literature related to cardiac evaluation, cardiac risk assessment, and prevention of post-operative myocardial infarction.

So those weren't highlighted in the literature review, but those are also part of our

literature.

DR. LYZENGA: All right. Well, if there are -- oh, go ahead. Dr. Cima.

DR. CIMA: I think to follow up on Dr. Romano's point is that about a time to OR, that -- I mean, that may not be the best measure because then you're going to have to get into, well, is it reasonable to say it's 72 hours, but what if a patient needs an extensive medical evaluation before -- to make it safer to go to the OR. I mean -- so, I mean, I think if you're going to do a mortality measure, you sort of do the risk adjustment, but you don't start putting in process measures in the middle of it.

DR. YATES: We are getting into the scientific discussion about the measure itself. But, again, this measure does not capture whether or not DVT prophylaxis was performed, whether or not there was a time to OR issue. This is strictly mortality, and the numerator and denominator are the type of fractures and the patient's age across a wide spread of hospitals.

So that is what the measure is measuring, but those other things are the processes that could justify the measure. And I would agree with the last comment in that it's -- all of the papers that show 48 hours being some sort of a magic time period for getting someone to the operating room and repaired, all of those exclude those patients that have significant, reversible, other medical issues that could be improved upon before they have surgery.

So that is an important point, but that's going on into the scientific validity.

DR. SIPERSTEIN: So just a quick comment that one of the values of this type of measure obviously is for the institution itself to know what their own rates are, so they can then do an internal reflection in terms of, are there processes in line in terms of improving it. And time to OR, pre-op evaluation, all of these factors, may play a role. But we -- I see the value primarily as the institution knowing their rate, being able to benchmark, and then figure

out whether they have an optimum process in place.

DR. FLEISHER: So, John, do you have a comment about the evidence? Because what I would like to do -- this is a review of an approved measure, correct? This is a currently established measure. So --

DR. HANDY: Well, I would just -- as I recall from my reading of this, it's a small proportion of patients. But transferring it to another hospital is an exclusion, and that's the way you can take that patient and move him downstream. And so that's going to be a description of the measure versus the evidence.

DR. FLEISHER: So can we vote on evidence, and then we can just --

DR. LYZENGA: These are really important points. So maybe we can use this as a little test case, go through this and see how it works. Do you want to take us through it, Alexandra? All right.

So I'll read it off. All right. So,

again, just to remind you, we are voting on whether a rationale supports the relationship of the health outcome, in this case hip fracture after surgery, to at least one health care structure, process, intervention or service.

One indicates yes, two indicates no, and we'll start voting now. So tell us -- Alexandra, how many is that? How many have voted?

MS. OGUNGBEMI: Yes.

DR. LYZENGA: And what's in the right corner? Oh, it's the timer. You're not timing this, so -- it looks like we've got 22 votes. I think that's it.

All right. So clear result there, and we can move on to -- scientific acceptability is next, or actually -- sorry, we've got more in importance to measure and report yet. Can you skip to the next slide, Alexandra?

So performance gap here. That's the next point of discussion, whether data demonstrate a variation or overall less-than-

optimal performance across providers and/or populations and groups. Basically, what we're trying to get to here is whether there is a gap in care, an opportunity for improvement, that the measure is not so to speak topped out or otherwise unable to achieve improvements in performance.

Is there any discussion on this? Go ahead.

DR. YATES: To make up for the last section, I would say yes. That's easy.

(Laughter.)

DR. LYZENGA: Anything else from the committee on this, or should we go ahead and vote? Yes, let's go to the vote. Alexandra? You can go ahead and start.

Can we see the vote? Okay. Yes. So it passes.

And then we'll move on, and, again, this is -- go ahead, Karen. Good point. Can you back to the last slide, Alexandra? Just for the purposes of the transcript I'll read off the

vote. Can we do that? Oh, we can't. All right. Well, we'll go back and correct that later. Seventy-nine percent high? All right.

So, again, we're not actually voting on high priority, so we can go ahead and skip over this one. And we're not looking at a composite now, so we can skip that one as well.

So now we're at scientific acceptability, and this does include the specifications, whether the measure is specified precisely, whether it has been tested appropriately and with adequate results to demonstrate reliability and validity.

So we'll -- I'll hand it over to the lead discussants to talk over this issue.

DR. KO: So this measure is an administrative data measure. The numerator and denominators are specified. Numerator is the number of deaths among cases meeting inclusion and exclusion rules. The denominator is discharges of patients 65 and older, principal diagnosis with ICD codes with a diagnosis of hip

fracture from their evidence.

The hip fractures, there are 350,000 discharges, and 300,000 or so of them have an ICD of -- primary diagnosis of hip fracture.

The exclusions are -- well, I won't go through that, but the -- you can see what those are. The fact that it's an inpatient mortality measure means that -- obviously, that this is just the inpatient aspects of mortality and cannot get any longer or post-discharge mortality rates. And so that may be a concern. I know that clinically a number of people have brought that up.

Do I get into -- is reliability testing in this section as well?

DR. LYZENGA: Yes.

DR. KO: So with this data and with the -- with the data and the analysis that was performed, the reliability testing, the reliability of distinction testing that was performed with this measure has been very good and very scientifically sound. And you can see

that -- from Table 2 that the cut point of .4 as a reliability is marked there. I mean, they looked at it very scientifically and methodologically well.

It looks like the deciles that meet the .4, however, is the eighth, ninth, and tenth deciles, and the first seven deciles do not. So that is an issue with this measure. That is probably an issue with a lot of measures in terms of looking at reliability to that degree. I think that there are a number of measures, just as a comment, in -- that are evaluated by the NQF that do not look at reliability at all.

So the fact that they did is a plus. And I think that their numbers are -- their findings are probably in line with a lot of the measures that we have in the endorsed category.

So I will stop here, unless there is another section in this that I'm missing.

The validity, is that in this area?

DR. LYZENGA: We will go to validity next. They are two separate votes, actually.

DR. YATES: The issue of reliability is I think open to question, in that it is -- it may be that they are able to demonstrate mathematically that there is reliability. But the -- not to belabor the point, it is very difficult for even orthopedic surgeons to decide what level of fracture they are dealing with. So there may be several different things on the margin here that wouldn't be called "a hip fracture." Proximal femoral fractures, and the like, sometimes are lumped into that category.

But aside from that, there are some reliability issues in that I'm not sure that the lower percentiles showing -- or the higher performing percentiles showing zero hip fractures indicates that they are capturing the right hospitals in that percentage. There may be -- I mean, I don't understand for a measure that's looking at patients 65 and older why they would include pediatric hospitals. I don't -- I mean, they are obviously going to have zero.

So I worry -- it would be nice to see

a graphical distribution of that occurrence across those hospitals. I assume that it's going to be skewed. Likewise, in the reliability portion, they include as part of the measure open fractures. It takes an incredible amount of energy for someone to have an open -- femoral neck, open intertroch, or open subtroch fracture, and those would represent patients that might be, at 65 and over, in a trauma hospital.

And I don't believe that's the intention of this measure, to look at high energy trauma patients as opposed to, say, grandma and grandpa with osteoporosis falling and having a hip fracture in the community.

So, yes, you can -- the reliability is mathematically correct, but the reliability of what we are looking at I'd just call into question from just what the numerators and denominators are capturing.

DR. LYZENGA: Amy?

MS. MOYER: I believe the inclusion of the pediatric hospitals is somewhat related to

the feasibility. You don't have to then take the data set and look at, is this a pediatric hospital, is this an adult hospital. But then the removal of all patients under age 65 effectively takes those out -- those hospitals out of the measure. So it does remove them, because they don't qualify for the measure in the denominator.

DR. LYZENGA: Collette?

MS. PITZEN: Great. I have two comments. One is in general about mortality measures, and a consideration for having a 30-day mortality rate. And I think we will be getting into that more as time goes on.

But just, for example, if you have someone that is discharged after 120 days, that has a different kind of a case than perhaps a 30-day mortality rate.

And the second comment I'd like to make is about the reliability score. I applaud the measure developers for providing that performance score. I think that's an important

value, and many of the measures that we are looking at over the next couple of days do not have that performance score.

I am a little bit concerned about at .4. We publicly report all the clinics in the State of Minnesota, and our reliability scores for doing those comparisons clinic to clinic are in the .8 and .9 range. We like to see something at .7 or above, and we start to get concerned when a reliability score starts dipping down below that .7.

Thanks.

DR. FLEISHER: Rick, and then we'll get responses.

DR. DUTTON: Yes. A couple of questions for the developers. As Dr. Yates mentioned, there may be a hard time distinguishing signal from noise in this measure. The distinction between high energy and low energy fractures is one. Another one brought up was patients who have a fracture but choose not to have treatment, for instance, end-stage

Alzheimer's patients who might become DNR, or the final -- be transferred to hospice or a skilled nursing facility before dying.

Do you have any comment about how we might improve the measure to better discriminate the population we are trying to get at?

DR. FLEISHER: Comments from the developer?

DR. ROMANO: Sure. Okay. Let me -- I will try to hit them all, but we'll see. So, first, with reference to pediatric hospitals, yes, they're in the source data set. They're in the original data set, but they are effectively excluded. So the 2,72 1 hospitals are basically the hospitals that had at least three patients who were eligible, patients who had hip fractures who were over 65.

So the original number of hospitals in the data asset is close to 4,000, right? Do you know, Carol? Over 4,000, right? So 2,721 is a subset of those.

Second is in terms of the definition,

yes, this measure was deliberately specified in consultation with an expert panel that included specialists from a variety of relevant disciplines. It was specified to include both patients who are managed surgically and patients who are managed medically. And that was deliberate because of the fact that some patients may opt for non-surgical management.

There may be some particular contraindications to surgical management, and the type of management in this case is actually included in the risk adjustment. So that there are -- the risk adjustment model is actually quite fully specified with the C statistic of 0.893.

For those of you who are familiar with risk adjustment models, that's a very high discrimination, and it reflects the fact that the model incorporates the type of procedure that was necessary, whether the fracture was open or closed, and whether the patient was treated medically or surgically. So that's incorporated

into the risk adjustment. We can talk about the pros and cons of doing that, but that was the decision that was made in consultation with the expert panel.

In terms of reliability scores, yes, this is a problem across I think a wide panoply of measures that are in the NQF portfolio. And we do want to be honest about this, and sort of recognize that this measure is not going to be reliable. Most procedural mortality measures are not reliable for hospitals that are in the lower part of the volume distribution.

So the way that we account for this methodologically is to do what is called smoothing, and this is basically the same approach that Yale CORE uses for CMS measures, so that the risk-adjusted rates for these lower volume hospitals are shrunken or smoothed back to the overall mean.

Again, this method has certain strengths and limitations that we can discuss, but what it basically means is that for low

volume hospitals the publicly reported metric -- for those who choose to publicly report, the publicly reported metric will basically look very similar to the average.

As hospitals move into higher volume categories, then the hospitals' own experience becomes the primary driver of the reported metric, the smooth metric. So we in fact adjust for the difference in reliability that you see across the volume deciles in the smoothing process.

So basically we recommend that when hospitals are using the data themselves, or within their organizations, that they should focus on the unsmoothed rates to reflect their own experience. For public reporting applications, we generally recommend use of the smoothed rate to account for the variable reliability across volume thresholds.

Another approach to this problem, of course, would be to set a volume threshold, a single arbitrary number, and to say that this

measure should not be recorded if your volume is less than X. That's an approach that we have not taken in the AHRQ quality indicators program, sort of leaving it up to users, presenting the data, and allowing them to make their own decisions in their own context about what the right threshold is. But that would be an alternative approach would be to dictate or to suggest a single threshold for volume.

So, and then, finally, so in terms of patient preference, yes, we certainly recognize that this is an important issue. I will point out that age is a powerful factor in the risk adjustment model. So, for example, patients who are over age 85 have 2.4 times higher odds of mortality, as you would expect. We also find that hospitals that receive patients transferred in from other hospitals, there is a factor in the risk adjustment model to account for that. And so those patients also have somewhat higher mortality.

So as far as -- we don't have any way

of knowing specifically what the patient's values were, one thing I will say that we are also exploring analytically, but it's not reflected in the current specifications, is using information about hospice enrollment as a proxy for patients who have chosen palliative care as their approach. It is consistent with their values.

We still have some uncertainties about our ability to capture that information accurately across all payers. These are intended as all payer measures. So it's not reflected in the current specification, but it is a topic of ongoing analysis.

DR. FLEISHER: Thank you. I don't think you addressed the -- well, Larissa, a comment?

DR. TEMPLE: I just have a point of clarification, and this speaks to sort of not being a hard-core methodologist. But when you talk about your smooth rates, that includes the risk adjustment or does that not?

DR. ROMANO: Yes. It starts with the

risk adjustment, correct.

DR. TEMPLE: Okay.

DR. FLEISHER: Any comments back? Cliff?

DR. KO: I have maybe a clarification from you or from Bill, that when -- data source is clearly an important issue, and this is a billing data source. And there is pros and cons, and we all know what those are.

But how should we look at this? In what perspective? So as a methodologist, you know, if I take a data source and I'm like, well, there are good things about it, and there is not such good things about it, and I just like, okay, we just acknowledge that, and then we go forth with that data source, and, you know, AHRQ and Pat has done a great job in doing the methods with that piece.

But if there are issues like what Dr. Yates brought up that there are clinical issues that are just not there, or there are risk adjustment issues that you cannot address because

you can't tell the level of fracture or patient preference, and things like that, how should we look at this in totality?

Because as a methodologist doing the best you can with what you have, but if you were working in the real world it misses a lot of things that are important to this topic of, in this case, hip fracture mortality. So the perspective of the way we look at this is very important as to whether we think this is going to cut it or not.

DR. FLEISHER: I think you actually outlined the question that you -- this committee has to decide from a reliability standpoint. I mean, this is -- we can't change the measure unless we turn it down. I mean, we have an up or down. They can respond to questions, and I go back to either Andrew or Karen for further comments. If you feel that the reliability is not high enough with this data set, at this point in time that's a decision you make as a member of this committee and pass it over to CSAC.

Other thoughts? So I -- you know, it would be great to have the ideal patient preferences. We have actually published on the fact that there are racial differences that I'm not sure are patient preference differences based upon work by Neumann's paper. So, I mean, this may not be preferences, and this would be important data that can be found in this measure.

So I think the answer to your question is you have to decide if you think it's good enough, as currently stated. And if there's a better measure -- this gets back to the harmonization. If there's a better measure, could be unintended consequences of using this data are too great and there's something better out there, then you should I think vote for the better measure. If there's nothing else to compete, then the question is, is it good enough?

DR. YATES: We're going to move on to validity separately.

DR. KO: Yes. That will be the next one.

DR. YATES: Or are we going to vote on reliability first?

DR. KO: We'll vote on reliability first.

DR. YATES: Okay.

DR. GUNNAR: But I think specific to what -- and Cliff's question is a good one -- is, how do we evaluate this from a perspective of, is it as designed have an underpinning of reliable information that allows it to then have the observed outcome?

And the answer, that it's easier I think in a measure that has been around since 2008, which this has, if the data has been treated essentially the same from its creation, and we don't believe the world has developed an enormous workaround, that there is improvement. That's what they're showing, that the observed rate per hundred has improved over the last -- when you look at the totality of hospitals that they're acquiring information from.

I think the questions -- or the

granular questions, then, raised are, what are people doing in response to this? Are they reacting to this measure and modifying their behavior to the good of the outcome, or are they modifying it to the good of "I can find a way of not being seen or that mortality being counted." So I think that gets to the reliability and the validity. I hope I hit that on the head.

DR. YATES: Well, and along those lines, one big wraparound gets back to Collette's point, which is you'd have to know what the length of stay was in that -- in terms of knowing whether or not there was really a reduction in mortality or was the mortality outsourced, if you will, to the skilled nursing facility, which that would be a trend that is also very much concurrent with length of stays all dropping.

DR. FLEISHER: John, and then Barb.

DR. HANDY: I mean, while Cliff's point is incredibly dead on, almost nobody else has a resource that doesn't involve or that involves a very detailed clinical database,

because almost -- most measure sponsors are using administrative data. So it's a really important point, but you would say that we, therefore, will only go with somebody that has a very extensive clinical database to be able to overcome those points, but --

DR. LEVY: And, more importantly, we don't know that mortality isn't the outcome that the patients wanted. I mean, to -- to the earlier point, this may not be improvement, that we are discharging these folks alive. So, you know, this is a philosophical point that we are going to need to talk about. But the fact that mortality is a measure of quality may or may not be an accurate assumption on our part, and we don't have an opportunity to know what the outcome was that these patients really wanted to have.

So I'm not sure that the assumption that we've had improvement in mortality is a workaround or is an improvement in process, but it may actually be detrimental to the true

outcomes that we're looking for.

DR. LYZENGA: I just want to note that a lot of these issues I think are -- kind of relate more to validity. So just for voting purposes, again, the first vote we're going to take is on reliability really, while the information can be collected in a sort of reliable way to sort of form a reliable foundation for the measure results, again. And then validity will really be talking about the -- whether it is a true reflection of quality. So maybe we should vote on reliability at this point, so we can kind of get into some of these validity issues?

DR. ROMANO: One final response, which I just got by email from another member of our team. So just to clarify that -- so when we are looking at these measures of reliability, there are several different measures out there. And so it's a little bit caveat emptor to some extent.

So just to be clear, the measure that we use is a measure of signal to noise, which

basically compares the between hospital variance to the total variance, including both between and within hospital variance. So this is a measure that focuses on whether there is a signal that hospitals can be identified as having a higher or lower than expected mortality.

Now, this is a little bit different. Some other measure developers use a test/retest measure, which is simply looking at whether there is consistency and performance over time. And we are actually doing some comparative analyses right now to understand the relative performance of these two different approaches.

But in these types of outcome measure applications, signal to noise metrics generally give you lower reliability numbers than test/retest measures. In other words, it's easier to show that the performance of a hospital is consistent over time than it is to show that it's statistically distinguishable from the performance of other hospitals. So that's just --

DR. FLEISHER: Thank you.

So, Alexandra, you ready?

DR. LYZENGA: All right. So now we're voting on reliability. Is the measure precisely specified and tested with an appropriate method and scope with adequate results?

We have 21. Last second to try voting again. We have 22. Alexandra, can you -- Andrew, are you reading or -- who is leading up the -- lead it off. Go ahead. Do we have the results? So we have 39 percent high, 59 percent moderate -- sorry, 57 percent moderate, four percent low, zero insufficient.

And now we can go ahead and jump into the validity questions, again, sort of really more focused around whether this -- the measure results reflect quality of care and whether valid conclusions about quality can be drawn from the measure results.

DR. FLEISHER: So Barbara's comment really feeds into a lot of questions that will come up tomorrow again with STS, and we wanted

STS also to be here, so we'll have a brief discussion about that, which we will get to STS to be prepared to have a more full discussion, but it is something that has obviously been commented on, including in The New York Times recently.

Cliff?

DR. KO: Well, I will just go through it. A lot of the testing they had done was on face validity. They performed a panel, like a RAND panel, of 14 clinicians, and basically they had acceptable indeterminate agreement of overall usefulness rating as a quality improvement metric, overall usefulness rating as a comparative reporting metric.

The issue that was brought up previously about the caveats of indicator use suggested by the panel is exactly the length, which is use of 30-day mortality measure would offer additional information and reduce the bias. But overall they thought that -- from this panel that there was high face validity.

DR. LYZENGA: A.J.?

DR. YATES: Thanks. The issue of validity is probably more than anything else what drives the discussion we will have on usability and feasibility and reportability, because the reliability is something that most of the developers can provide mathematically through statistics and the like. I think that's a given.

And the issue on validity is that unfortunately we can't vote three different ways. I think this is a very valid study in terms -- or valid measure for national trends. I think this is a very valid measure for individual hospitals to look inside at themselves and see what they can do for improvement.

But in terms of the validity of this measure to measure the quality of one hospital versus another, I don't think that the risk adjustments, which I see as being offered as being -- as being level of fracture, type of fracture, and also age, and transfer in or out, is enough to truly decide that one hospital is

more highly -- has higher quality than another.

There are -- what I don't see here is within the black box out of the administrative data set codes that could have been captured is any indication of comorbidities or other things that are important. There will be community -- there will be specialty orthopedic hospitals that might take a transfer of a hip fracture that wants to have a total hip replacement that is in perfect health, an ASA-1 at age 66. That's a different population than, say, a hospital that has a large oncology and cardiac population, such as ours, where I may operate on four ASA-4s in a row, and trying to get them better and get them out of the hospital.

And I don't see where there is any risk adjustment for those type of issues. The validity can be adjudged in terms of, yes, it -- death is a -- death is a great dichotomous 01 thing, and I see that as being very valid.

But I would say that in terms of the risk adjustment, for the purposes of comparing

hospitals' quality, I would be hard-pressed to say that I can say it's valid.

DR. FLEISHER: Thank you. Before we go forward to others, just from the perspective -- and I'd like Andrew's perspective -- if we approve this, we should also be providing the developers with our concerns for the next time this comes back. If we don't approve it, we should also be providing the developers with the key things that we need fixed if they come back, to bring it back.

So it will be important that I assume, Patrick, you are also taking notes of what are our greatest concerns going forward.

DR. YATES: And, Lee, just to add to that, you are making public comment on those that might use such a measure as to whether or not the NQF has concerns that can be expressed in this forum for the use of this, which can be something that can be used for payment penalties to hospitals and also public reporting. And so, on that note, I want to make sure that when we say

something is valid, it may not be valid for certain uses. We don't have the ability to distinguish that in this process in terms of voting.

DR. FLEISHER: So let me let Marcia quickly comment, because CSAC and the board are wrestling with this made for -- measure fit for purpose. So do you want to comment?

MS. WILSON: Just very briefly. This obviously is of great concern to many stakeholders. And what NQF is doing is going to be looking at intended use, because right now it's the global language that we use as the measure can be used for accountability and quality improvement, and that has been our language for a very long time.

So, yes, we are aware of this problem. We are going to be looking at this. I don't think this is an issue we will solve quickly or in the near term, but it's very much on our radar and we'll be bringing together a group to talk about intended use.

Unfortunately, for today we are operating under the existing language, which is the point that you raise. So that's how we're operating today, but it is on our radar and we will be looking at the issue.

DR. FLEISHER: And just, Andrew, in the report, will there be comments about whether or not a hospital internally, whether the committee felt strongly one way or the other about internal quality improvement versus external? That will be in the report?

DR. LYZENGA: We can certainly reflect that in the discussion in the report, yes.

DR. FLEISHER: But thank you for bringing that up. That will be important.

Collette and --

MS. PITZEN: Just a couple of comments. With an administrative claims-based type measure, one could construct a measure that is a 30-day mortality rate, as I talked about before. And perhaps setting your location would not matter as much if you're using that claims

data. So that would be one recommendation to make the measure stronger.

And I just want to share some thoughts about how I feel about mortality, perioperative mortality measures in general. I think they do have their place in terms of kind of a monitoring function. Are things going okay? Are things perhaps getting out of hand? And when I'm looking at the various measures coming through here, I want to understand, is this a fairly high volume procedure that makes a difference to a lot of people, or are we looking at something that is really rare and hard to measure anyway?

And then, is there any potential for a small improvement on that rate? Has it been demonstrated over time? Or are we looking at something that now has an incidence of .28 percent in the population?

DR. GUNNAR: So that is a really great question. If you look at the numbers, if you say -- if you assume that we have now improved by half a patient per hundred, and 200,000 cases a

year, or episodes, you've got 1,000 people per year who have benefited from this measurement just if it's the same measure and the same data and no one has done a workaround in the last, what, five years that they have measured it.

So in and of itself, to your point exactly, I think they -- when looking back, you actually can say that it has -- mortality as a broad indicator of quality has resulted in 1,000 patients this year that are alive that wouldn't have been alive before. If I've got the numbers right.

Dr. Romano, do I have the numbers correct? About?

DR. ROMANO: We wouldn't take any credit for that, but --

(Laughter.)

DR. GUNNAR: But if I'm reading your evidence correctly, that's --

DR. FLEISHER: We can get back as -- Rick?

DR. DUTTON: Yes. A quick comment on

the mortality thing. This is not a surgical measure. This is a disease measure of a patient with a hip fracture that may be managed with surgery or not. And to Barbara's point, I am much fonder of mortality measures where the patient has chosen to have an operation. So if they are presenting for a CABG, they have made a decision already, as opposed to you don't decide to have a hip fracture; it just happens. And you may have patients who don't want surgery, who want to be DNR at that point.

DR. SAIGAL: Two points. I was definitely struck by Barbara's comments as well. I actually looked into this while we were sitting here, and there is a survey of older men and women, and 80 percent of them preferred to die rather than be discharged from a hospital with a hip fracture to a nursing home and lose independence.

So that may not be, you know, representative of everyone in this country, of course, but it may be a significant thing to

think about.

And the other question I had was about Dr. Yates' comment about the open fractures. Maybe the way to handle that would be to risk adjust for that and include open fracture as a mediating fracture in a model, in terms of mortality.

DR. FLEISHER: Clifford, did you --

DR. KO: I wanted to ask Pat, and maybe the rest, that if -- if we know if the reason for the improvement. I know you don't take credit for it, but do we know? Because I know a lot of institutions that I know I visited, they decreased their mortality rate by transferring everyone to hospice. We definitely did that, if we needed to.

So, I mean, is this something going on here? Because I'm sure some of our patients that were on that track were transferred to hospice. That's the first thing.

The second thing is, the issue that Dr. Yates brought up about the accountability and

quality improvement measure, and whether this is for, you know, reporting or for tiering for payment, because then, you know, if we're -- if we're grading on a pass/fail and a C minus is a pass, you know, you have a medical student who gets a C minus, they pass, they graduate.

But, you know, sometimes we have that cut point as an A. We want that cut point to be between an A minus and a B plus, and we just want A level care. And so how -- when we think about accountability, how should we think about that as we vote, not just for this measure but for all measures? Because it's a little different how we look at that.

DR. FLEISHER: That is part of what we're wrestling a lot at CSAC, and I assume the same questions are at the boards. And we have talked about, are there different levels of passing? So I think that needs to be reflected.

If we decide to go forward, since we're not the graders, the graders are CMS and the -- and other end users, I think we should

think about the measure itself -- but reflect in the report, and we should all read it, but it accurately reflects those concerns, if that's the concern.

Would you agree, Marcia and Karen? Okay. Kelsey?

MS. McCARTY: Perhaps we will bring this up later in the gap discussion. But I was looking in the NQF database and there isn't a comparable measure for patients under the age of 65, which I know Dr. Yates brought up earlier. I know that AHRQ has that older population at the core of who they care about, but we're talking -- earlier someone mentioned, you know, grandma and grandpa, they have falls, whatever. That's the population we cared about.

But another big at-risk population are the younger patients that have obesity and are at risk for a pulmonary embolism. And so that's not getting -- if that's the kind of thing we care about, improving those processes about medical optimization or choosing the surgery correctly

for those types of patients, then they are not captured at all under what NQF has as its measures.

So I don't know if that's relevant to this or to later, but I thought I would mention that.

DR. FLEISHER: Unless there is another measure, which there is not, correct? But that sounds like that should be in the gap, and I would ask you to help us write that section related to that measure. So when you do -- okay. Fred?

DR. GROVER: I guess this is getting to be redundant at this point, but I have concerns, too, in a population that -- part of which may be doing -- turning down an operation really leading to the consequences, are we really going to improve quality with this measure? And particularly without the risk adjustment with the comorbidities, because this is a population that so frequently has a lot of comorbidities that really can impact the hospital, and does there

need to be avoidance of risk, and that type of thing.

DR. FLEISHER: Melissa?

MS. THOMASON: I just wanted to weigh in. As a patient with lots of -- I've spent tons of time in a hospital bed, open heart surgery, three in a year, aortic dissection, I mean, over and over and over. And most of the work I do is inpatient-centered care.

I would never want a measure to imply that a hospital delivers lower quality care because they honor my wishes as a patient. You know because if I really am DNR, and I do have that hip fracture, as we were talking -- and I know that's a philosophical question that we will probably get into later, but I certainly wouldn't want to imply that a hospital is not doing its job by honoring my wishes and by listening to me.

DR. FLEISHER: Thank you. I think we have had a lot of comments about that within all of these measures. So as we vote on the validity, I think that's something that cuts

across almost any hospital mortality measure. But it will be good to identify that here.

Two quick more comments, because we -- I want to move on.

MR. MARKMAN: So with this specific measure, what -- I mean, is there public reporting? This is going back to Clifford's -- what is this used for?

DR. LYZENGA: So we -- that is, again, something that we are currently wrestling with, how to kind of address these questions. As a standing committee, dealing with the question of endorsement, we really are -- our guidance so far has been to try to kind of stay agnostic in some sense to that question of use.

Not exactly agnostic because, again, endorsement does imply that it is suitable for a range of purposes, including public reporting, payment, and quality improvement. But we -- you know, this committee is supposed to be looking at the measures, sort of scientific properties, the -- you know, is it a good measure? And sort of

staying, again, to some degree agnostic about what we would want it to be used for. But it's -- you can't avoid, you know, having that question in your head.

MR. MARKMAN: This is in the public -- I mean, you know, it's in that -- it's in the body of the statement of, you know, is there public reporting of -- I know we're jumping ahead towards, you know, a more --

DR. LYZENGA: Yes. The usability section, we should get some information about how it is being used and how it is -- how AHRQ and others intend to use it, so we'll address that to some degree.

DR. FLEISHER: I would like to focus on new points related to this section, so that we can move forward. If you do -- Amy?

MS. MOYER: Well, I certainly don't want to put measures out there that, you know, might drive patient's wishes be overridden. I think when measures get used is when they really get looked at, and when data really gets improved

and things really happen. I know we've certainly seen that when we've asked for measures or reporting from providers we work with.

Suddenly, it's, "Oh, we were putting that into the registry wrong." Or, "Oh, you know, we looked at our data and we found this." And so I guess my concern is if we just say, "Well, we're not going to measure things like this because there is this issue," the impetus for resolving that issue goes away somewhat, if, you know, we kind of back away and take things off the table.

I don't want to put things out there that are misleading, but I also don't want to let, you know, perfect be -- or I don't want to strive for perfection and not get something good out there as well I guess.

DR. FLEISHER: Thank you. You're paraphrasing it clearly.

DR. MOSS: I just wanted to expand on Melissa's point, which I think is very well stated. The issue is wrapped into all mortality

measures. Where does patient choice figure into that? You can make a conclusion that while that applies to all mortality measures, it is just built into the system, and we just have to accept it.

But I think it really is of very differential relevance depending on the patient population and the measure. I mean, patient choice to die in an elderly patient with a hip fracture is a much more significant issue, for example, than mortality in something like congenital heart surgery.

So I would just suggest that for this measure in particular that's a highly relevant point, and probably plays very strongly into how we might rate validity.

DR. FLEISHER: Thank you. I think we have outlined a lot of the issues.

New points. Patrick, do you want to make one quick comment before we vote? Go ahead, please.

DR. ROMANO: So, first of all, there

are some fundamental misconceptions about the risk adjustment that I need to address. So, in fact, all of the things that were mentioned here are included in the risk adjustment. So the risk adjustment approach that is used here is -- includes not just age and transfer status and gender, and the type of procedure that was done, but it includes a method that was developed by 3M that is called APR DRGs, risk of mortality score.

So some of you who work in hospitals are familiar with this. So the risk of mortality score incorporates comorbidities. So if you look in Table 5 that is shown in the materials, you can see that there are different levels of risk of mortality -- minor, moderate, major, and extreme. And those levels of mortality risk are essentially based on comorbid conditions.

So the -- as a result, the overall performance of this model is a C statistic of .893. And so those of you who are familiar with mortality models, that's actually an exceptionally high C statistic. That is a

measure of the discrimination of the model.

What it means in lay terms is that if you take a randomly selected survivor, and a randomly selected person who died, that 89 percent of the time a person who died would have the higher risk of mortality than the person who survived.

So the model actually does quite well in terms of discriminating different levels of mortality risk, precisely because it takes into consideration all of the things that were discussed today, including open fractures, which have 50 times higher odds of mortality in the risk adjustment model.

With reference to the patient's choice, so in consultation with the expert panel, the decision was made specifically to include patients -- there are about five percent of patients who opt not to have surgery. And to be honest, there was some concern that that may differ across hospitals, and that there may in fact be some tendency, if you ignore those

hospitals, for surgeons to discourage very high-risk patients from having surgery.

So, in fact, it was a deliberate decision to include the patients who don't have surgery to avoid this opportunity for gaining, and to sort of level the playing field across hospitals so that hospitals where the surgeons are more enthusiastic versus hospitals where the surgeons are less enthusiastic essentially get treated the same way.

And then the choice of whether to perform surgery or not, and what type of surgery to do, is then incorporated into the risk adjustment rather than as an exclusion.

Finally, with reference to the availability of 30-day mortality, I would just point out that this is only possible if you have data from a single payer. So Medicare can do this with their own data. Medicaid plans can do this. But the essence of the AHRQ quality indicators is that they are built on multi-payer data to capture all of the patients who cared for

in hospitals.

And then, unless the state has set up some specific data system to allow linkage to death certificates, which a few states have but most haven't, then there is no way to capture those post-discharge deaths.

Finally, the age cutoff of 65, this again was a deliberate decision/recommendation from the expert panel, because in the younger age group the hip fractures are boosted with people who have a pathologic condition, so have particular reasons to have early osteoporosis or degeneration of the hip joint. Many of these patients have cancer or other disease, myeloma, that is invading the bone. And so it creates a more heterogeneous and atypical population.

So for mortality measures, we would like to have in general less heterogeneous populations. And so that was the rationale for the age restriction. So I think hopefully that should clarify.

And then, finally, DNR, if we -- we

have the issue with DNR that some patients choose to become DNR after they have experienced complications in the hospital. So this is a problem, because we would like to know just whether the patient came in knowing that they didn't want to have any intervention and knowing that they wanted to go directly to hospice. That would be the ideal approach to perhaps separate those patients.

Unfortunately, again, we don't have consistent, accurate data collection with respect to the post forms or DNR status at admission to the hospital, which is what we'd like to know.

DR. YATES: The APR DRG that you're signing, is that on admission, or is that after the hospitalization?

DR. ROMANO: Correct. That is based on the conditions that were identified by the hospital as being present on admission.

DR. YATES: On admission.

DR. ROMANO: Yes.

DR. YATES: And so there is -- and

critical to the APR DRG that you're using is what they qualify the diagnosis as, whether someone has been predetermined to get a hip -- how do they know they're going to get a hip replacement for their hip fracture?

DR. ROMANO: Well, the focus of this measure is on patients who are admitted for hip fracture. So patients who have hip fracture in the hospital are excluded. So this is patients who are coming into the hospital --

DR. YATES: Right. The patient comes into the hospital with a hip fracture. How do they know they're going to have a total hip replacement on admission? How do they get qualified for total -- for hip replacement, minor or moderate?

DR. ROMANO: They don't know that. We're talking about the -- we're talking about the diagnoses that are used for adjustment. So those diagnoses -- for example, if a patient experienced a pulmonary embolus after hospitalization, we would not be adjusting for

that because that's a complication of the care.

DR. FLEISHER: I think that A.J.'s question is, do you then use the CPT code to also adjust, correct? What they actually had, the procedure?

DR. ROMANO: Yes. We use the procedure codes as part of the adjustment as well.

DR. FLEISHER: Okay. That's to move forward.

DR. YATES: But that's after the fact. You're now applying a pulmonary embolism as part of your APR DRG.

DR. ROMANO: No. The diagnoses have to be diagnosed at admission.

DR. YATES: Right.

DR. ROMANO: They have to be labeled as present on admission. The procedures are counted whenever they're done. That's -- whatever the procedure is done, it's --

DR. YATES: Right.

DR. ROMANO: -- that's the patient was

treated.

DR. YATES: And then there is a pulmonary embolism after the procedures, does that count toward the APR DRG, or is that assumed to have been there at present? Or it wasn't there at present?

DR. ROMANO: Okay. If the hospital reports that the patient came in with a PE, then it gets counted. If the hospital reports that the PE arose in the hospital, then it doesn't get counted in risk adjustment.

DR. YATES: Okay.

DR. ROMANO: But the procedure gets counted no matter when the procedure was done. Okay?

DR. YATES: I'm going to hold to my original statements.

DR. FLEISHER: Okay. That's fine.

The one question I have, Patrick, for myself is, do you have -- there are some states who have linked data, and there is Medicare who has linked data. Do you have any linked data to

say, if we look at 60 or 90 or 120 days, does this measure hold in that subset of patients, to see if there is any difference in how they are ranked from a quality standpoint?

DR. ROMANO: We have not ourselves done that analysis. There has been some empirical work in the area. And in general, as you might imagine, as you get longer time intervals following the event, the correlations weaken. And so the correlations are fairly strong at 30 days, but when you get out to 120 days, frankly, they weaken substantially, presumably because of outpatient care factors.

DR. FLEISHER: But you're not seeing -- my question was, you're not seeing people are getting 30 days and then putting them into hospice, so they avoid the 30-day measure, because at 35 days they die at a higher rate.

DR. ROMANO: Well, clearly, 30-day mortality is significantly higher than inpatient mortality. So, and clearly many of those deaths are occurring in skilled nursing or hospice

settings. So I can't deny the fact that some patients are being transferred out of the hospital with the expectation that they will die after discharge, either at home or in another setting of care. This is true.

But the question is, we haven't seen -- over this time period, we haven't seen a systematic change in length of stay or discharge distributions. Now, we did see that -- so back -- if you go back to the literature 20 years ago when the DRG system was introduced, there was a dramatic change in length of stay and in discharge patterns.

But over this period of observation, the last five years, we have not seen a significant trend towards shorter length of stay or a change in discharge distribution.

DR. FLEISHER: Thank you. I think -- that's one of the things I think Cliff and I were concerned about monitoring.

We should go forward and vote on validity.

DR. GUNNAR: Just one final -- so it would be possible for you to go back historically to 2008 through '12 and determine whether the 30-day or 60-day or 120-day mortality actually changed through any one of those years for that cohort.

MS. STARKS: I am not sure how many states achieved the data from that. We'll take that into consideration, I'm sure.

DR. FLEISHER: Thank you for our report. That will be important information.

Can we vote?

DR. LYZENGA: Yes. Let's go ahead and vote on validity. So your options here are high, moderate, low, and insufficient. Go ahead and vote.

DR. FLEISHER: Chicago rules, vote often, so that we can -- because it only counts once, correct?

DR. LYZENGA: Okay. All right. So we've got nine percent for high, 68 percent moderate, 23 percent low, zero percent

insufficient. So this passes on validity.

Now we can go ahead to usability. Sorry, we can skip past this. Or do we vote on each of these separately? Karen? I didn't think so either. We can skip over this, yes. We can go to -- oh, yes, skip this as a composite.

Okay. Feasibility. All right. So whether the data is generated during care or through electronic sources, data collection can be implemented without undue burden. That's kind of what we are addressing here.

MS. JOHNSON: Just real quickly, all of the other threats to validity, exclusions, that sort of thing, you did talk about risk adjustment. That's probably the biggest one.

But you -- we may want to just make sure from the committee that no one had any concerns about exclusions, just in case they didn't realize that that was part of the validity discussion.

DR. FLEISHER: Okay. So if you -- anybody have any specific concerns about

exclusions?

MS. THOMASON: I just have a quick questions for clarification purposes, Andrew. So when we say it passed on validity, does it pass on validity in the high, moderate, and low categories, and insufficient would mean it would not pass? Or low also means --

DR. LYZENGA: Low also means, I believe, it would not pass. So we're going with the two top and then the two bottom.

MS. THOMASON: Okay. Thank you.

DR. FLEISHER: Cliff, or A.J., any comments on feasibility?

DR. YATES: No comment.

DR. KO: Their Table 1 shows the number of hospitals. The one question I had is, why did the number of -- why is the number of hospitals going down in that table? It's just the data set. You have 3,500 in 2008, and down to 26-, 2,700 more recently.

DR. GUNNAR: They are all transferring to my hospital.

(Laughter.)

DR. FLEISHER: I think there is some regionalization. Do you have an answer or --

DR. ROMANO: Yes. So the rationale or the reason for that particular drop is in footnote -- the third footnote here, which is that basically between 2010 and 2011 we made a change in the hospitals that were considered eligible for the measure.

And so there were certain, for example, specialty hospitals and rehabilitation hospitals that previously were in the reference population for the measure, and so we cleaned that up and focused only on the acute care, the general acute care hospitals. And so that is the drop of the 800 hospitals from 3,500 to roughly 2,700.

DR. FLEISHER: Okay. Why don't we vote.

DR. LYZENGA: Yes. Let's go ahead and vote on feasibility. Again, you've got high, moderate, low, and insufficient as your options.

DR. FLEISHER: You had a comment before we voted?

MS. McCARTY: Yes. Based on what he just said. So if you've changed the methodology in terms of which hospitals are included, then in terms of the drop that we've been talking about, how there has been noticeable improvement over the past five years, was that analysis redone going back to 2008 to look at just the population you care about? Are we comparing apples and oranges with those two different cohorts?

DR. ROMANO: Well, effectively, all of the -- these are patients with acute hip fractures who are being brought in by ambulances. So, in effect, they are all going to general acute care hospitals anyway. So these 700 or 800 excluded hospitals essentially had no cases anyway, so it makes the data set more logical.

DR. LYZENGA: All right. So go ahead and cast your vote on feasibility. Okay. Sufficient votes. We have 74 percent high, 26 percent moderate, and zero for low and

insufficient. The measures passes -- feasibility. And we'll move on to usability now, use and usability.

So this is sort of a question of how the measure has been used, how it is planned to be used, if it has shown improvement during the course of its use, and whether it is usable for consumers and other viewers of the health care information.

DR. FLEISHER: Any comments from our developers first? No? Cliff or A.J.?

DR. YATES: My comment earlier is that we are not allowed to tier these. There is definite usability for the process. It's just a question of whether or not it's usable at a level that would imply public reporting across all hospitals and across possible payment and adjustments. So I --

DR. FLEISHER: Can you separate those two different just real quickly? I mean, public reporting --

DR. YATES: Well, public reporting is

-- there is 500 different public report cards out there on the internet that are commercial, and of course there's hospitalcompare.gov, which is a centralized Medicare reporting system.

And then that would be one set of consequences is that something is used to adjudge that public reporting by hospital and regions or across the country. And the second thing would be value-based payments or payments would be based on either CMS or on HMOs.

DR. FLEISHER: So I am just saying, do you have a different opinion or the same opinion for both?

DR. YATES: I'm sorry. I beg your pardon. I have the same opinion for both of those, but my -- I still think it's a very usable and important measure for general trends and for individual hospitals to assess themselves.

DR. FLEISHER: Quality improvement. Great.

Any other comments? Great. Let's vote.

Sure. Go ahead. Melissa?

MS. THOMASON: Have you guys looked at the -- do we know if real-world patients are using this data as they look for providers? Do we have any idea yet?

DR. ROMANO: There is limited evidence on that. So this measure is not a measure that is used in the CMS HQR, hospital quality reporting program. So, therefore, it is only available when the state health data agency has chosen to report it or when hospitals themselves have chosen, in the interest of transparency, to report it publicly.

So I would suspect based on that that there has been very little actual use by consumers. I think we see more consumer use of the measures that are incorporated into the CMS programs.

One that -- I just also have to say in terms of -- there is a general sort of question about process measures to outcome measures. It is the philosophy of the QI program to focus on

outcomes and to let the hospitals and doctors kind of figure out the processes. So a natural response to an outcome measure like this would be if patients are dying, well, let's look at why they're dying.

And if they're dying from PEs, what do we need to do in terms of pharma prophylaxis? If they're dying from infection, what do we need to do in terms of infection prevention? If they're dying from MIs or arrhythmias, what do we need to do in terms of cardiac risk assessment and risk reduction?

So it's understood that outcome measures really pose questions, and that it's our work -- and, again, putting on my doctor hat now, it's our work within healthcare organizations to kind of figure out what is going on, what is contributing to the issue, and to address those process factors.

DR. FLEISHER: Thank you. And just to reiterate -- to follow up on that CSAC, really the process measures -- really, if we have

outcomes measures, that's what we should focus on. And we should only approve process measures if they are -- as a substantial absence of outcome measure or the process measure would so move the field forward.

And I think this committee has seen that, where we retired or put in -- excuse me, we put on reserve status process measures that are still evidence-based but are topped out or in and of itself would not make performance improvement.

Cliff?

DR. KO: Yes. I just wanted to ask Pat, have you -- since this measure has been out, available for such a long time, and, you know, when you -- in your application you said it's being used, and you mentioned two I guess systems that is using it for public reporting, but why hasn't the uptake been much greater? Because it has been out there a long time. It's a relatively easy measure to get, to do, to perform. Have you received any feedback why it hasn't been taken up?

MS. STARKS: Well, I think, first of all, it's a very difficult thing to capture, what the take-up is. But there are a number of states that are reporting this particular measure. I don't know the exact number, but we have a program called Monarch. And software is used, and this measure is included in that software. So I wish that we did have better information about the uptake.

DR. FLEISHER: Amy, did you have a definitive comment?

DR. ROMANO: Sorry. I was just going to say I think that we are missing -- in this submission, we are missing some information about from an inventory of state reporting programs. And I can get you that information, but there are several states that are recording it. I just -- we don't know those states off the top of our head.

DR. FLEISHER: Okay. It would be great to get that for the -- any follow-up call or get out to the committee, if there is any

question.

MS. MOYER: I was just going to say as, you know, our population is commercial and we do public reporting. And I think what I would struggle with including this measure in our public report is, what is our population going to do with that information? I mean, they are not really the ones targeted by the measure.

It is not really shoppable, as we may call it. So I think it can be really helpful, you know, if you're a state and you're kind of trying to look at overall quality or, you know, how hospitals in the state are doing. But I think would really struggle from a commercial application.

DR. FLEISHER: Thank you. Shall we vote on usability? One more comment from --

MS. THOMASON: So I thought of it entirely different. I thought it would be very "shoppable" when we talked about mortality, and if I was looking at a hospital and where I wanted to have these things done, if I had broken my hip

or if my mother had, and all those things. I would want to know that.

DR. FLEISHER: It's an emergency operation and people go to the closest hospital. So I think that's the point that was made previously.

MS. THOMASON: So none of it is like a procedure at all. It's all an urgent --

DR. FLEISHER: No. This is an emergency.

MS. THOMASON: Okay.

DR. FLEISHER: And we can take that offline for further details.

DR. LYZENGA: Let's go ahead and vote. It looks like some folks have started already. Go ahead and enter your votes.

DR. FLEISHER: Missing two. Go ahead. Alexandra?

DR. LYZENGA: All right. We have 23 percent high, 73 percent moderate, five percent low, and zero percent insufficient. So the measure passes on usability and use.

So I think the next step is to go ahead and vote on the measure's overall suitability for endorsement, unless there are any other points of discussion that we haven't covered yet. We could probably just vote.

All right. Let's go ahead. Are you ready, Alexandra? Okay.

DR. FLEISHER: Andrew?

DR. LYZENGA: Ninety-one percent yes, nine percent no. The measure passes.

DR. FLEISHER: I want to thank everyone. I think it was an incredibly robust discussion.

(Laughter.)

I mean that in all sincerity, despite the laughter, because I think it really will help inform the next couple of days -- today and tomorrow -- it will inform the report, it will inform CSAC, since I will be able to bring it back as the Vice Chair. And as we go forward, I think it will be important to bring new comments, so that we can continue to stay on time.

We are done this --

DR. LYZENGA: Yes.

DR. FLEISHER: Why don't we take a break, and then Bill will take over for the next measure. We will take until 10:45, if you can come back here then.

(Whereupon, the above-entitled matter went off the record at 10:27 a.m. and resumed at 10:41 a.m.)

DR. GUNNAR: So, Andrew, do you want to take a dinner vote or --

DR. LYZENGA: I think we were going to try to sort of get a sense of that at lunch, how many people wanted to -- I guess we could actually just -- yes, might as well just do a hand count now. How many of you are interested in going to dinner? We can make a reservation at a restaurant around here, and just want to kind of get a headcount. Let's call it 15.

Thanks, everybody.

DR. GUNNAR: So the next measure to be discussed is 0360, esophageal resection mortality

rate. And just to be clear, this is -- this has been an endorsed measure since 2008 and reendorsed in 2011. So who are the discussants? Oh, I'm sorry, the -- I also want to know who the discussants are. Okay. Great. That would be wonderful.

All right. Developers? Dr. Romano?

DR. ROMANO: So this is an unusual measure in that we are going to say right up front that this measure is unreliable for the great majority of hospitals. And so best to get that out of the way.

(Laughter.)

This measure is intended for use in combination with the measure of esophageal resection volume, which is IQI 1, which is the next measure that will be under consideration.

And the notion here is that there is a very strong repeatedly demonstrated volume outcome association in a total of 29 studies, according to our latest literature that exists. So we know that from the patient's perspective,

absent any information about specific hospital quality, that volume is very important, that the hospitals that have more experience with this type of surgery show better outcomes.

Among the hospitals that have higher volume, there is a real difference in mortality. And so it makes sense, thus, to have both a volume measure and a risk-adjusted mortality measure, because for hospitals in the lower part of the volume distribution you focus on their volume and their low volume. For hospitals at the higher end of the volume distribution, you can actually learn something from their actual experience with risk-adjusted mortality.

And that mortality experience may drive the hospital's own self-examination as well as decisions by, for example, payers in contracting with centers of excellence. So that is just a quick background to really viewing these measures together.

DR. LYZENGA: Thanks, Patrick. And I think we have Melissa and Larissa as our

discussants.

DR. TEMPLE: Well, I will start. True disclosure, I found this measure very difficult to evaluate. And I do think that maybe as we go through it we can also have some input from the people who looked at the volume measure as well.

This is a simple measure in the sense that it is looking at the number of inpatient deaths per 100,000 discharges for patients undergoing esophageal resection for predominantly GI -- for esophageal and upper gastric cancers.

It is, as reported, high volume relationships, and the developers argued that in addition to a volume mortality relationship there is also variability within each of the volumes, to suggest that mortality, in and of itself, is worth measuring.

If we go to -- so as an outcome measure, there is plenty of evidence to suggest that there is a volume relationship. And they define, actually, that more than eight procedures per year seems to be the acceptable high volume

center.

So I guess we're looking first at the evidence. I was just looking for the slide to vote. Melissa?

MS. THOMASON: So do you want me to jump in, Larissa, for evidence? Okay. So, for evidence, again, there is that rationale behind it, just like we talked about with the last outcome measure. And it's an established fact that hospitals that perform more of these also have fewer mortalities of them. Thus, there is a relationship that exists. Therefore, we can say, yes, there are processes that have been identified that affect the outcomes, and so it has an evidential basis. Is that correct?

DR. LYZENGA: Any other comments or thoughts or questions?

MS. THOMASON: That was a stretch for me, I'll be honest. Sort of following the logic of that, coming into this as a non-clinician/third party, you know, to say this, then this, then this, and then there's this other

conclusion way over here. And until I heard I think it's Patrick speak earlier, it didn't make a lot of sense, but now it does.

DR. LYZENGA: So any other comments on evidence, or should we go ahead and vote? Sounds like we can vote. So we're voting, again, on evidence, whether there's a rationale supporting the relationship with the health outcome to at least one health care structure, process intervention, or service. One for yes and two for no. Just go ahead and vote. And one more. There we go.

We have 95 percent yes, two -- or, sorry, five percent no. The measures passes on evidence.

So now we can move to performance gap.

DR. TEMPLE: So under the script, using the opportunities for improvement -- right, Andrew? So the authors report that there has been -- that there is room for improvement with rates ranging from 59 per 1,000 to now 41 per 1,000 deaths per year. There is -- if you look

at the table they provide from 2008 to 2012, they do show the rates of esophageal mortality being relatively flat.

There is a drop from 2011 to 2012 from 59 to 41 per 1,000, which I don't think that they can -- they didn't -- we really don't know why there was that drop. I'm curious if the measurers have any comments on that. But there is certainly room for improvement.

They do demonstrate that there are -- there is variability in the outcome based on age, gender, insurance, region, and income, to suggest that, again, there are opportunities for improvement.

DR. LYZENGA: Thanks, Larissa. Any other comments or questions about opportunity for improvement? Go ahead.

MS. PITZEN: Great. Thanks. I just wanted to make a comment. This is a relatively low, low volume procedure. If I'm looking at the numbers correctly, less than 5,000 cases nationally on an annual basis. So it's really

hard to measure and discriminate changes between practices, hospitals, et cetera.

DR. DUTTON: This may be the opposite of the situation Dr. Yates was referring to earlier where this is a very important measure for public accountability. We need to tell the public we are not killing people in esophageal surgery, but where there is very little opportunity for quality improvement out of it, because the discrimination is so low, because the numbers are so low. I mean, the improvement from 51 to 49 is a tenth of a patient per facility doing this or something like that.

So there is no -- there is not going to be a quality signal at the facility level. What Patrick was saying, it's not reliable to discriminate hospital performance. But at the same time, I think it is an important measure for public accountability.

DR. GUNNAR: Yes. Dr. Romano?

DR. ROMANO: Yes. I can address one of the comments, which is that this is actually -

- these indicators have gotten a fair amount of attention, because the Leapfrog Group actually publicly reports similar information for esophageal and pancreatic surgery. And I think some payers in fact have focused on contracting with centers with excellence for this type of cancer surgery.

So what we have seen between 2011 and 2012 is a decrease from 198 hospitals performing this surgery to 155. So basically 43 hospitals dropped out of this market, and those turned out to be higher mortality hospitals.

So that basically is what explains the drop from 51.9 to 40 -- or 5.2 percent to 4.0 percent in terms of in-hospital mortality.

DR. TEMPLE: So I'm going to actually ask the developers to put that in the comments in that section, because it is not clear there, because that is actually very, very important data.

DR. LYZENGA: Any other comments on performance gap, or should we vote?

DR. GUNNAR: Yes. I just want to -- this comes up in the -- may be Fred wants to comment on this -- the STS's comment about low volume facilities for CABG surgery. And it's not that low volume facilities can't perform excellent surgery, but if there is a -- but there is a higher tendency in the lower volume facilities to actually accumulate mortality. Let's put it that way.

So we just don't -- but that doesn't mean that there aren't a substantial number of low volume facilities that are actually doing great work. That is -- this gets to the issue of, is it -- is the measure prohibiting, you know, essentially the free practice in those facilities to be able to do good work? I guess is the fundamental question. Or do we care?

DR. GROVER: I would strongly -- I think for sure -- in cardiac surgery, that is very true, and particularly in things like coronary bypass and straightforward valve surgery. There are people -- there are surgeons

in low volume hospitals that do very good work, and so we have always gone by the real outcomes.

And I think esophagectomy and Whipple procedures are an area that is different. And Cliff obviously knows a lot about that, too, from the NSQIP database.

But so I think the logic here actually, Bill, probably is reasonable, because you can still be -- you can still stay open. And I'd ask this to you, Patrick. You can still continue to practice at a low volume hospital if your results are good, right? I mean, that's -- this is just kind of alerting people to the fact that -- well, you could have a surgeon who came out of a training program, say, in Michigan where they have always done a lot of esophageal work, and maybe in a smaller volume hospital might do quite good work. So --

DR. GUNNAR: So let's -- and then the last piece to this is, before I turn it back to Melissa, that -- it's about access. So when you have now decreased the number of -- when you --

we don't know that a decrease in the number of hospitals from 244 to 155 is -- was actually higher, lower volume. Your assumption is that those were low volume hospitals, now moving volume to what now is higher volume hospitals, and that -- those other hospitals benefited from the acute additional volume.

But we don't know what that did in relationship to access or distance to travel. These aren't simple procedures. And they -- it is really nice, given the potential for post-operative complications and having to manage those complications, to be closer to your physician. So I guess the question is, how have we have impacted access in relationship to this driving to a centers of excellence model?

DR. FLEISHER: Just one comment, which gets back to the public reporting. And I'd be curious, as we think about this -- public reporting may not be patients. It may be hospitals making decisions. So as we look at usability, it would -- a robustness of that data

could help frame some of the questions going forward.

Thank you for that information.

DR. GUNNAR: So hearing any additional -- yes, Dr. Cima.

DR. CIMA: You know, the one issue with this is, you know, we've talked about the volume and stuff, but the one issue that we -- and this is mortality measure, but the people who have taken care of esophagectomy patients know that mortality measure is one thing. The morbidity measure is a bigger thing. These patients can linger and go months and months after complications of the surgery and die. So that is not going to be picked up here.

And so that is one of the issues here is, you know, yes, you can do excellent -- I mean, an excellent operation, do it six times in a year, and, you know, you're going to have 100 percent survivability. It's going to look great. But you're not capturing the process of care which these centers of excellence do very well to

manage and mitigate complications. And that is really quality improvement.

I hate to say it, you know, because it is such a low volume thing. Mortality is not the best marker of quality here. You know, it is one thing to do 150,000 colectomies a year across the country and use that as a marker. But when you're doing less than 5,000 esophagectomies, that is probably even a high number, because a lot of patients now are getting neoadjuvant therapy, they are getting things, and they are not coming to surgery, other things.

That is the big issue with this is that -- is it really a reflection for a low volume hospital as mortality -- even if it's stellar or poor, is that really a marker of quality of surgical care that we can't risk adjust or at least monitor complications? I mean, is there a way of doing that for you?

DR. ROMANO: Yes. No, this is a very good point. Certainly, the complications after major esophageal surgery can be quite serious and

quite difficult to deal with and can lead to a protracted hospitalization. Obviously, this is an inpatient mortality measure. So if the patient dies after a series of complications, that is captured. But if they go home, and die at home, that would not be captured.

DR. CIMA: Or what if they -- more likely they go home and come back.

DR. ROMANO: Exactly. Yes.

DR. CIMA: They go to the nursing -- the rehab and come back.

DR. ROMANO: Right.

DR. CIMA: Six or seven times.

DR. ROMANO: Yes. This -- we have seen this, yes. So, I mean, all I can say is that here we have the advantage that this is an elective surgery, where the surgeon and the patient together make the decision to go for this kind of aggressive surgical resection. And so they presumably do that knowing the risks of the surgery, which are quite considerable. And they do that, you know, with a curative intent,

basically.

So this is clearly a minority of patients who have esophageal cancer. Most patients with esophageal cancer don't get this type of resection surgery. So, but these are the patients who are willing to travel. These are the patients who are seeking out the Mayo and Cleveland Clinic and the centers where they can go to experienced surgeons and experienced hospitals in the treatment of these conditions.

The other thing I would say is that there is -- there has been repeatedly demonstrated a correlation between -- so we have volume, higher volume is better consistently for esophageal surgery. Higher volume has been correlated with both morbidity and mortality outcomes, and length of stay. So, and those correlations are strong.

So we know that, in general, these things are tracking together. Morbidity measures, mortality measures, and length of stay are all tracking together with volume. But,

obviously, there may be some exceptions in individual hospitals that may differ.

DR. GUNNAR: That is a gap. So --

DR. LYZENGA: Every time you make these suggestions, we are writing your name down.

(Laughter.)

DR. SIPERSTEIN: So I kind of look at this as different measures for different focuses. I mean, this data is very easy to collect. It is all inclusive. But gives you -- I mean, the price you pay is that you have a more limited dimension of what you're looking at. On the other hand, the flip side, if you're participating, for example, in a -- you know, a procedure-specific risk-adjusted database, where you may have more limited participation, but you get more granular data in terms of your morbidities and mortalities, that may serve a different purpose.

So I would see this as kind of two ends of the spectrum where I think there is a value to this just because of the ease and

completeness of data collection, and that all hospitals can be easily represented.

DR. GUNNAR: Shall we vote on performance?

DR. LYZENGA: Yes. We have gotten a little off track here on discussion. So just to remind you, we are voting on performance gap here. So your options are high, moderate, low, and insufficient. And I think you can go ahead and vote now.

DR. EREKSON: Andrew, can you phrase the question, please?

DR. LYZENGA: Sure, yes.

DR. EREKSON: Before we answer the --

DR. LYZENGA: Yes. That's a good point, and I think actually we need to start up the vote again. But so we're asking here whether data demonstrate considerable variation in performance or overall less than optimal performance across providers. Again, sort of getting here to whether there is a demonstrable performance gap that this measure is addressing,

whether there is an opportunity in improvement -- opportunity for improvement in care that this measure is addressing.

Have we started up yet, Alexandra? I suppose while we're waiting maybe we could get started on our discussion of reliability. Never mind. We'll complete the vote. Scratch that.

DR. GUNNAR: We don't have -- it's technical difficulties.

DR. LYZENGA: Yes. It's just waiting, computer --

DR. GUNNAR: It's getting close to reboot.

DR. LYZENGA: Yes, go ahead, Melissa.

MS. THOMASON: So when we were speaking earlier and you said that so there are places -- facilities with low volume that you may actually get really good care at --

DR. GUNNAR: Exactly. Yes.

MS. THOMASON: -- but then you said, but mortality accumulates here. And so is that --

DR. GUNNAR: No, no. What I'm trying to say is if you take the -- if you just look at populations of hospitals by, in this case, esophageal -- we are really trying to -- we are combining mortality and volume relationships by virtue of the way that these are written by the developer.

So this is mortality rate, and the evidence is that there is an association between volume and mortality. This discussion really should be left for the volume measure, which is 0361. This is -- we really need to refocus ourselves, and this is a good -- maybe this is good that we've had the white screen, the fact that this is really just mortality rate.

So that if you are accumulating mortality in your center, regardless of the volume, that is a measurable outcome. It's an outcome measure that is related back to quality, and that has been endorsed since 2008.

MS. THOMASON: So, as a patient, I guess -- and it's really in all of these

mortality measures, it almost seems limiting that we look at it as a facility and not as a provider. So I'm not even sure that -- but when you say you have -- you know, you have this doctor that comes from this place where they do all of these procedures and he's so great at them, and then he comes to this place, this facility, with this -- with low volume. So it looks like, you know, he is in a place with numbers that don't really represent what he is capable of doing.

As a patient, I would love to know that he is so really great at it. But it seems that by making mortality these facility-designated measures, we limit that.

DR. FLEISHER: Lynn, did you have a comment related to that?

MS. REEDE: I do. And I appreciate the concern, Melissa, that it being provider-specific, but this really is, as we talk about, a team sport. And particularly this procedure is very complex in how the team cares for this

patient. So it does look more across the facility than just the person who provided the surgical intervention. At least that's my perspective.

DR. FLEISHER: Dr. Moss, did you --

DR. MOSS: Another way to answer that is, in aggregate, all low volume centers taken together will have a higher mortality rate. But if you look at them individually, there are going to be isolated low volume centers that do very, very well. But these kind of analyses can't prove that statistically because the volume is low.

So they might do well, but you're not going to be able to confirm that with this kind of information.

DR. FLEISHER: Amy?

MS. MOYER: One other thing I would add, I agree it would be nice if we could look at both surgeon and facility. I think in some ways it could a shortage -- a shortcoming of the data set we are using. At least I know the one we get

for the State of Wisconsin, we can't identify the surgeon. We can only tell where the facility was where this happened.

MS. THOMASON: And I guess I'd ask, because one of the things I'm concerned about -- like I co-chaired NQF's Committee on Consumer Affordability last year, and we talked about access to data, access to cost data, access to quality data, and all of these things, and a lot of this quality data we're talking about today.

So having access to the data and it being usable by someone who needs these procedures, it just seems that -- like if it's a team sport, and it really is, so we know in these -- especially in like these procedures. So maybe it's not advantageous for me to go to a low volume facility, even if there is a provider there who may provide excellent care. You know? So maybe it is an accurate representation from that perspective. Yes?

DR. GUNNAR: Again, that could be just to refocus ourselves on this -- this is the

mortality measure. So this is really about -- I would refrain that not about volume but about mortality rates.

MS. THOMASON: So I guess the question comes in -- my question with that is, so is mortality a good representation of quality in this instance? I guess that's where my question would --

DR. GUNNAR: And I think the -- and we're going to have to speak to Dr. Romano, but it -- I think they have -- that has been asked and I believe it has been answered in the evidence.

DR. LYZENGA: And we will vote on that again in validity, actually. So that -- and we'll have a question and you can weigh in on that question specifically.

DR. GUNNAR: Okay. So the -- yes, Dr. Romano?

DR. ROMANO: Yes. Just two things I can address quickly from some good literature that has been published from John Birkmeyer's

group from Michigan and Dartmouth.

So in The New England Journal of Medicine, in 2011, they described that 32 percent of the decrease in adjusted mortality for esophageal cancer surgery between 1999 and 2008 appeared to be attributable to higher hospital volumes, so directly to your point.

So one-third of the decrease -- one-third of the observed decrease in mortality appears to be due towards -- due to a concentration of patients in higher volume hospitals; two-thirds presumably to other factors.

And partitioning out the physician versus the hospital effect, it appears that physician and hospital volume are correlated with each other. So physicians who go to higher volume hospitals tend to get more volume, as you would expect. But statistically 46 percent of the hospital volume effect is explained by surgeon volume; 54 percent is not.

So there is a correlation to some

extent. Hospital volume is proxying for or picking up the effect of surgeon volume. To some extent it is an independent effect, reflecting the team concept of care that Lynn has described.

DR. LYZENGA: So I think actually as we are trying to resolve this technical difficulty, we may just do a hand vote on this one, so we can go ahead and move forward. We do have four options here, so I guess I'll just go one by one and ask people to raise their hands for each. So, again, we're voting here on opportunity for improvement and the options are high, moderate, low, and insufficient.

So first I'll ask who wants to rate this measure high? I've got nine for high. And who would like to vote moderate?

PARTICIPANT: One vote from the phone.

MR. LYZENGA: I've got 14 including the vote on the phone. Thank you.

And low and insufficient. Okay. So the measure passes on importance to performance gap and we can go ahead and move to reliability.

Next vote we should be able to use the system.

DR. TEMPLE: Can I see the next voting slide? Is that okay?

MR. LYZENGA: Yes. Alexandra, can you skip to the reliability vote slide?

I'll record it just by hand here and then we'll add it into the report. It's in the transcript as well.

DR. TEMPLE: So the reliability, I know we want to stay away from the volume piece of it. I'll do my very best to do that. But it will come into play a little bit.

They did the reliability testing. They had data from 36 states. It included 82 percent of hospital who do esophageal resection. It was 655 hospitals with 4,331 patients. So they did the reliability testing. They state that the data is very reliable when there's more than eight patients discharged with esophageal resection per year.

But when they start to look at their

signal to noise ratio, it's really 0.14. So it's very poor. They have done the risk adjustment with the smooth curve statistical modeling. And that does report out that there's some reliability.

But I think we're going to struggle with this because it's reliable in higher volume centers and not as reliable in lower volume centers. But yet there's no exclusion of low volume centers in this measure. So I put it out to the Committee to discuss.

DR. GUNNAR: Any argument? Dr. Romano.

DR. ROMANO: Yes, I mean we have to accept the fact that it's not reliable. And the question is is it then despite the lack of reliability useful together with the volume measure. And we argue that it is because it allows you to identify on either end among the higher volume hospitals the difference in performance between those that are both high volume and provide the highest level of care and

those maybe that provide a lower level of care.

But there are a variety of -- I should say there's been some very interesting work that's been done again by Birkmeyer's group to try to combine these two measures into a single measure. And some of you may be familiar with that.

It's been picked up by the Leapfrog Group as what's called a mortality predictor. The idea is to use a Bayesian approach to actually combine these two measures together into a single measure that might be more -- that would have the reliability advantages of the volume measure but would still incorporate information about mortality.

So that's something that is currently under evaluation as part of our current work with AHRQ. And it may be a direction that we go in the future.

DR. TEMPLE: Can I just ask a point of clarification? And I'm sorry to go into the low volume piece, but I think we have to with this

measure. If your volume is low, is the metric of mortality reliable? And it's not. Right. So it's only helpful for --

DR. GUNNAR: You framed it in relationship to the mortality measurement. So it's correct. There is the reliability of this data at a low volume center is poor by admission.

Dr. Moss. Oh, did you -- No. Anyone? Dr. Yates.

DR. YATES: That is true for any small sample size. You're going to have as part of how physician profiling that we have to throw out the small volume physicians because the sampling is so wide in terms of center of deviation. I think it's just a mathematical necessity of that being the case.

DR. GUNNAR: Any other discussion?

DR. CIMA: But then that gets to the usability. I mean down the road it's for patients and even for payers or health business in general.

DR. GUNNAR: But to be clear that's

why if any one of these fails to meet. My understanding, Andrew, is that if we failed any one of these measures or any one of the components then the measure fails. Correct?

MR. LYZENGA: Only the first two criteria, importance to measure and report and scientific acceptability. Those two are must pass. Usability and feasibility actually are not must pass.

DR. GUNNAR: Amy, did you have something?

MS. MOYER: I guess I had a question in looking at the volumes that are in the measure and the reliabilities that relate to volume for this. I would almost want to probably run this as a multi-year if I were using this using more than one year of data. And I was curious if you'd done that and looked at that and how that might not impact the performance of the measure.

DR. GUNNAR: So question for the developers. Have you considered or would you consider multi-year analysis for the mortality

relationships when you're dealing with low volume reliability issues?

DR. ROMANO: Yes and I believe -- I'm going to check. Is Sheryl Stanford on the line? Do we have the lines open?

MS. FELDMAN: Can you let us know --

MS. DAVIES: Yes, I'm here.

DR. ROMANO: Thank you. Sheryl, can you address whether this particular analysis was done with two years of data or one year?

MS. DAVIES: This is done on our typical measure set, which is done on one year.

DR. ROMANO: Yes. In answer to your question, yes. For some analyses we use two years of data. In this case, clearly I think reliability would be higher with two years of data or three years of data. We could estimate those numbers and bring them back to the Committee.

DR. GUNNAR: Melissa.

MS. THOMASON: I know when we talked about the last measure and we talked about low

volume and less reliable and you went through the reliability adjusted rate and smoothing and all that. How much does that help with this measure?

DR. ROMANO: The same approach is used here with smoothing so that hospitals are smoothed back to the overall average. And in fact the practical result of that is that most hospitals' performance appears to be average. And that's a limitation of a measure that admittedly has low reliability. The smoothed measures come back to the average. So it looks like the hospitals are indistinguishable from the consumer's perspective except again for the high volume hospitals where the smooth measure is primarily reflecting the hospital's own experience.

DR. GUNNAR: Any other discussion? I think we're ready to vote on reliability.

MR. LYZENGA: Are we ready to vote?

DR. GUNNAR: Yes. Go ahead and start.

MR. LYZENGA: This is a test for anyone who votes moderate or high.

All right. I think that's everybody. Can we get the results? All right. So we have zero percent high, 17 percent moderate and 70 percent low and 13 percent insufficient.

That means that the measure does not pass on reliability which means we stop the discussion I believe. And that means the measure will be recommended -- The Committee will recommend that this measure loses endorsement. And that saves us some time at least. We'll go on to the next one.

DR. ROMANO: So I guess I'll just ask a question. Is there -- So we'll go on to discuss the volume measure now. My question would be whether there -- Well, maybe we'll discuss in the context of volume measure whether they may be an opportunity to come back to the Committee with a combined measure or some other approach.

MR. LYZENGA: We can certainly discuss that.

DR. FLEISHER: So that would actually

be coming back after the next Surgery Standing Committee as opposed to within this project.

MR. LYZENGA: We would have to talk to AHRQ about the time line and feasibility of doing that, but likely within the next project I would expect. It kind of depends on funding and other factors. So we can't really say with too much certainty when the next surgery cycle will be. But we would expect it to be within the next year or two I think.

DR. FLEISHER: So we --

MS. JOHNSON: I'm sorry, Lee. We should probably talk about that a little bit later offline with Patrick. If you guys could do it quickly, there is a small possibility it might get looked at after comments. But we can -- It doesn't -- yes.

DR. FLEISHER: Right. My question to Carol and Patrick is do you want to go forward with the volume measure alone if we're not going to consider the outcome measure. That's just a question. Because if you think you could come

back within this cycle, then we should probably discuss it. And I don't know if you could come back by the conference call that we're going to have.

Is there an interest from the original discussants that if they came back with a paired measure we should look at it?

DR. TEMPLE: Yes.

DR. FLEISHER: Yes, that's a pretty resounding yes. So my question to AHRQ. And if that's okay, NQF Staff?

MS. MURPHY: I was just going to say that as endorsed currently they are endorsed to be reported as a pair.

DR. FLEISHER: So what's the implication?

MS. JOHNSON: Patrick, you are talking about bringing back a measure that in one measure incorporates the volume, not having a separate measure that's volume versus -- Am I correct?

DR. ROMANO: Correct. So the current endorsement does stipulate that the measures are

always paired. And NQF staff can better explain what that means in practice.

But what we could do is to come back with a single measure. It would combine both volume and that preparatory work has actually been already by Birkmeyer's group. And we've had multiple discussions about how that would be operationalized. So it could be done. It's just that we haven't yet -- We wanted to see how the discussion flowed in this process before implementing that.

DR. GUNNAR: Could I ask a fundamental question? So is the reliability issue going to plague the volumetric measurement as well?

DR. ROMANO: We'll find out, but I don't think so.

DR. GUNNAR: Under your own data understanding now. The mortality at low volumes suffers a reliability issue. Does the volume measurement of a low -- is that impacted as well? Okay. So you can identify reliably low volume centers, just not the mortality within that low.

So how that will help a composite or a combined measurement?

DR. ROMANO: The approach is that as I mentioned in a smoothed measure the hospital's actual risk-adjusted mortality is smoothed or shrunken back to the overall need based on the distribution that's observed in the reference population.

Now in the approach that John Birkmeyer's group has developed, the risk-adjusted measure is smoothed back not to the overall mean, but to the mean of hospitals with the same volume. So what this means is that it's really incorporating volume and mortality together into a single measure.

The assumption is that if you don't know anything else about the hospital you go with the volume. And you go with the fact that on average low volume hospitals have higher mortality. If you know a lot about the hospital from the hospital's own experience and their risk-adjusted mortality, then you put more weight

on the observed experience.

This is a Bayesian design of the measure where the prior is driven by volume stratified mortality. And then that prior is adjusted according to the hospital's own experience. And Birkmeyer's group has shown that this approach has very high reliability actually and repeatability over time.

DR. CIMA: Let me make a comment. I mean in some ways we have a conundrum of any low volume event being able to statistically measure it. And I think in the zeal to be able to measure the mortality rate of a hospital who does four cases a year in some ways it's throwing out the baby with the bath water here.

And the question is statistically if you have a low volume hospital should they simply have an asterisk that says within competency intervals we can't give you a result for this particular institution. I mean there are other ways to deal with this because there is a reliability for higher volume institutions. It's

simply I think the difficulty that we're having is do we have reliability for every institution included in the metric.

DR. FLEISHER: It sounds like both Birkmeyer's group and actually Jeff Silber and Paul Rosenbaum are submitting a manuscript on the part of my group that has also done a new Bayesian model. I think there will be new statistics coming out.

DR. SAIGAL: But I would agree with that point. I mean basically you're putting like lipstick on a pig in some ways. You've got very little data on small volume hospitals and you can sort of say that they're all the same. But you don't really know that.

So from a consumer point of view I think it's not helpful to say that you think these are all the same. Higher reliability it's confused by that we just don't know. So I'm not sure that even with a Bayesian approach you're really getting to the consumer's point of view.

DR. FLEISHER: John was next.

DR HANDY: If we are going put on a reserve status esophageal outcomes and an operation that's infrequent but does have the potential for high mortality and lots of morbidity, this is an area that cries out for a composite measure. Getting to Dr. Cima's point is that most patients don't die. Most patients have morbidity.

I don't know that fast-tracking this into lesser surrogates is really the best way to go. Maybe a rethinking with a more comprehensive composite is a way to monitor this important clinical activity.

DR. FLEISHER: So this would not be eligible in my mind for reserve status. This is an up or down. Reserve status means the evidence is high. Everything else passes except gap.

DR HANDY: Gap, okay. Got it.

DR. FLEISHER: So this would essentially lose endorsement as the decision.

DR HANDY: Right. So what I say still stands though.

DR. FLEISHER: Amy.

MR. LYZENGA: We're on 361, right? We've moved past 360.

MR. LYZENGA: We're having a general discussion right now. We haven't jumped into the measure quite yet.

DR. FLEISHER: Amy and then we want to come back to Patrick and Carol about how they want to approach, whether they want to move forward with this measure after these comments or how they want to approach it. Amy.

MS. MOYER: I have one comment on process. I'm wondering if -- You know last year we had those series of calls that felt like it gave the developers a little more lead time on potential issues we might have.

And I'm not saying like the measure could be respecified. But if we said, "Hey, we want to see this for two years of data" there was a time frame that might have allowed for that. I'm not saying I don't like the new process and do not miss all the calls. But it's something

that potentially was lost.

And if both measures are coming back, I just wanted to touch on one thing that confused me which was it appeared they had different denominators. So for the mortality rate, the inclusions all required not just the procedure but also the cancer diagnosis.

And then for the volume it was any resection regardless of diagnosis. And then gastrectomy for a cancer diagnosis. That to me was a little confusing and I don't want you to bring it back and then blindside you with that concern.

DR. FLEISHER: Collette.

MS. PITZEN: Thanks. I just wanted to share that I understand the relationship between volume of procedures and technical proficiency. But I think the measure science has evolved beyond reporting structural measure of volume when you could have good outcome measures in place. And given our prior discussion about the unreliability of a low volume procedure and

mortality rates, I'm just kind of throwing it out there.

I know for some of the measures that we do we're actually reporting functional status outcomes post on knee and hip -- I am sorry, on knee and spine surgery. And we're actually reporting the volume along with the outcome rates, not as a separate measure but together.

DR. FLEISHER: I would like NQF staff to just comment because Fred checked that there is a risk-adjusted mortality and morbidity, NQF approved measure from STS which we have said that if there are two competing measures and they're both valid we should approve them both.

But it also gets to in this space of what the right thing to do is because we are a standing committee now, not a measure-specific. So thoughts if I'm making myself clear.

MS. JOHNSON: It's a little bit difficult because we ask you to do something that's a little hard to do which is consider measures with blinders on, not thinking about

whether there is a competing measure or not. And then if you like both measures, then to have the discussion about competing or related and that sort of thing.

I'm not as familiar with the STS measure. I haven't to look at it. When measures are competing, you as a committee can decide if you think that there's justification for having both or if there's a reason to pick a superior one. I think if you're going to look at that measure tomorrow, we'll just have to remember -- You're not going to, okay. So this is outside of -- Okay.

DR. FLEISHER: The reason I'm just bringing it up and I wouldn't bring it up if we didn't have the problem with the first half of this measure. Essentially, Patrick and Carol, any thoughts on it? I'm happy to continue proceeding. The question is given the comments how would you like to proceed? Is that appropriate?

DR. ROMANO: I think we can't commit

AHRQ here and now. Part of it is that our lead program officer is not here and we'd need to have a broader discussion. What I would say is that if the staff and this Committee are open in the current cycle to potentially considering a composite measure that would be a composite of volume and mortality, then we should go ahead and discuss the volume measure. We would want to incorporate any questions or concerns that the Committee might have regarding the volume measure into what we bring back. If that is not an option, then we'll take both measures off the table.

MR. LYZENGA: Again, it depends on AHRQ's capacity to do the work within this cycle. We'll maybe have to loop back with you on that.

I should also ask. The measure is officially paired with the volume and mortality measures. So that would actually under NQF policy, if they are paired, that means if one goes down the other goes down is my understanding.

MS. JOHNSON: I think I want to confer with somebody else and see if that's the case or not.

MR. LYZENGA: Yes, we'll have to check on that.

MS. JOHNSON: That one throws me every time. I have to apologize.

MR. LYZENGA: But potentially a question for you is whether you're open to dropping that pairing so that this one could maintain endorsement, if that is the case, without the other.

DR. GUNNAR: I think, to the developer's benefit, why don't we have the discussion on the next measure, the volume measure, and then make a determination if the fact that they're paired should have influence on our decision. And then we can move on.

DR. FLEISHER: I agree. Although I don't think we can actually make a final determination of an endorsement. But we can actually give you insight to what it would take

to get over the hurdle this time. Because if we don't have that, we'll fail at the next call if you decide to make it. If that's okay, the discussants are? Okay. And who wants to start?

DR. SAIGAL: I am happy to start if you want. This is a measure about resection volume. The outcome is a structural measure of quality in terms of the evidence. They present direct evidence of a relationship between hospital volume and outcome. There's also evidence that the surgeon volume is an explanatory variable here.

There are a few studies that show a lack of relationship. But the preponderance of data indicates that there is a relationship. I did find one site I didn't cite from the NIS that showed the C statistic from this dataset was about three percent. So very little of the variation in mortality was explained by just the hospital volume in that study versus the national study. Probably the surgeon volume is also very important and other factors are being measures in

terms of its relationship to mortality.

That's the evidence. I don't know if there's anything else to say about that.

PARTICIPANT: Three percent?

DR. SAIGAL: That's a C statistic, yes. The best model was like more than 15 procedures in the C statistic which was three percent. So it wasn't very high. That's from NIS.

PARTICIPANT: The difference?

DR. SAIGAL: No, that absolute C statistic was three percent. This was something that came out in 2009. And the C was 3.87 percent and varied between different models by 0.64 percent. So the overall volume which was a less than one percent variance was due to the volume of the hospital patient mortality.

I didn't see that study in this. So this wasn't a comprehensive review that was submitted. It was basically a scan of what they found in more of an ad hoc way it seems like.

DR. FLEISHER: What I might suggest given one measure has failed of a paired measure

is that we just have a discussion without a vote on any individual aspect of evidence, reliability, validity if the Committee is acceptable. We essentially inform AHRQ so that they can respond. Does that work for staff? If you want to continue, Chris.

DR. SAIGAL: Okay.

DR. FLEISHER: And any other comments. And then we have a robust discussion and then call it.

DR. SAIGAL: Sure. There is a gap. There's a very large variation that hospitals reported from 2 to 25 surgeries annually. It's claims-based. In terms of validity testing, I don't really see any besides face validity basically showing people the measure to say "Does this look meaningful to you?"

And it is a claims-based or administrative approach looking at IC-9 codes as someone said which are procedure codes that are less specific than CPT codes. I didn't see any evidence that there was a chart abstraction at

any point to say whether they were capturing the surgeries. I thought they were capturing the surgeries from a chart review. There is just no data about that that I saw in their application.

In terms of its -- One thing about usability and use, it's being used in three states by one AMC and one health system for a couple of years. They did not report that I saw in this application that there was a shift in the distribution of volumes of hospitals.

Someone else said earlier that there was a drop of low volume hospitals that Leapfrog was measuring. But I didn't see it in this application. At least in what is here, there wasn't data to say that the impact of the measure had changed the distribution of procedures in different hospitals. But maybe that wasn't presented in this application.

DR. FLEISHER: That was Patrick. Do you have data to answer that, Chris'?

DR. ROMANO: This particular analysis as reported here is only with the 2011-2012 data.

In the other data that we just reviewed for IQI 8, that analysis goes back to further years and shows that change in volume.

DR. SAIGAL: I would suggest if you have better data to say it's impactful you should include it. And if that's the case, then I would say that's great. According to this application, there's been no impact. I guess I'll defer until it's been revised. Those were my overall comments.

DR. FLEISHER: Amy, comments?

MS. MOYER: My biggest question had been the discrepancy between the denominator definition on the two measures. If we're saying -- If we're somehow pairing them, it feels like we should be looking at the same patients for both instead of including additional people.

I was wondering if a way to increase the volume covered by the mortality rate would be to drop the cancer diagnosis. But then my guess was we're potentially trying to weed out traumatic or emergent patients potentially by

limiting it.

DR. FLEISHER: Allan.

DR. SIPERSTEIN: If we're unhappy knowing the mortality rate of each hospital, we're even less happy just knowing the volume. I mean it's even a looser surrogate of quality outcomes. And it kind of begs the question because we know if you go under the hood you're going to know what the morbidities and the mortalities are. Obviously, a structural measure like this gives us even less information or guides the patient with even less information in terms of how to seek care.

DR. FLEISHER: Thank you. Other comments? Are individuals comfortable with not voting and asking AHRQ to come back to us quickly with whether or not you want to develop a composite in this cycle which it sounds like we are open to? Robert.

DR. CIMA: Just one question for Patrick. Is there a specific reason to pinpoint? I know you're looking at esophageal resection

volume of all causes. And then you separate out mortalities.

What was the rationale behind doing that? That seems to me to that point is if you're saying there's a direct volume relationship, then the denominator should be the same across both. Even the title of that doesn't say esophageal resection for cancer patients. It seems like -- I'm just trying to figure out why.

Now I'm starting to get concerned that Allan's point is we're taking something away that may be useful for patients out there even though it's not a good marker. If you pair them together, it might be. But why was it done differently and probably in the future it shouldn't be.

DR. ROMANO: That's an interesting question. The answer to that question reflects the unreliability of the NQF review process.

(Laughter.)

No criticism. I edit a journal and I know unreliability of review processes. When

this measure was originally developed, the denominators were paired to be identical and to be limited to cancer patients. But in the previous endorsement review, the Committee felt that surgeons should get credit for procedures that they do for benign disease.

They felt essentially the volume, the total volume, for both benign and malignant disease was the relevant metric of volume. It's essentially the same surgical technique for benign disease versus malignant disease.

Whereas, for the mortality measure, there was a feeling that the cancer patients have particular challenges in terms of mortality. And that cancer in particular is a focus of public attention and referral.

And, of course, most of those procedures are done in affiliation with designated cancer centers. So there was an interest in keeping the focus on cancer for the mortality measure.

But I think that this is obviously an

issue where reasonable people could disagree. And I would say that from the methodologic perspective I'm always in favor of a broader denominator if it makes sense because we can deal with cancer through risk adjustment. And it strengthens the reliability of the measure as a whole.

DR HANDY: I was just going to make the point that when you look at esophageal surgery risk assessment with cancer patients that's a risk for doing worse.

DR. ROMANO: That's easy to handle through risk adjustment because it's well known the patient has cancer.

DR. EREKSON: I just have a theoretical or ethical question for the group to consider. And as a surgeon who does mostly elective surgery I think it's a really big point. It's not as necessary for cancer surgery.

But when surgery is very patient preference sensitive, we have to be very, very careful of structural measures that just measure

volume. That will really influence how the surgeon and the center counsel those patients on what their options are for treatment.

I really like the fact that these measures are tied to also outcomes so that this was a volume measure tied to an outcome. And maybe we could get into that the outcome shouldn't be mortality and maybe it should be morbidity and functional status which is a great direction to head in.

But I really, really think it's important on these volume measures that we consider having it linked to something else.

DR. FLEISHER: Thank you. I think that should be included in our report. And when the report gets written, hopefully you can ensure that that's accurately reflected in the report.

Other comments?

(No response.)

What I'd like to do is take a simple hand vote on who thinks it's important for AHRQ to bring back a composite or paired measure at

this time.

(Hand vote.)

It looks unanimous to me. I think from the standpoint -- and Carol and Patrick -- we would like to see it. And if it can come back within this cycle, I think there would be great interest in the Committee looking at this. I think that was a great discussion. Thank you.

MR. LYZENGA: Thanks everyone. Next we'll be moving to --

Also an AHRQ measure. So we'll ask our developers to -- We'll give them a moment here.

DR. FLEISHER: Background. I just wanted to say one thing, Patrick. Your observation about NQF we were just discussing was very helpful which is the reason of a standing committee for those of you who were not aware. They used to be ad hoc.

Hopefully, the people here can have memory as to the approach we take. For those if we do change our approach, people should call us

out on that so that we stay consistent.

That's actually -- Just to be aware, that was a Kaizen to actually -- This is one of the outcomes because of that reliability. So thank you for pointing that out.

DR. GUNNAR: Just to tee this one up, it is an endorsed measure as of August 2009. Dr. Romano.

DR. ROMANO: Okay. Now we're switching gears to a measure of morbidity. This is part of the patient safety indicator or PSI module. These measures key off of complications that arise during the care of patients in the hospital. This is a measure of post-operative respiratory failure. It's one of several measures of different types of post-operative complications.

This particular measure uses both diagnosis and procedure codes in a manner that I'm sure we'll discuss to identify patients who had evidence of respiratory failure after what's called a major operating room procedure. And it

is based on evidence of course that there are a variety of interventions that can reduce post-operative respiratory complications.

DR. GUNNAR: Any other comments? Discussants?

MS. THOMASON: Is Dr. Asher on the phone?

MR. LYZENGA: Is Dr. Asher on the line? I just want to check to see if he's available.

(No response.)

DR. DUTTON: Go ahead, Melissa.

MS. THOMASON: I just wanted to make sure if he was on the phone to give him an opportunity to weigh in. This is Measure 0533, Post-Operative Respiratory Failure. It's a measure of respiratory failure, mechanical ventilation, reintubation cases per 1,000 elective surgical discharges. It's a facility measure and it's intended to identify adult patients with particularly significant adverse events that are at least partially preventable.

I do have one -- I'm going to ask the developers really quickly for clarification. Does it only pertain to elective surgical discharges because you can clarify there was no respiratory distress going into that? Or why do we cut out the emergent procedures?

DR. ROMANO: Currently, the measure has a variety of exclusions that are intended to exclude patients who have respiratory failure before surgery or where respiratory failure is the indication for surgery.

There are also a wide variety of exclusions for patients who have underlying severe lung disease and heart disease. But there is no exclusion based on the urgency of the surgery.

DR. DUTTON: This is both an outcome measure and a marker for further bad outcomes. Post-op respiratory failure itself is painful and dangerous and highly costly. It's also associated with significantly higher subsequent morbidity and mortality.

This is an important thing to measure. The rate is high enough. There is a significant gap in care here and significant variability among centers.

DR. GUNNAR: Any other discussion?

MS. PITZEN: I don't know if this was brought up. I'm just going to continue the discussion about the exclusions. And this measure actually came up through our last phase.

MR. LYZENGA: Actually, can I cut you off?

MS. PITZEN: Yes.

MR. LYZENGA: Just because that's part of the validity discussion.

MS. PITZEN: Okay.

MR. LYZENGA: Take a vote on evidence first.

DR. GUNNAR: We need one more.

MR. LYZENGA: Again at the discretion of the Chair. I think we can probably call it.

DR. GUNNAR: There we go. We got them all.

MR. LYZENGA: All right. Ninety-six percent yes and four percent no. The measure passes on evidence. We can move on to opportunity for improvement or a gap in care.

DR. DUTTON: I thought gap was on the last vote. It was on the last slide.

MR. LYZENGA: Was it on the last slide?

DR. DUTTON: Yes, I thought we were doing those altogether.

DR. GUNNAR: Do you want to go back and vote again or? Anybody want to change?

DR. DUTTON: No, that's fine. I don't want to discuss it again.

DR. GUNNAR: Any other discussion regarding having this already been addressed on this? All right. Vote?

Always seems to take that little less. Let's see what it says if there's a -- One additional vote wouldn't have made a difference.

MR. LYZENGA: Seventy-three percent high, 27 percent moderate and zero for low and

insufficient. The measure passes on performance.

Now we'll move to reliability.

DR. GUNNAR: Any comments from the discussants on reliability?

MS. THOMASON: For reliability we run into the same conversation that we've been having. The reliability testing score was 0.744 of one. But it's significantly less reliable for lower volume hospitals.

DR. DUTTON: I am a little less concerned about the reliability based on the size of the hospital. Since this is a common enough problem, there will be plenty of data and very low volume hospitals are going to be ones that aren't dealing with that kind of patient. I don't think that's as big a deal.

If you read the measure specifications, there are multiple denominator exclusions in this. And I understand why they're there, particularly if the measure is going to be used for public accountability.

Some of them I think are very

reasonable and very objective. Others though, exclusion for specific co-morbidity codes of chronic heart or lung disease, that's going to be very hard to nail down. And I think it offers a large potential for gamesmanship in the measure. As I do my hospital coding, I may have a lot more patients with chronic lung disease as a result of having to report this measure.

MS. THOMASON: And just to echo Dr. Dutton, I think, from a reliability standpoint even with the low volume issue that the developers pointed out, it certainly isn't a concern in that area. But the exclusions were another thing that I wanted to hear from the developers.

MR. LYZENGA: And exclusions are actually I think in the validity again. For purposes of voting, we'll focus on reliability here.

DR. GUNNAR: So I guess to frame that correctly, could a hospital or is there any evidence that low volume hospitals use the

exclusion list to mask outcomes?

MR. LYZENGA: First, let's vote on reliability.

DR. GUNNAR: Yes absolutely. Did I capture that correctly for the record? Hearing no argument, I'll assume that that was correct. Yes sir.

DR. SAIGAL: Thank you. A question for our patient representative as well and basically this issue with low volume hospitals, is it for a consumer looking at hospitals and for us looking at how to identify problems, is it not more honest to not report data where hospitals have low reliability and just say the volume is too low to say? That in itself is information for everyone to use versus other approaches which I think might be providing information that isn't there. I was just wondering if that's a way to address the reliability issue with the low volume hospitals that people find interesting.

MS. THOMASON: I can't speak for everyone, but I will say that certainly from my

perspective I would definitely prefer for the low volume hospitals for it to say not reportable or not numerically significant or these facilities. Then I can say okay instead of adjusting the numbers or smoothing it over or bringing them to average so that I have the sense of they're all operating at the same level.

DR. GROVER: To follow up on that, I agree. I think an asterisk or something. You could report it, but you could say that the numbers were, however you want to phrase it, too small to know if this is a significant difference or something or not significantly higher.

MS. MOYER: We actually publicly report this. We're listed in the application as a user of the measure and that's what we do. We just have an asterisk and indicate there's not enough volume to really reliably calculate performance for this.

DR. GUNNAR: What do you use for your reference point for not enough reliable? What are you using? An own internal?

MS. MOYER: I believe the software does that for us is my understanding. I can check on that with my programmer.

DR. GUNNAR: Because the measure is capturing it. It's not making that distinction. It's incorporating all data regardless of whether that data came from low volume or not. It's not excluding low volume data. Is that correct?

But we are voting on reliability. And I think the discussants' impression is, at least as I hear it, is that their recommendation is that the number of low volume facilities does not impact the overall reliability of this measure due to the fact that it's a relatively high volume event. Any other --

DR. ROMANO: Maybe I'll just clarify a couple of things. One is that this of course is a strikingly different situation from the measure that we just reviewed in that the overall reliability here is 0.74. And it's not until you get down to the bottom 30 percent of hospitals that you get into this low reliability range.

For the great majority of hospitals this is a reliable measure.

Now how to handle that low volume situation, obviously different entities in the public reporting space make different choices. And AHRQ software tried to support a variety of choices that users may make.

Some users choose to implement a threshold. I think CMS for example in its measures has generally used a threshold of 30 in the denominator if I remember correctly. Some users, other users, have adopted that. But that is a matter of user discretion.

I do want to explain that there is a conceptual and statistical advantage to this approach of using all the data and doing this shrinkage. And that is it gets back to a nice example that a famous statistician described, Ephron, related to batting averages.

If you imagine early in the season somebody has just come up to the major leagues and their batting average is based on a limited

number of times at bat. And maybe their batting average is 0.500. And you say, "Wow, this is an incredible baseball player." And you pay that person a lot of money. But then as the season goes on, the batting average drifts back to the average, so-called regression to the mean.

We often see this phenomenon in practice where there's a period of time when a hospital looks really good and it drifts back. So we know that we're always using data that are a year or two old to provide information to the market about current performance. That's the nature of the beast here. What we want to do is to maximize the correlation with current performance which of course we won't measure until a year or two from now.

It's been conclusively shown that to get the best prediction of current performance we would use both the information about the hospital's own performance in the prior period as well as the information from other hospitals, because on average, hospitals will tend to

regress towards the mean.

In fact, this statistical approach, it's not just mumbo jumbo. It's based on the practical issue that we're trying to give the field the best information that we can about what's really going on right now. But we don't know that yet.

The best way we know that is by looking at how that same hospital did a year or two ago and how other hospitals did, knowing that there will be a tendency for hospitals to regress towards the mean of all the hospitals that are in the population of interest. That's the conceptual approach.

DR. GUNNAR: And I appreciate that. Just to get back on to reliability and this vote though is that the discussions about volume and the reliability and the ability to predict current events, it's easier in higher volume events. And this is a relatively high volume event. Can we have a vote?

MR. LYZENGA: Let's go ahead and vote

on reliability.

There we go. So we have 30 percent for high, 65 moderate, four percent low and zero insufficient. The measure passes reliability. So we'll go onto validity. And this would certainly include concerns about the exclusions. Maybe we can start with Collette.

MS. PITZEN: Great. Thanks. Now the right time for discussions of exclusions. This came up during our last cycle. Our small group actually had the opportunity to review. I have the same feedback that I had back then.

I'm concerned that some of the exclusions for this measure are a bit broad, in particular, MDC 5 diseases/disorders including procedures of the cardiovascular system. This would exclude every coronary bypass patient, every pacemaker. I think that there needs to be some careful thought in terms of having that as an exclusion.

I feel the same way about the MDC 4, although perhaps not as strongly. Just for

background, an MDC is a major diagnostic category that's a grouping, a very large roll up of the DRGs at the hospital base. So I would hope that there would be some careful consideration of exactly what types of patients you're pulling out of the measure.

DR. DUTTON: I already expressed a concern about the exclusions. And I wanted to say something about the risk adjustment model because I think that's an exception as well. I guess I favor -- I think that we do have to risk adjust measures when they're going to be used for public accountability. So payment is on the line. It's important to level the playing field.

But at the same time risk adjustment has the risk of throwing away the very data that we need to take meaningful action and just the very simple observation of patient age. All of these measures are adjusted for age.

But I would actually like to know that my older patients have a higher risk of post-op pulmonary dysfunction because I may concentrate

more of my hospital's resources on that in order to improve that. I may have more physical therapy with my old patients, etc.

I just want to raise the concern. This applies to all of our risk-adjusted measures. As the CSAC starts to look at tiering of measures or uses of measures, we may have a different model for measures used for public accountability versus measures used for quality improvement. And one of the big differences would be how much or how hard you risk adjust.

DR. GUNNAR: Amy.

MS. MOYER: I had a question on the exclusion of the MDC related to obstetrical care. And I was curious if that would exclude caesarian section patients from this measure. If so, why?

DR. ROMANO: I can address some of these exclusion questions and I just want to say joining me here at the table now is Dr. Garth Utter who's a trauma surgeon and critical care surgeon on our team. And he's written some of the papers related to this indicator. He'll

participate as well.

We actually share the concern that some of these exclusions are a bit broad. They were to some extent inherited from a progenitor version of this indicator which was developed by Lisa Iezzoni and her group under the rubric of the complication screening program.

We went through two rounds of expert panel review with AHRQ with multidisciplinary panels including surgeons, critical care physicians, nurse anesthetists, anesthesiologists, so forth. And there is a tendency frankly in these panels reviews for people to throw out exclusions and say "Oh, you ought to exclude these people. You ought to exclude these people." And then the exclusions start to add up. So this is a legitimate concern.

But I do want to point out that the MDC exclusion is based only on the principal diagnosis. We're not excluding people who had a co-morbidity of severe lung disease or a co-

morbidity of severe heart disease. We're excluding people who were admitted specifically for treatment of lung disease or heart disease.

The idea is that if the patient is being admitted for example for a lung cancer resection, then presumably the surgeon has had some discussion with them about the risk. And they're undertaking this procedure aware that by losing some of their lung they may go into respiratory failure. And similarly patients who are undergoing a procedure, for example a transplant procedure or a procedure for severe heart disease, they're doing so obviously recognizing that we're operating on the chest. And that confers a very high risk of respiratory failure.

These MDC exclusions are based only on the principal diagnosis, why the patient was admitted to the hospital. You're right that obstetric patients are excluded across the board from these measures.

Part of that is because obstetric

patients are subjected to different coding rules. And of course post-operative respiratory failure is a very rare event in the obstetric setting. It would require modeling in a completely different way. It's such a different patient population that it just made more sense to exclude it.

Garth, do you want to add anything on the exclusions?

DR. UTTER: Only the point that some of the exclusions concern airway protection issues. And this is a particularly challenging issue. It creates a conundrum of trying to avoid penalizing people for doing something that is really intended to prevent the problem in the first place. That's the only point I'd add.

DR. YATES: I just have a technical question for the developers. In one of your previous measures, you used the composite risk adjustment variable of an APR DRG. In this measure, you use specific co-morbidities. Is that because there would be so many APR DRGs? Or

is this an older measure? Or why the difference in methodology?

DR. ROMANO: Yes, two reasons for that. One is that the APR DRG system that we use is the risk of mortality or ROM classification. And that is designed specifically for predicting mortality. That would not be exactly appropriate for a morbidity complication of this type.

Now there is a different tool in APR DRG. It's called the severity of illness subclass. But that's really designed more to predict length of stay. And again this is not exactly length of stay.

The first reason is a tailoring. The APR DRG approach is tailored to the outcome for the mortality IQIs. It's not tailored to this outcome. We felt that developing a tailored model approach was better.

The second reason is that for the patient safety indicators we want to be particularly careful about not adjusting for things that arise during the patient stay in the

hospital. We actually discussed earlier for hip fracture how using the APR DRG framework we are adjusting for the procedures that the patient had regardless of when those procedures were performed. And we accepted that for hip fracture because we know that certain types of hip fracture require certain types of procedures.

In this case, we have a very, very broad range of procedures that are being done and it would as you point out lead to a large number of APR DRG parameters to adjust for all those. And some of those procedures are being done to treat complications that actually arose during the hospital stay.

Now it would be really difficult to sort out was the procedure being done to treat the hip fracture. Well, we know the hip fracture was there on admission. But was the procedure being done to treat something else that was happening during the hospital stay?

So APR DRGs implicitly adjust for procedures. To avoid adjusting for procedures,

we only want in this case to adjust for the main reason the main procedure why the patient came to the hospital and not for all those secondary procedures. That's the other reason we're not using those.

DR. YATES: Then to follow up on that, those co-morbidities that you are capturing ideally those would be captured as present on admission. Now how do you separate present on admission and this administrative dataset collection? Some of those co-morbidities may develop during the admission and be collected retroactively or after the event.

DR. ROMANO: Correct. Right. The hospitals are required to report in the case of Medicare on CMS claims. And most state health data agencies also require hospitals to report for each diagnosis whether that diagnosis was present on admission or not.

Those of you who are in the clinical documentation improvement side know that sometimes there is disagreement and sometimes the

coders are unclear about whether the condition was actually present on admission or not. And we've been doing some sensitivity analyses to explore that further.

Fortunately, it doesn't really seem to make a difference in the modeling. So there are some patients where it appears that the diabetes developed after admission. We know it didn't. But it turns out that that kind of miscoding is so uncommon that it doesn't make a difference in the modeling.

MS. McCARTY: The explanation around the reason for the exclusions makes me wonder with everything that is risk-adjusted for -- I guess two questions: One is if we can risk-adjust for all these things, why can't we also risk-adjust for the MDC 4 and MDC 5 indications?

And if that isn't a possibility then it sort of indicates that maybe there needs to be subsequent measures that just focus on those groups. Because to say that it's a high risk and known outcome, I think we care just as much about

knowing to what extent that that does occur and if we can improve the frequency of those known risks happening. So potentially coming back with a measure structured like this but with just that group as the focus.

DR. GUNNAR: But marked as a gap. Right, Andrew?

MR. LYZENGA: Yes.

MS. THOMASON: As is, the exclusions, post-op respiratory failure for any coronary bypass patient, it's excluded. Right?

I would love to hear other clinicians in the room weigh in. I know that certainly having tons of surgery doesn't qualify me to weigh in from a clinical perspective especially you, Dr. Dutton. You know, from a clinical standpoint, is that valid?

DR. DUTTON: My take is Kelsey's thing. I think I would favor a lot fewer exclusions and then risk adjustment around things like cardiac bypass surgery or pre-existing COPD or what have you.

DR. GUNNAR: I think that may be for future to refocus on this measure and the validity associated with the data. Any other discussion on the validity?

DR. GROVER: Bill, it was brought up earlier about the importance obviously of the co-morbidities being captured at admission.

One problem that was revealed back to us in the 80s was that, at least in our specialty, there are a fair number of things that can happen after admission but before the operation. People can go into shock, whatever.

They can be taken to the operating room where if you just capture these risks on admission people can really deteriorate, and it doesn't capture their risk as they go into the operating room.

I'm just curious how you handle that. You could just say when that appears in the chart as long as it's before the operative day rather than admission.

DR. ROMANO: Garth reminds me that one

of the limitations in the denominator for this measure is that it is limited to surgical discharges that are identified as elective. So I may have misspoken earlier on that question, so that addresses Dr. Grover's comment.

DR. GUNNAR: Just to reframe it for myself in relationship to -- this is a measure where the denominator is a group of patients who the expectation of respiratory failure is low.

I appreciate the other comments about the exclusions, but from a quality improvement sort of environment of care perspective, I personally don't have an argument with the way this was created.

Other measures should look at that excluded population but potentially for a different purpose.

DR. ROMANO: I think AHRQ would agree with that, that it would make sense to look at cardiopulmonary operations separately, but obviously those patients are much more like to come in with some degree of respiratory

compromise and that would have to be carefully considered in the adjustment approach.

DR. GUNNAR: Lee and then Dr. Cima.

DR. FLEISHER: One of the things, getting back to Rick's comment, this is part of PSI 90 but unweighted, which is interesting how that happened which I think gets to reliability because that is used in the current payment model but not used. I haven't figured that one out yet.

Can you comment as to what happened to this measure as it became part of the composite? Do you have any clue?

DR. ROMANO: I'm just conferring with staff. I'm not trying to be evasive. It's just that the issue is that the composite measure is a different measure. It's being reviewed by a different committee, and so we are trying to keep the processes separate.

I would say that PSI 90 was endorsed before PSI 11 was endorsed. This is why PSI 11 is currently unweighted in PSI 90. It's the

original endorsement by NQF of PSI 90 predated the original endorsement of PSI 11.

DR. FLEISHER: Thank you. That's sufficient.

DR. CIMA: I just want to bring up probably the dirty little secret about PSI 11. And it's not about how you wrote it or how you did it, but as you know the reality is that I have visited a number of institutions, sent down by my institution to go say, boy, they did great. Their PSI 11 dropped like this.

And the reality is when you go there it was all because of how they changed their documentation. It had nothing to do with the number of patients that were on ventilators, and I think everyone including the developers know that that's a significant portion of this problem here.

When we talk about validity, that's not captured in the measure, and I'd like the developers to discuss the reality of what this whole process had done now. It's that people are

able to significantly alter this PSI by just how they describe the post-op patient on the ventilator.

DR. UTTER: I think it's safe to say this is a topic we've given considerable thought to, our team.

And in summary, approximately 20 percent of records flagged by this indicator are flagged on the basis of the diagnosis codes only. They do not have one of the procedure codes for either the length of intubation or reintubation.

I don't know that there's any clear cut right direction to go on this. We have considered just eliminating the diagnosis code component. However, we realize that that will probably leave uncaptured some cases that truly represent quality deficiencies.

DR. CIMA: To follow up, how can you address the validity then? On face validity, there's a problem.

DR. UTTER: Yes, I think we have to admit that there is an issue with the validity

because of the diagnosis code options basically.

This is not the only code that might be used. There are other codes that might be applied to the same clinical phenomenon that will not trigger this indicator. We acknowledge that.

DR. CIMA: My point is that when we talk about reliability that's one thing, but to go to the public and say we have a measure that has a validity issue you can't put an asterisk by that. It's either valid or not.

And that's been my big -- I mean I was shocked. You know, 50, 60 percent improvement in people's PSIs literally within a quarter or a year based upon just what they told people to document. The nurse practitioners in the ICU never write this down for this patient.

I mean I'm not saying it's ethical. We're not talking ethics. We're talking validity of a measure that's going to be used for public reporting, and I have very real concerns about the validity here.

And that's the only concern that I

have about it. The exclusion I agree with, William and everything, but I think that's a real issue we need to put up front.

DR. GUNNAR: I saw Melissa, then Kelsey, then Dr. Yates and Dr. Grover.

MS. THOMASON: Patrick, did we go back to ---- it said the denominator does specify elective procedures? Correct?

So what was the train of thought behind that? Is it because if it's not elective we don't have a standing respiratory point to compare it to or?

DR. ROMANO: It was Dr. Gunnar's point that these are patients who should be very low risk, where there's no expectation of respiratory failure after surgery. So it may provide a quality improvement opportunity.

DR. GUNNAR: Kelsey.

MS. McCARTY: Just to the point about the workarounds, I think it's related to a comment that Amy made earlier.

By making things a measure, no one is

surprised that those behaviors start to happen and I think it's a known thing across a lot of measures. Maybe in the first quarter or in the first year or maybe in the second year, they're able to get those huge improvements, but then after year two they can't -- that kind of exhausts itself.

It maybe reestablishes the baseline, and there are some behaviors in there that we don't want, but I think those workarounds will happen. I think it just maybe reestablishes the baseline, but then there's still potential for improvement from there forward.

DR. GUNNAR: Microphone, please.

DR. CIMA: To allow that to go on, that's a problem. You're facilitating them to do that.

Let's say if they go from a high outlier to as expected, everyone's happy then, but you're still having a whole bunch of people that aren't treated, but they now have the justification of saying, well, we're doing fine

on this PSI.

That's my only point. I mean I'm not trying to -- but I'm just saying this PSI is inherently one of the ones that can be gamed masterfully. And you're not helping people. That's the point. It all becomes then just a display. It becomes window dressing.

DR. GUNNAR: Dr. Grover.

DR. GROVER: With this question having been raised, I guess I'm wondering what your audit process is. Do you audit? How do you know the data is accurate? What are the consequences of gaming the codes?

DR. ROMANO: I can address this in several ways. There have been a number of audit studies that have been done. They're cited in the validity review. Dr. Utter has authored a couple of them as well as the VA, Dr. Borzecki and colleagues from the VA.

One thing that we learned -- again just to sort of address Dr. Cima's question. So, right now there are two potential ways of

identifying patients with respiratory failure. You can identify them based on procedures that they had, a reintubation after they were extubated from surgery or they had a prolonged course on ventilation. So you could use procedure codes or you can use a diagnosis of acute respiratory failure or now an acute post-operative respiratory failure.

Now in the design of this indicator we chose to use both, and we did so based on an earlier validation study in which we used NSQIP data from the VA, VASQIP data, that were linked to the VA's patient treatment file, and we showed that if we used the procedure codes alone, the sensitivity of the indicator was suboptimal.

So to get this balance between false-positive/false-negative error we had to use or logic to capture either the diagnosis or the procedure. Fundamentally, the postoperative respiratory failure sometimes is treated without mechanical ventilation. Sometimes it's treated with very high FiO2, with CPAP or other kinds of

interventions.

This is the problem. So when we tried to apply a clinical definition from the VASQIP program, we found that the best balance between sensitivity and positive predictive value was to use this or logic, using either the diagnosis code or the procedure code.

Now we have about 20 percent of the numerator cases that are identified based only on the diagnosis code. I find it hard to believe -- I mean it's not impossible, but I find it hard to believe, given that 80 percent also meet the procedure criteria, that there are more than a handful of hospitals that could reduce their rate by 60 percent just by addressing the coding of the diagnosis.

Maybe there are a couple out there at the extreme, but overall we have only 20 percent of the numerators being captured by the diagnosis codes. The great majority are being captured using procedure codes.

And as part of the audit study, we did

look at a stratified indicator of positive predictive value -- what we're going to call the marginal positive predictive value, which is what is the PPV among the patients who have only one criteria.

The patients who had only the diagnosis criteria, their PPV was 79 percent versus the patients who had the procedure criteria, their PPV was 83 percent. Again, we concluded that the difference between 79 percent and 83 percent was not high enough at this time to exclude the diagnosis code from the definition of the indicator.

It was the results of the audit -- of a series of audits, that have led to this kind of incremental process of trying to refine the indicator definition, but this is an ongoing process and of course we welcome input from the field.

DR. CIMA: The only thing I would say is if you had asked the U.S. Army what an IED was ten years ago, they had no plan for it.

So basically all these audits were done, when this was developed -- which was almost 10 years ago, if not more. These insurgents have developed a way around it and you have no data to suggest that's not the case now, and that's what I'm saying.

That's the problem with it, is when these were developed in all good honesty and integrity and everything and you guys were looking at it to develop it through an already existing database, that's great to develop it, but I'm just telling you what the reality is in the field. And you have no data to suggest otherwise now, that it's not being gamed, and we know it's being gamed.

I'm just saying that's the problem. This has been recycled and recycled multiple times. Circumstances on the ground and in the field are different, and now we have to understand whether it's valid. And I don't think there is any data to suggest that it is.

DR. GUNNAR: The question I had raised

for the force of the Committee and the way forward is in evaluating the validity, do we evaluate the validity in perspective to its raw, granular data or do we evaluate the validity in relationship to the intent of the measure?

That there is enough data, it may not be perfect, but does it drive improvement? Or are we in an isolated way just looking at is this valid information? I'd like a little on that, and then Dr. Yates.

DR. YATES: To address that question, I think you look at the validity in terms of the intent and whether or not it has some -- it captures what it's supposed to.

I think the granularity is critically important, but I think that's part of the -- these measures are all going to evolve, and how they're responded to is going to evolve. And that evolution over time is something that has to be captured and looked at.

When you see dramatic evolution, you bring up the things that Robert has brought up,

and you address that and perhaps by better capture.

This is a public forum. I just wanted to say that not every change in reporting is because of people being devious or trying to hurt people or trying to hide things, which could be implied, but I know you're not implying it.

But the fact is that regionally some hospitals and locally some hospitals will have the wrong language, and their coders pick up on the language the wrong way. If they adjust the language or they adjust how they're dictating or how they're recording things, it doesn't mean that they've changed their practice. They're just better at coding.

There's a certain regression to the mean that occurs with people learning that coding counts. That may be part of this as well and that's -- you're going to see that with the puncture and laceration one that we're not talking about, but if you add the words that this was inherent to the procedure in entering

something it eliminates a lot of punctures and lacerations.

DR. FLEISHER: Yes, I actually would be curious. We want to be consistent. I don't know if the staff has any comments, but the second part is I think you've sent a clear signal that we should make sure in the report that when this comes back if this is approved they have to do that testing is maybe a comment that I'm hearing to ensure for threats to its validity in the future.

And we should be clear in the report if that's the comment. I just want to be consistent across how we look at validity beyond us.

MS. WILSON: I think you're making an important point, Lee, which is if there are threats to validity, if there is something that you want the measure developers to consider, we capture that in the comments of the report, and it's a way of informing.

And just as AJ said, the measures are

evolving, and this is a way of informing them and help drive that evolution and to keep these issues in front of us as we continue to look at these measures.

DR. GUNNAR: Dr. Yates, one last comment.

DR. YATES: No, that's all.

DR. GUNNAR: Okay. Can we go ahead and vote on validity then? Amy looks like she can't.

MS. MOYER: I was just asking ---- related to what we're asking to be captured, are we asking for anything that's measured with a nonmedical record source to be validated back to the medical record? Is that the level of validity question we're asking? I just want to make sure we're clearly stating.

DR. GUNNAR: I think all of that is fair game, but you as a committee member voting on this, you vote on it with the knowledge that there may be issues with regard to validity that make it your rating of high, moderate, low or

insufficiency, but also don't forget the impact that it's that intent of the measure in relationship to that data. Does that make sense?

Can we vote now?

MR. LYZENGA: All right. So we're voting on validity. Your options are high, moderate, low or insufficient.

I think that looks like all we're getting. We can close it out.

We've got five percent high, 59 percent moderate, 36 percent low, zero percent insufficient. So that passes.

And you can skip that. Skip that, too. I think we're on over -- here we go. Feasibility. So feasibility of implementation is what we're considering here.

DR. DUTTON: It's drawn from administrative data. So it's certainly feasible to get that out of records. The fact that you need a lot of different codes to know whether the case is in or out makes it harder. So the burden is relatively high, but I would say it is

feasible.

DR. GUNNAR: Hearing that, any other discussion? Seeing none, go ahead and vote.

I think Cliff was out. So do you want to vote on this, on the feasibility?

I think we go ahead and record it and see.

MR. LYZENGA: Fifty-nine percent high, 41 percent moderate, and it passes on feasibility and we move to usability.

The extent to which audiences use or could use performance results for both accountability and performance improvement activities.

Any comments before we vote? It doesn't look like it. Let's go ahead and vote?

MS. THOMASON: Is this -- I have a question. The questionable validity and all the conversation we had surrounding that, does that come into play here at all, to say, at the end, is this a usable measure? Or do we negate all of that validity conversation and say --

DR. GUNNAR: I might frame it the other way. What I hear is that yes, they're using it. They're reacting to it and potential reaction to it may actually diminish its validity if you exclude your patients.

And then the last comment which came from Kelsey was you can only do that so long. At some point, you're going to have to react to the actual people that are being included and who are being measured.

Can we restart usability please unless there are any other comment? Yes, Dr. Romano.

DR. ROMANO: Just again we heard two things loud and clear that I think we'll take back to the whole team.

One is that there is a desire and interest to have a broader denominator that would include a larger set of patients, including patients with underlying lung disease or heart disease. That may require a different measure or different stratification of the measure.

The other thing that I think we heard

loud and clear is that there's a need for updated evidence regarding validity. We can't rest on the evidence from five, six, seven years ago. And that before this measure comes back again that NQF will expect a revalidation of the measure to reflect current practice.

DR. GUNNAR: We'll go ahead and vote on usability.

MR. LYZENGA: Do we get numbers here? All right. We're going to have to go back and examine that one, but it looks like it's a pass. So now we're at overall suitability for endorsement.

DR. GUNNAR: Any final comments? Hearing none, please vote.

MR. LYZENGA: All right. Does the measure meet NQF criteria for endorsement? Yes or no?

We have 91 percent yes, nine percent no. The measure passes, and before we break for lunch, we wanted to make sure to take a moment to get public comments.

Operator, can you let us know if we have anybody on the line to make public comment?

OPERATOR: Yes, sir. If you would like to make a comment please press star, then the number 1 on your telephone key pad.

There are no public comments at this time.

MR. LYZENGA: Okay. Is there anyone in the room who'd like to make a public comment? It does not appear that way.

So I think we are going to at this point set aside Measure 236 for the moment and allow you to grab a bite to eat for lunch, but we're going to do a shortened lunch. Just about 15 minutes and then come back.

You can eat while we're continuing our work, but we'll jump back into things around 1:00 p.m. So go ahead and grab some food and come back to your places around 1:00 p.m. And we'll get back going. Thank you. Off the record.

(Whereupon, at 12:43 p.m., the meeting went off the record, and resumed at 1:08 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

(1:08 p.m.)

DR. FLEISHER: Our next measure -- we've done public comment. I think we're doing great. The next measure is actually -- I think 0236. Correct?

DR. GUNNAR: Yes.

DR. FLEISHER: I'm going to recuse myself because I worked on this I think originally. I was on the TEP. Is this -- which measure is this? Was this developed by Pennsylvania?

PARTICIPANT: Yes.

DR. FLEISHER: I was on the -- I chaired the original TEP. So perhaps you should -- and I should recuse myself from this measure.

DR. GUNNAR: Do we have it? The measure is 0236, CABG: Preoperative beta-blocker in patients with isolated coronary bypass graft surgery. And CMS is the developers. If you would like to begin.

DR. BERG: Sure. I guess as you did

this morning you want us to introduce ourselves and we have some people on the phone as well.

I'm Sven Berg, I'm the Chief Medical Officer at the West Virginia Medical Institute. We are the parent company for Quality Insights of Pennsylvania, who is the original measure developer for this measure, and with me is Jane Lucas.

MS. LUCAS: Hi, Jane Lucas. I'm a project manager at Quality Insights of Pennsylvania.

DR. BERG: And we have some folks on the phone as well. So I'll ask them to chime in and introduce themselves.

MR. CRAWFORD: This is Al Crawford from Thomas Jefferson University.

DR. BERG: Thank you, Al.

MS. DETLANA: This is Hiro Detlana with Quality Insights of Pennsylvania.

DR. BERG: And, Gary, are you on the phone?

MR. REZEK: Gary Rezek, yes. Thank

you, Dr. Berg. My name is Gary Rezek. I work for Quality Insights of Pennsylvania.

DR. BERG: And our person from CMS actually had to drop off because she had another call at 1:00 p.m. Quality Insights of Pennsylvania will represent CMS today.

I was going to open with something, since we were originally scheduled to be doing this this morning. I was going to say something about being the only thing between you and lunch, but I don't have that at this time.

Now I can just say that I remember back to medical school days and eating on the run when I was doing surgery rotations, and I see that nothing has really changed from that either. You were only given ten to 15 minutes to have lunch before reconvening, but we appreciate being here this afternoon and having the opportunity to present this measure.

On behalf of CMS and the measure developers, Quality Insights of Pennsylvania, I'm pleased to reintroduce NQF 0236: Preoperative

Beta-Blocker in Patients with Isolated Surgery for consideration for NQF reendorsement.

The measure was first implemented in the Physician Quality Reporting System, PQRS, in 2007 in an effort for specialist to report measures that address the relevant clinical strategy. Since then the measure has been expanded to include use by anesthesiologists.

The intent of this process measure is that a beta blocker would be received within 24 hours prior to an isolated coronary artery bypass graft surgery. The denominator of this measure is isolated CABG surgeries for patients 18 years and older.

The reporting requirement is each time an isolated CABG procedure is performed during the 12 months reporting period. And this is done by way of administrative claims or a registry.

The American College of Cardiology Foundation and the American Heart Association 2011 Clinical Guidelines on Myocardial Revascularization support the use of beta

blockers and the use of them administered at least 24 hours before CABG to all patients without contraindications to reduce the incidence or clinical sequela of postoperative atrial fibrillation.

Postoperative atrial fibrillation is a common complication following cardiac surgery occurring in 25 to 40 percent of patients and has been associated with increased rates of postoperative morbidity and mortality and consequently increased costs.

The prophylactic administration of beta blockers has been shown to reduce the risk of postoperative atrial fib and mortality following isolated coronary bypass graft surgery, and a review of the literature revealed that there was an 19.5 increase in preoperative use of beta blockers from 2000 to 2009.

We appreciate the opportunity to review the measure with you today and we look forward to your comments and questions.

DR. GUNNAR: We have Dr. Olsen and Dr.

Roth.

DR. OLSEN: If we just start with the evidence, the risks of postoperative atrial fib, about 25 to 40 percent. And as we've already heard, there are clinical practice guidelines that recommend that beta blockers be administered anywhere from 1B to 2B to 1A recommendations.

I would say that there's also a 2012 meta-analysis that says although there's a substantial risk reduction that it was not statistically significant. Another paper published in 2014 said although the burden of supraventricular arrhythmias can be reduced there is really unclear evidence on mortality, AMI, stroke, heart failure, hypertension, bradycardia, and a couple of other new papers with atrial fib were not statistically significant as well.

DR. GUNNAR: Dr. Roth.

DR. ROTH: Not much to add to it. Again, the concern is that actually disparity of the literature, although the literature that was provided supported mostly Class B evidence.

That's not surprising considering that the meta-analysis not being statistically significant certainly caught my attention.

DR. SIPERSTEIN: I'd just like to comment in general about beta blocker payment. Obviously, some of the initial studies on beta block paid for noncardiac surgery were exceptionally optimistic, and then a whole host of subsequent studies have basically come to an opposite conclusion that it really plays minimal if any risk on perioperative morbidity and mortality, and obviously a lot of those measures have retrenched as well.

I guess the question I really have is for those have studied the literature in detail what is the evolving state of the art in terms of the literature, in terms of whether this practice really improves morbidity and mortality despite the fact that at its inception various societies endorsed it.

Yes, it's become more of a routine part of practice, but I'm really interesting in

whether the follow-up studies have really continued to support its use.

DR. GUNNAR: That is an interesting question, and Fred should address this as well and Dr. Handy. The overall mortality risk now in elective coronary bypass grafting is remarkably low. It's one percent or less. And from the STS for VA's data from whatever you want to look at, that's come down substantially in the last 10 years.

The question is where does any single component of that play in, and does beta blocker use 24 hours beforehand, what part does that play? And I think you're never going to be able to distinguish that at this point forward. And I'm not sure there would be -- would anyone ever put a randomized control study of -- nowadays no beta blocker versus beta blocker?

DR HANDY: The literature is best vis-a-vis strictly A-Fib for amiodarone. Most programs are focused on amiodarone, not beta blockade.

DR. GROVER: You're right. I think there was more initial enthusiasm for this than now. Having said that, however, I'm not quite sure that the administration preoperatively -- the way it's being done is the way envisioned by the ACC/AHA guidelines.

And the measures -- and we struggle with this in our STS measure a little bit too. It's hard to show a difference in mortality and yet if somebody goes into atrial fibrillation postoperatively, they have a lower ventricular rate and they're -- I think, more easy to control. You usually do can most of that by just increasing or modulating up the beta blockers.

But it is right now still a guideline, ACC/AHA guideline, and I think it's useful from that standpoint, but probably we need to determine how many of these patients that are say they have pre-op for either you or us get it in a very short period of time and how many are loaded over a day or two beforehand. In other words, the dosing, it's not just being done

acutely right preoperatively is an issue.

I mean you're an anesthesiologist, I might throw this back to you, too. What are your thoughts with the anesthesia?

DR. FLEISHER: Actually, Chair, for the last two versions, the noncardiac guidelines. We've written extensively on this and I don't think you should extrapolate from the noncardiac to the cardiac. So that's my only comment.

DR. SIPERSTEIN: That wasn't my intention at all, but simply the fact that there is some initial reports that are very enthusiastic guidelines get set and then subsequent studies may not support them.

We saw a little bit of that here in that we've seen some subsequent studies that did not seem to show a statistically significant difference.

DR. FLEISHER: There is actually a paper out of Mark Neumann from Penn showing how frequently ACC/AHA guidelines get modified. And it's only in a level of evidence B class 1 it's

about seven percent in subsequent guidelines.

But you're right. As of now, I know the guidelines exist.

DR. GROVER: I wouldn't change things right now. I think it's still a personally useful measure. I just think we ought to define it a little better and be sure that it's being given as recommended in the guidelines.

DR. OLSEN: I can stand corrected, but I believe this is an extension of the 0127. And now the proposal here is to expand it into the outpatient area as well.

DR. FLEISHER: I think you're talking about the SCIP measure because this I don't think is the outpatient. There is an outpatient measure, but that's for noncardiac surgery.

Can we look up 0127?

DR. ROTH: Actually, is it 0117, the beta blocker at discharge that we're referring to? And you have to help me with this. On the agenda, there's also a harmonization discussion. Is that what it's referring to is 0117.

MS. FELDMAN: You have a document that was provided in your copies that is a measure comparison table, and the first page is 0127 and 0236.

MR. LYZENGA: Remember, we're discussing this measure first and then should we recommend endorsement of it, then we'll move onto the discussion of harmonization with 0127.

DR. GUNNAR: And just remind me. 0127 comes up this time or is it a endorsed measure that's currently endorsed?

MR. LYZENGA: Currently endorsed.

DR. GUNNAR: Okay. So the harmonization is not another measure that we will consider. It's to a measure that's currently endorsed. Okay. We should bookmark that, address this separately.

Any further discussion on evidence? Amy.

MS. MOYER: My question was more on the outpatient item. I believe what this is is it's not a measure of things happening

outpatient, but the claim source is outpatient from the physician practice group. Is that correct?

DR. BERG: This was originally developed as an outpatient measure and is an outpatient measure.

DR. GUNNAR: Any further discussion? Let's vote. Technical difficulty.

MR. LYZENGA: Just wanted to make sure we had the right slide. In fact, we've come to a decision that we're going to leave that fifth option off for now.

If it comes up that we have insufficient evidence in the voting on the first four options, then we'll consider an exception, but right now, we'd ask you just vote on those first four: high, moderate, low or insufficient evidence.

DR. GUNNAR: For this process. We'll miss one. So 22 is a complete vote.

No one else is recused, right? I think that's right. All right. Here we go.

MR. LYZENGA: Did we get results? Okay. I think we're going to need to revote on that one. If you could cast your vote again on evidence.

I think that was it. It will be a 21 if you can stop it there. Oh, 22. All right. We have 27 percent for high, 55 percent for moderate, 14 percent low and five percent for insufficient. So the measure does pass on evidence, and we can move on to opportunity for improvement.

Again, this is whether a performance gap demonstrates quality problems and opportunity for improvement or overall low performance.

DR. GUNNAR: Dr. Olsen.

DR. OLSEN: In 2012 we were at an average of 95.5 percent compliance rate. Rural with 98.2 versus urban 91.8, and 31 percent of the providers were reporting.

It certainly appears the high compliance rate with what's currently being done.

DR. GUNNAR: Dr. Roth, any other

comments?

DR. ROTH: Nothing to add.

DR. GUNNAR: I'll be proactive. Is this measure topped out?

DR. SAIGAL: I would say that it's not because you've only got one-third of the people reporting.

DR. DUTTON: Yes, exactly my point. It's topped out among people who reported.

DR. YATES: Yes, words out of my mouth.

DR. GUNNAR: Let's go to the vote. Is there any other discussion?

Okay. This is performance gap on this process measure. There we go, 22.

MR. LYZENGA: I think we can close it at this point. So we have 23 percent high, 45 percent moderate, 32 percent low, zero insufficient.

So the measure passes on performance cap, and we can go ahead and move on to reliability. It's the first part of scientific

acceptability.

DR. OLSEN: In reliability testing, the averages were about -- the reliability score was 0.85 with a 1.0 max from the registry reporting and 0.99 from claims-based reporting. Of course, this is based on the cohort that was reporting to the registry.

DR. GUNNAR: Any other discussion? We'll go ahead and vote.

DR. ROTH: It's the same situation as the reliability of those that are reporting.

DR. GUNNAR: We will go ahead and vote. Yes. I think we collected the data and it's reliable.

MR. LYZENGA: Did we get a result on that? All right. I think we have to retry again. Sorry, one more vote on reliability.

Is that it? Can we close it up? I think we're having technical difficulties again. Sorry, we thought we had a smoother system this time. It's not turning out so great. Maybe we can take a hand vote on this one again.

So let's see a count of hands for reliability -- those who think it has high reliability.

I've got five for high. Okay, and moderate.

PARTICIPANT: One vote from the phone.

MR. LYZENGA: Okay, 15 for moderate, and low.

One for low, and insufficient. Zero. So that passes on reliability. We'll move onto validity.

Are there any comments on validity? This is whether the specifications are consistent with the evidence and the measure yields credible, valid results about the quality of care. This also includes exclusions, risk adjustment and other threats to validity.

DR. OLSEN: I think as a process measure it does what it's set out to do and capture the number of patients that receive beta blockers, and exclusion was left up to the prescribers. About 4.4 percent of patients were

excluded from the denominator.

DR. GUNNAR: Any discussion?

MS. PITZEN: Just a question and a comment. Let's see.

The validity testing was from several years ago in 2010 and the inter-rater agreement rate between what was submitted and what was abstracted was at 64.2 percent. That's a little bit on the low end and would tend to lead one to maybe question the ability or feasibility to collect that data, and then an additional comment.

This is just in general. I know many of the PQRS measures have a wide, open physician-can-document-any-reason for contraindication. It's been our experience that when those contraindications are better defined and specified you have a stronger measure.

DR. GUNNAR: Any other comments? Dr. Yates.

DR. YATES: Given the fact that only about 30 percent of the surgeons are reporting,

we have to wonder if those that aren't reporting are doing so because they don't use beta blockers or they're reusing amiodarone or because the literature has evolved to where they don't feel the need to give it.

And it's not going to change the validity for those 30 percent that ---- in the statistics that support that 30 percent that respond, but again this would be one of those tiered answers in validity.

You wouldn't want to make this something that becomes a required PQRS or applied for hospital quality in comparing hospitals, if there's only 30 percent of surgeons replying.

Now I'm correct in that 30 percent, right? And if that's the case, then this should be one of those tiered questions. So I would answer it's valid for those people that answer.

It means something to them and they have answered correctly or incorrectly, but for those that don't report we don't know, and it needs further study. Someone should do a poll of

STS members as to how many people use beta blockers anymore.

DR. CIMA: Can I turn that around to say that due to the fact that 60 some percent don't report impacts of validity of the measure in its entirety in relationship to its outcome and driving quality.

DR. YATES: The surgeons have voted with their feet in terms of how valid they think it is. So I have -- I'm always loathed to be one of the ones that throws a measure out of the bus, but I'm trying to say that I think that's a big deal that so many people aren't replying.

It means they've decided it's not a valid thing to be worried about and reporting.

MR. LYZENGA: This can also be considered a question of use again, in the use and usability of the measure, but I mean if you're construing it as a question of validity I think we can --

DR. YATES: Yes, it probably goes to usability more than validity because the validity

is inherent to what you have statistically, but I'm going to face validity. Does it pass the sniff test? And so that's just raises an issue.

DR. GUNNAR: I would say that the validity is impacted by the data that you receive in relationship to its overall -- I don't know how valid 95 percent is in relationship to the fact that only 30 percent of the -- we only get reporting of 30 percent.

DR. SAIGAL: Are we testing the validity measure as reported and how that measures, you know, what we think that measures. And it's not reported we can only count on it that it's not used.

MR. LYZENGA: Right. I would say that's accurate, and that's a decision made by clinicians they have ---- as I understand the PQRS program, they can select from a slate of measures that they choose to report on.

DR. SAIGAL: If they chose to report it, the data we have now is what we have now. You had a question about the validity of that

data. You said that you had some concerns about that. Is that right?

And would you say that on the data that we have now is it strongly valid or weakly valid? Or would you vote for it or not?

MR. LYZENGA: Is that question for everyone?

DR. SAIGAL: For the presenter.

MR. LYZENGA: Sorry. Could you repeat the question?

DR. SAIGAL: I think it was Collette, am I right, that mentioned that? Did you say that you had a problem with the validity of the measure?

You had some data that you were concerned about? So would you vote for it or not?

MS. PITZEN: I would tend not.

DR. BERG: Let me provide just a little additional information because there was additional ---- and I believe this is part of the package as well that when the medication

administration record was looked at as well the documentation was found therein.

So the inter-rater reliability with the addition of the MAR to the data it improved from the 64 percent that you're talking about to an 87.8 percent.

MS. PITZEN: Thank you for the additional point, but of what of the whole does that represent?

Because when you're saying your inter-rater reliability agreement is 64 percent, then to me that would mean that not many people were using the medication administration record. Is that a correct assumption?

DR. BERG: With the medication administration record, originally when our reviewers ---- so when we look at inter-rater reliability we're comparing between different reviewers and we're comparing our reviewers to the other reviewers as well, the people who were reporting.

With the addition of the MAR, our

reviewers were not originally looking at that. When we made it a requirement to look at the MAR as well, then the inter-rater reliability increased.

This measure comes from registry data and et cetera, and so what we are doing when we're looking at that and assessing the reliability and validity of the measure then is we're trying to determine whether the information in the record supports the decision that was entered into the registry by the hospital or the claims data. We needed that additional piece of information to show that the measure was valid.

Obviously, we can't tell you what was in the minds of the people who were actually reporting it to the registry, but what we can report is that when our reviewers were looking at the MAR, then the IRR improved.

MS. PITZEN: Okay. This is Collette. Then I amend my previous comment because the MAR would be the source of the data.

And just kind of comment or feedback

is we do a lot of validation in Minnesota as well and we're validating against data received versus what's actually in the medical record. Anything you can do in the future to outline that would be great. Thanks.

DR. REZEK: This is Gary Rezek. If I could just say a quick word about that. I think that 64 percent speaks more to our methodology than it did the actual validity of the measure because our measures are typically outpatient measures.

This of course is ---- our sample of providers who we requested documentation from were from part B record. We requested medical documentation to support the numerator code or the code they reported, and we think what was going on is the medical records were often in the inpatient setting, and I think it was just difficulty in obtaining those records and submitting them to us.

We didn't often get the parts of the medical record that we were requesting and it

took several rounds of requests for further information before we began to really get the documentation that we needed to validate the measure.

DR. GUNNAR: Any other comments? Seeing none, let's go ahead and vote on validity.

MR. LYZENGA: All right. Fingers crossed here. Yay, all right.

We've got nine percent high, 64 percent moderate, 27 percent low and zero insufficient. The measure passes on validity, so now we'll move to feasibility.

DR. GUNNAR: Any comments, Dr. Olsen?

DR. OLSEN: The only thing I would say is that if we think that the records can be accessed readily then the feasibility should go right in line with that.

DR. GUNNAR: Gary.

DR. ROTH: The same comment other than of course what we're hearing with the disparity of the data that we're receiving. I might suggest that feasibility has been limited for

now.

DR. GUNNAR: Any additional comments?

Go ahead and vote. Yes, Dr. Yates.

DR. YATES: I was just going to make the comment that that's part and parcel part of the problem with PQRS for specialists which hasn't been a smooth transition for this process. It hasn't been real easy to involve specialists with the PQRS process. It hasn't been smooth.

DR. GUNNAR: Go ahead and vote.

MR. LYZENGA: We are voting on feasibility.

We have nine percent high, 73 percent moderate, 18 percent low and zero insufficient. So the measure passes on feasibility, and we'll move onto use and usability.

DR. GUNNAR: And usability.

DR. OLSEN: Since it's a process measure, all it is is measuring whether people were compliant with beta blockers or not. So there's no strict out outcomes associated with the process measure.

DR. GUNNAR: And I believe there was other discussion you can apply to this. Amy.

MS. MOYER: Circling back to earlier, I decided my comment belonged better here, and I think it builds on what Dr. Yates was saying.

I don't think we're necessarily seeing surgeons voting with their feet away from this measure. I think we're kind of seeing them voting with their feet away from PQRS. As I understand it, participation has been low in general in the program.

So it's not like they're saying, oh, I don't want to use this measure. It's not good.

DR. ROTH: Not only are they voting away from PQRS, but also just that the management ---- I can use the state of Michigan as an example.

We have a very extensive statewide cardiothoracic collaborative where every hospital participates in it that performs cardiac surgery, and it's all about amiodarone. They don't even discuss beta blockade, as was mentioned earlier.

I think if you looked at our state

we'd see that we probably are part of those that are not participating in this measure.

DR. YATES: And just to follow up what I said, it's not something that they can walk away from. They're going to have to participate in PQRS as time goes by, and there hasn't been a lot of different specialty-specific PCQMs or other vehicles to report PQRS through.

If you have one and it's gone through all these steps up to this point, I would beg that everyone allow it to stay in place for the cardiac surgeons to have at least another PQRS that they can use when they report that's at least applicable to what they do.

DR. GUNNAR: I think it would hard to say it's not a measure when it's still a process measure still in the guidelines. Fred.

DR. GROVER: I am conflicted here, but it is part of the STS composite score for a process of care.

When my colleagues are here tomorrow, it might be interesting to ask them what

percentage of people ---- I'm pretty sure virtually 100 percent report it because it's one of our metrics in the composite.

The question would be what is the level of compliance. It can be a tough one if you're getting patients transferred in and out of other hospitals.

DR. GUNNAR: I think that's interesting because my impression was it was topped out. I mean I was ---- a little from the STS perspective, but that was my bias. So I shouldn't bring that in.

DR. GROVER: As I said, I think the enthusiasm for it isn't as high as it was a few years ago as has been state, but I don't know that anybody's come up with the evidence to eliminate it by any means.

DR. GUNNAR: Right.

DR. GROVER: And it's in the guidelines.

DR. GUNNAR: What's interesting is that the data in the noncardiac patient

population hasn't influenced, at least in the thirty ---- in the people who report this, it hasn't influenced their actions. Right? The compliance is still going on.

DR. GROVER: And this doesn't mean that you can't still just amiodarone in the postoperative period for the purposes of atrial fib.

DR. GUNNAR: All right. Any other -- Amy, yes.

MS. MOYER: I went to the STS's site because I was curious and this is ---- you know, perioperative medicines is supposed to be reported. And there's at least enough radiation to have all of the star groups represented in the publicly reported results. There's ones and twos and threes.

DR. GUNNAR: Any other comments? Chris, you had your -- okay.

Let's go ahead and vote.

MR. LYZENGA: We have nine percent for high, 73 percent moderate, 18 percent low and

zero insufficient information. The measure passes on usability and use.

We can go ahead and discuss the question of overall suitability for endorsement or just vote on it if there are no further comments.

DR. GUNNAR: I think we can go ahead and vote.

MR. LYZENGA: 77 percent yes, 23 percent no. The measure passes. Thanks to our stewards.

DR. FLEISHER: I think you heard the concerns of the Committee with regard to what would be required the next time it comes back.

DR. BERG: I agree, and it certainly will be comments that will go to our technical expert panels the next time as well.

Just thinking about it, there may be increase in exclusion criteria utilizing other alternative -- appropriate alternative medications, et cetera.

So perhaps the use of amiodarone as an

exclusion criteria might be something that would increase the validity of this, but we'll go back to the TEP and see what they say.

MR. LYZENGA: Now we had placed an item on the agenda to discuss harmonization. I don't know if everybody has had a chance to take a look at this comparison table, but we have sort of a side-by-side look at the two measures that we've identified as being related or potentially competing, 0127 and 0236.

The 0127 is preoperative beta blockade. That's an STS measure. I understand that's in the composite, and then the measure that we just discussed.

Really we're just actually looking for some feedback from the developers on whether this has been -- the specifications here, harmonized with the STS measure. To the extent that they have not, is there a plan to do so? Or is there a justification or rationale for not doing so? Share your thoughts on the need for harmonization in general.

MS. LUCAS: Yes, we've reviewed the other measures that we submitted on the form and several of them, 0017, is beta blockade at discharge. So that has the patient coming from the inpatient setting, and the target population is the same.

However, as I said, it's a beta blocker at discharge and our measure focuses beta blocker within 24 hours prior to surgical incision.

As far as 0117, that is a registry reporting only option, and our 0236 is a claims and registry reporting option. 0119 is risk adjusted operative mortality for CABG. So the measure focus is different.

I think after reviewing even the composites we felt that the closest one to 0236 is 0127.

MR. LYZENGA: I believe so, that's the one we had identified as well, 0127.

MS. LUCAS: Right. However, that one is -- I believe that is inpatient and registry

only and ours is outpatient.

MR. LYZENGA: Just for our sake, to clarify, this is an outpatient only measure.

MS. LUCAS: Yes.

MR. LYZENGA: I think the submission form does indicate that it is applicable to the hospital, acute care facility. So maybe we'll just ask you to correct that.

DR. YATES: We are discussing harmonization. My one comment ties into what I have said earlier in that to date, at least last year, we heard STS say time and again on each of their measures that they didn't feel comfortable breaking that out into surgeon-specific data. They felt only comfortable going by program or hospital.

As such, if this is a measure that addressed the ability of the surgeon to report to PQRS, they're meeting this measure then I think there is room to keep both measures and there's a need to because there's again a dearth of surgical specialty measures within which to

report in the vehicle of PQRS. So I think it should be protected.

DR. DUTTON: I will go even a little further. One big difference between these two measures is that anesthesiologists are able to report this one because CPT codes anesthesia for cardiac surgery have been included here, whereas they're not included in the STS measure.

It makes a difference from our perspective and anesthesiologists are very often the ones giving the beta blockers or writing pre-op medication orders.

DR. GUNNAR: Any others?

DR. CIMA: If they were to have them rechange it from the acute setting and take that out, then the anesthesia providers would be excluded because they're giving them the beta blocker in the acute setting and not in the outpatient.

I'm still trying to figure out how a CABG is an outpatient. In my level of training I have yet to see an outpatient CABG done. I'm not

sure about why we are even discussing this as an outpatient.

DR. YATES: I think that means that they're outpatient up to the moment of day of admission.

DR. CIMA: That's on the anesthesiologist side.

DR. YATES: Right, but it depends on your definition as to when inpatient starts.

DR. DUTTON: Bob, this may explain the difference between 64 percent in the outpatient pharmacy record and then we add the MAR, which is where it would appear when the anesthesiologist administers it.

DR. CIMA: Because if you use the SCIP data on administration of beta blockers before CABG that was basically almost topped out. So I'm trying to figure out how those things are different.

This is not saying the patient came to the cardiac surgeon on a beta blocker. This is just saying did they get a beta blocker within 24

hours. It just depends on how you define that 24 hours. Usually it's within incision, but a lot of institutions it was the morning of. If they didn't take the beta blocker within the last 12 hours, they're getting a beta blocker in the pre-op hold.

I'm just trying to figure out why does it make a difference if this is an outpatient or whatever. Why not just use the same definition? Twenty-four hours before incision. Some hospitals don't count incision as an inpatient activity. It's only when they hit the recovery room or the ICU.

How you decide whether outpatient or inpatient it's dealer's choice, and so it's basically 24 hours before incision. That's sort of what it comes down to.

If you take it from an outpatient setting, then the anesthesia providers are going to get -- I'm not trying to see -- I'm trying to figure out what the difference are here.

MR. LYZENGA: I guess I would turn to

the developers.

DR. BERG: It's a good question.

MR. LYZENGA: It sounds like there is some consensus that we're comfortable having both measures. The 0236 is important and it provides an avenue for specialists and individual clinicians to report to PQRS outside of the inpatient setting

DR. GUNNAR: Do you want to vote on that?

MR. LYZENGA: I don't think we need to take a vote on it. This was really just for purpose of discussion, clarification, and sort of to inform our report.

That can be a recommendation of the Committee if you would like to recommend harmonization between the STS measure and the CMS measure. This was a good question about the care setting. I'm not sure how harmonization works in that instance.

Usually we would say if it's two different care settings then they're not

competing measures, but I guess there would still be an opportunity for harmonization there I would imagine.

So I guess that's something we can pose to the developers and to STS tomorrow what kind of opportunity there is for harmonization of those two measures. I don't know.

I turn to Melinda and Marcia, too, if you have any thoughts.

MS. MURPHY: I would say to Dr. Cima, can you articulate what harmonization of the two measures would be that you might recommend for the Committee to consider.

DR. CIMA: I don't know. Maybe just to have more measures to have more measures is probably not a good idea. Why not just have one measure that does the same thing and that's what these basically do. So why have two on the records, just have one?

I'm sure we will hear it from the STS tomorrow, because they always say it, 98 percent of all cardiac surgeons are members of the STS.

So the data is there. The last time we were here the reason we had to have separate ones from CMS was because we only care about 65 years old, but STS has age in it.

We can figure out ---- with an Excel sheet, I can sort from age and then I can draw a line. I'm just saying we should simplify as well as do the right thing, and they're the same measures. They're basically the exact same measures.

It just has subtle differences in wording which to lawyers makes sense, but to people on the street makes no sense.

MS. MURPHY: Are you then saying --

DR. CIMA: There should be one measure.

MS. MURPHY: Yes, one measure which is different than harmonization which would be having them look very similar. We're talking about having a single measure.

DR. FLEISHER: If I could just comment because CMS is not here and this is a CMS -- no,

CMS is not at the table.

MS. MURPHY: Right.

DR. FLEISHER: Besides STS.

MS. MURPHY: Right.

DR. FLEISHER: And what you're really asking is can CMS use the STS dataset, which I think is a question that is beyond the scope of this committee, but is something that we can communicate.

And if you can also communicate to CMS, I think that would be a ---- is that your question? Okay. No, if that's a great question, we should move on but forward that to CMS.

MS. MURPHY: And we can do that and have a conversation with the STS and then the Committee when it next convenes, a conference call, whatever, to consider. Okay.

DR. FLEISHER: Thank you. Did you have a -- you always look --

MS. MOYER: I'm thinking. I guess in additional question I might have is this is a PQRS measure, and I believe that would then allow

an individual surgeon or anesthesiologist or whoever to say, yes, this can be posted on physician compare -- when we get to that, and this could be publicly reported. How useful a general person might find that, I'm not sure.

I'm wondering what the similar requirement might be to individually report, if the surgeon wanted to, their results out of STS. If they need to get like facility signoff or signoff from their group, that potentially could be a barrier that could exist for that source that might not exist for this. I don't know that it is, but it's something I'd be interested in hearing about.

DR. FLEISHER: I think that's actually beyond the scope of what our focus is, but we can forward that on.

Fred, did you want to comment?

DR. GROVER: I was just going to say in terms of PQRS we for a number of years have had a way where the STS database can -- when our members request that, that data is transferred

without them having to do anything. Other than that, it's straight into PQRS.

We don't count that as part of our regular public reporting business because it's an individual surgeon's preference, and what we count as public reporting is what we do through Consumers Union and through our own website.

DR. FLEISHER: Thank you. We have two hours to get through five measures. We seemed a little postprandial, but hopefully we can get some energy back. Thank you very much.

We next have perioperative anti-platelet therapy, SVS. Is SVS here? Great.

I think the next four measures are process measures. Correct?

MR. LYZENGA: Yes.

DR. FLEISHER: And then we have an outcome measure. Marcia, did you have a comment? Dr. Jarrett?

DR. JARRETT: Yes, I'm on the phone.

MS. WILSON: Hi, Dr. Jarrett. This is Marcia Wilson with the National Quality Forum.

Earlier this morning, all the Committee members had a chance to do an oral disclosure of any professional activities that might be relevant to the subject matter before the Committee.

If you would be so kind as to state your name, who you're with and disclose any activities that we need to know about, please.

DR. JARRETT: Sure. I'm Dr. Mark Jarrett with the North Shore LIJ Health System, and I have no conflicts to disclose regarding any of these measures.

MS. WILSON: Thank you so much.

DR. FLEISHER: Great. And our discussants are Alan and Dr. Jarrett. Okay.

DR. JARRETT: Yes.

DR. FLEISHER: Okay. Dr. Johnson, do you want to quickly introduce yourself and tell us what you're presenting?

DR. JOHNSON: I'm Brad Johnson, a vascular surgeon from the University of South Florida, a professor of surgery. And I'm on behalf of the Society of Vascular Surgery. I'm

here and there should be two of my colleagues or at least one. Vivienne Halpern should be on the phone today with us.

DR. FLEISHER: Is she on the phone?

MS. HALPERN: Vivienne Halpern. I'm also here.

DR. CIOCCA: Rocco Ciocca, I'm also online.

DR. JOHNSON: All right. I'm here to present the NQF Measure No. 0465, Perioperative Anti-Platelet Therapy for Patients Undergoing Carotid Endarterectomy.

It was originally endorsed by NQF in July of 2008. It is a reporting of a percentage of patients undergoing a carotid endarterectomy who are taking antiplatelet agents within 48 hours prior to surgery and/or prescribed this medication hospital discharge following surgery.

This measure is reported by the Society of Vascular Surgery/Vascular Quality Initiative registries. VQI participates, receiving benchmark reports on this measure to

see how they are performing relative to their peers and to the quality goals set for the measure of 90 percent antiplatelet uses for carotid endarterectomy procedures.

While progress has been made towards this quality goal for this measure, there still is a gap based upon our VQI reporting data. Our VQI data goes through an annual validation process to assure validity and reliability of the data.

I'm happy to answer any questions that the Surgery Standing Committee has.

DR. SIPERSTEIN: In the interest of time, I won't repeat that very nice summary, and ---- you know, some of the key aspects.

Simply a process measure, looking at medication use and discharge prescribing in many ways similar in format to the measure we just discussed. One of the key issues in the original submission and then in some of the supplemental material is the impact of compliance with reducing morbidity and mortality.

There are a number of different papers that have addressed this. They come to somewhat different conclusions both on the morbidity as well as the mortality aspect of things, but in aggregate, there is a tendency to reduce complications by the use of these agents.

DR. FLEISHER: Mark, any comments?

DR. JARRETT: Yes, I do. I agree with everything that's been said. I think the evidence is not super-strong. What bothers me the most is looking at the data and looking at the history since this has been around for a good number of years. There's two groups. There's two groups: there's -- almost everybody is getting on with giving these drugs at the time of discharge, but there seems to be a group of, you know, practices or --

DR. FLEISHER: Mark, we're focusing just on the evidence. You were not here this morning.

DR. JARRETT: But I think the evidence is there. It's weak but it's there.

DR. FLEISHER: Okay. And we'll get back to your other comments. Thank you.

DR. JARRETT: Okay.

DR. FLEISHER: Any other comments?

(No response.)

DR. FLEISHER: Can we vote? We'll call it?

Yes, go ahead.

MR. LYZENGA: So we have 5 percent high, 81 percent moderate, 14 percent low.

So the measure passes evidence. And we can move to opportunity for improvement.

DR. SIPERSTEIN: So I would say yes. As stated, half the centers did not achieve the 90 percent benchmark, and 20 percent of the centers were like under 80 percent. And there was a distribution among centers. So, yes, there is an opportunity for improvement.

DR. FLEISHER: Mark, I think your comments now.

DR. JARRETT: Yes. I agree there has been opportunity for improvement but the question is why are the centers not doing it --- is that they are looking at the literature, reading it one way, while

other people are looking at it another way, looking at it, you know, that maybe the evidence is there since it's not super-strong. And I don't know, since looking at the data it looks like there has not been a budge on the pre-op usage really significantly since, you know, all the years. And I don't know if continuing to measure that and making that endorsement is going to really change behavior, which is really the goal.

DR. SIPERSTEIN: I had a question about that also, and actually looked up a more recent abstract. And that indicated that as the number of centers participating in this program increased, the newer centers tended to be lower users as they came on board.

DR. JARRETT: Right.

DR. SIPERSTEIN: And over time they built up their compliance. So if you'll simply look at the aggregate compliance over time, that's why it actually went up initially and then ticked down a little bit.

DR. JOHNSON: Yes, can I reply? I would reply that VQI has evolved then from, you know, maybe

30 to 40 centers to over 335 centers now. We've gone from -- so far this year we reported 40,000 carotid endarterectomies. As far as the benefit and the evidence to show that it does reduce stroke and death, pre-op and discharge, I've got some other papers of course, but we have shown that it does do that.

As far as usage, and the importance of this is the fact that it does reduce stroke, it does reduce death. And now that we have increased the number of centers participating in the registry I think you will see a move forward toward increase in compliance and more use of antiplatelet agents prior to carotid endarterectomy.

DR. SAIGAL: Can I make a comment? To clarify, we are just saying the vascular surgeons believe that it works.

DR. JOHNSON: Correct.

DR. SAIGAL: Is it non-vascular surgeons in the survey don't believe it works? Or what's the distinction?

DR. JOHNSON: I'm not sure. But if you go to board certified vascular surgeons who are now doing

the majority of carotid endarterectomy across the nation, yes, they believe it works. So, yes, you do have some people that aren't board certified, aren't fellowship trained that are still a little leery of using -- they are worried about bleeding complications. And that was one of the questions, comments here.

I've got a paper here from Stone in 2008, with use of antiplatelet agents across the board in vascular procedures, the bleeding complication is negligible and certainly does not warrant not using antiplatelet agents for carotid endarterectomy or bypass operations.

DR. SAIGAL: And I would add that if these guys think it works then the rest of the committee should be convinced that mechanism is measuring this to see what's going on with their performance, in my view.

DR. HALPERN: I would also add that I can -- most of the neurologists, they're not surgeons but the neurologists who do strokes are on board with peri-operative use of antiplatelets.

DR. FLEISHER: Thank you. Very helpful.

Oh, Tom?

DR. HANDY: Do non-specialist non-board certified surgeons use this registry? I would imagine that it's primarily the specialists that use this registry.

DR. JOHNSON: It's primarily vascular surgeons but there are also cardiologists. It's not an exclusive registry of vascular surgeons. So, yes, if you look at the type of centers participating in it, of the 330 or 40 we have now, one-third are academic, one-third are teaching-affiliated hospitals, and one-third are private. So it's pretty representative of a community of people performing vascular surgery.

And as this progresses along I think we will see more non-board certified vascular surgeons that are still performing carotid endarterectomies, they will be using more platelet, antiplatelet agents.

DR. FLEISHER: To follow up on that, do you know the percentage of vascular surgeons, the percentage of vascular operations that are being

performed in the United States that are part of this database? This gets to the STS comment that we usually hear.

DR. JOHNSON: I know we are increasing but I can't give you that number, no.

DR. FLEISHER: That might be useful going forward.

DR. JOHNSON: Yeah, that would be going forward. Yes, sir. I agree.

DR. FLEISHER: To find out that piece of information.

DR. JOHNSON: I agree.

DR. FLEISHER: Right. Correct; the neurosurgeons' performance.

DR. HALPERN: I think it also varies by area of the country. So in the big cities it tends to be more board certified vascular surgeons. In the rural areas there are probably -- because I saw data on this not too long ago, but there are not a huge percentage, but maybe 20 percent general surgeons who are also doing vascular procedures or other specialties.

DR. FLEISHER: Thank you.

My comment refers to the ability to drive performance since the group who would probably need the biggest incentive may be the group outside, so.

DR. HALPERN: Yes.

DR. JOHNSON: I totally agree.

DR. FLEISHER: Shall we vote? We will call it.

MR. LYZENGA: Thirty-five percent high, 61 percent moderate, 4 percent low, zero insufficient. So the measure passes on performance cap.

Move on to reliability.

DR. SIPERSTEIN: So just a couple quick comments. Obviously the inclusion/exclusion were clear cut. This is for elective carotid only, for example not combined with CABG. This obviously excludes emergency cases, those with drug intolerance. And so all of that seems perfectly reasonable.

They presented evidence on reliability testing by doing chart abstractions and corresponding -- correlating with what was in the database, and it was a very high, approximately .9 level. It's not a

very hard thing to measure actually.

DR. FLEISHER: Mark, any comments?

DR. JARRETT: No comments. Agree 100 percent.

DR. FLEISHER: Okay. Can we vote? We can see the results.

MR. LYZENGA: Eighty-three percent high, 17 percent moderate, zero low, zero insufficient. The measure passes reliability. And we will go on to validity next.

DR. SIPERSTEIN: So again I would say that there is a high degree of validity with this measure. I mean obviously we've had a little bit of discussion in terms of whether the, you know, literature has shown a huge benefit, but in terms of the actual measurement, the exclusion, there isn't, the risk adjustment's not applicable, all make sense.

DR. FLEISHER: Mark?

DR. JARRETT: I agree. Nothing to add really.

DR. FLEISHER: Any other comments?

(No response.)

DR. FLEISHER: Please vote.

Okay. If we can get the vote, please?

MR. LYZENGA: Fifty-seven percent high, 43 percent moderate, zero low, zero insufficient. The measure passes validity. And so we will move on to feasibility.

DR. FLEISHER: We will let Mark go first.

DR. JARRETT: Feasibility is, you know, I mean it's just basically the same as the other STS databases, I mean as the STS databases, the same type of thing. I don't see a problem with it. I think it's very feasible to do this. It's not a -- it's an easy measure to measure.

DR. FLEISHER: Allan?

DR. SIPERSTEIN: Yes. If you're part of the database, obviously very, very easy. If you're an institution that wants to do this independently, obviously getting the data together would be more of a challenge. But for the growing number of participants, very straightforward.

DR. FLEISHER: Larissa?

DR. TEMPLE: What's the cost to belong to

the registry?

DR. JOHNSON: Our cost for Tampa General, at my institution, is $14,760 this past year. It comes in modules. Each module is $2,100. Most institutions will have like a carotid endarterectomy module, endovascular aneurism module, so you're paying $2,100 per module. So our cost was $14,000.

Data entry again is the big problem. And that cost right now is -- originally at our institution we have medical students. Most people have a nurse entering it. And currently Tampa General has now hired, I've got two people that work in their IT who are doing my data entry for me. So data entry, of course, but all things is a cost.

DR. TEMPLE: So in looking at who is participating in the registry and who is not, it would be very helpful to see both the solo providers or the smaller hospitals and what the uptake in the registry -- because it doesn't sound like it's a measure you can really collect outside of the registry very well, right?

DR. JOHNSON: Yes, you are correct, it's

difficult outside the registry.

DR. FLEISHER: So if we can make sure, Larissa, that in the report we reflect what we would like to see back --

DR. JOHNSON: Okay.

DR. FLEISHER: -- when this measure comes back, should it pass.

DR. JOHNSON: Okay, thank you.

DR. FLEISHER: Amy?

MS. MOYER: I was just curious if it's billed on kind of a per facility basis or if it matters, you know, how many individuals are in the registry. So, for instance, if a facility purchases access to the carotid endarterectomy module can they enter all their people who perform that or is there kind of a incremental fee for that?

DR. JOHNSON: No. It's per module, so number of physicians does not increase costs for anybody at Tampa General. Cardiologists, interventional cardiologists, vascular surgeons, general surgeons, they all participate and they only pay that one fee.

DR. FLEISHER: Yes, A.J.?

DR. YATES: Given the fact that it's a yes/no dichotomy, I mean you do it or you don't do it, wouldn't it be simple to extend the measure in such a way that it can be reported in other venues other than just the registry?

DR. HALPERN: We are actually working, we were actually working on G-codes for PQRS. And I can't, I don't remember what stage we are in that, frankly. But we did have, we were working on G-codes for that.

DR. YATES: Okay.

DR. FLEISHER: Great. It sounds like that will be an important answer -- important to answer at the next phase.

Collette?

MS. PITZEN: I'm just curious how many data elements are coming through in the registry?

DR. JOHNSON: You mean data -- why don't you define for me better what do you mean by data elements?

MS. PITZEN: Well, a little bit for this

module, for cardiac endarterectomy.

DR. JOHNSON: For how many entries, how many things are we entering?

MS. PITZER: Correct.

DR. JOHNSON: Yes, for carotid endarterectomy, yes, a huge -- not huge, but it's a number of elements involved, anywhere from their history of smoking, significant, four or five pages just for carotid endarterectomy. So it's a 15- to 20-minute per patient entry.

MS. PITZER: Can I make an additional comment? Just something for the committee to consider as more registries are being used, and I know STS has a big uptake in the country, also has lots of data elements, this is a process measure that could be collected using other sources of data in perhaps a more efficient way. So I just wanted us to keep that in mind in terms of feasibility.

DR. FLEISHER: Sure.

Marcia, from an NQF perspective where would that discussion occur or where could that occur?

MS. WILSON: I'm sorry. It would come

under a recommendation or comment from the committee and be highlighted in the report as such, raised under the feasibility section.

DR. FLEISHER: And perhaps again under the gaps, it could be highlighted as a general comment --

MS. WILSON: Absolutely. Yes.

DR. FLEISHER: -- throughout the document of surgical registries and the initiative.

Barry? I'm sorry, I'll get you next.

MR. MARKMAN: You're the first registry -- what could be done for the October ICD-10 coming in versus the nine codes that we --

DR. FLEISHER: What can be done?

MR. MARKMAN: Yes. I mean have you prepared? A lot of these registries work on codes; right? So ICD-10 is coming in in October. So how, have you done anything in preparation for that?

DR. FLEISHER: Vivienne, can you answer?

DR. HALPERN: In terms of how the registry gathers this data?

MR. MARKMAN: Yes, yes.

DR. HALPERN: Yes, the company that helps

run the registry they've already started changing over to ICD-10 coding. So they will be ready.

MR. MARKMAN: Okay.

DR. HALPERN: They are well aware of that.

DR. FLEISHER: Great. Liz?

DR. EREKSON: I just wanted to play off of Collette's comment just slightly, aspirin is over-the-counter, and so this isn't something that we are going to be able to get in all things because doctors aren't prescribing it in terms of e-measures and things like that. So for this particular measure, the way it's collected right now seems very valid to me. But I would have concerns if we started morphing into an easier data set to grab things off of.

So for this particular measure I think the way it's written is the right way. And when we are thinking about future measures or making things easier, when patients have access to things that may not be in the medical chart or may not be prescribed by physicians in the EMR, we just have to consider that.

DR. FLEISHER: Thank you. We will add

that to the comments.

Cliff?

DR. KO: I just have a question of clarification. Is the measure has to be put into the VQI, or it's just the specs and just put it in however you can?

DR. JOHNSON: It has to be put in through the database through VQI.

DR. KO: And we don't know what percent of carotids are in the VQI now?

DR. FLEISHER: That is what we are asking for them to provide that data. It perhaps could be useful, given you can go into the mix as it goes through --

DR. JOHNSON: Correct.

DR. FLEISHER: -- the process to have that for the CSAC.

DR. JOHNSON: I will, yes, I will get that number.

DR. HALPERN: Yes, it should be fairly easy for us to do because we can take it -- because we did -- the VQI had 48,000 carotids captured, so we

could get -- at least from CMS, we could get an idea of how many more are being done.

DR. FLEISHER: Okay. We can vote.

MR. LYZENGA: So voting on feasibility: high, moderate, low or insufficient.

DR. FLEISHER: Do we have the numbers?

MR. LYZENGA: Thirty-three percent high, 54 percent moderate, 8 percent low and 4 percent insufficient. So the measure passes on feasibility. And we go to usability.

DR. FLEISHER: Comments on usability?

DR. SIPERSTEIN: Yes, very briefly. The document indicated that probably the primary users for institutions themselves get their own report cards and work internally on improving compliance, not publicly reported. There was a statement that it was PQRS approved but not currently used.

DR. FLEISHER: Mark?

DR. JARRETT: This is Mark. Yes, if I'm not mistaken I think it may be part of a composite of PQRS. But I may be wrong on that. But otherwise, again, there is no issues with the usability that I

see. It really should be used locally for improvement.

DR. FLEISHER: Amy?

MS. MOYER: I'm not sure if this goes in this section, but I am looking at measures that are out there for carotid stenting and seeing there's not a similar measure I'm not sure what percentage of patients choose an endarterectomy versus stenting, but it would be nice to have something similar that looks like there's similar medication requirements around that.

DR. HALPERN: In fact there's additional because most, most carotid stents need aspirin and Plavix post-procedure. So, yes, I believe that they, the VQI also collects data on carotid stenting, and I know we are actually actively working on several more measures for various vascular procedures. So I think that's one of them in the future. And I think we are also trying to work with the American Heart Association so that it can be a composite, so there are many other interventional groups. And can also be SCIVR, the Society for Interventional Radiology,

because many other groups are also doing carotid stenting.

DR. JOHNSON: But Amy, just to give you it number-wise, so far as total procedures captured so far this year through March, carotid endarterectomy 48,000, carotid stents just 7,000.

DR. FLEISHER: Thank you. I think we will also take it back as a potential area of gap to ask a question of stents versus carotid.

DR. JOHNSON: Right.

DR. FLEISHER: And whether that's harmonization, or whatever the appropriate term is, thank you for bringing that up.

Are we prepared to vote?

MR. LYZENGA: Voting on overall suitability for endorsement now. I believe --- oh, are we still on usability? I thought we'd gotten through that already.

Usability and use. I'm sorry.

We have 29 percent high, 63 percent moderate, 8 percent low, zero insufficient information. So it passes on usability and use.

And now we can go to overall suitability. Any other comments before we vote?

(No response.)

MR. LYZENGA: All right, go for it.

DR. FLEISHER: And the verdict is?

MR. LYZENGA: Ninety-six percent yes, 4 percent no. The measure passes.

DR. FLEISHER: Thank you very much.

DR. HALPERN: Thank you.

DR. FLEISHER: Dr. Johnson, thank you.

DR. JOHNSON: Thank you.

MR. LYZENGA: Oh, I'm sorry, we actually -- they've already hung up now, I think -- yes, harmonization. And I don't know that we have much to talk about here. This one we do have another STS measure, I believe, of antiplatelet medication at discharge. Although this one, it appears, applies to patients undergoing isolated CABG, whereas the STS measure is carotid endarterectomies.

DR. SIPERSTEIN: Yes, I don't think those two overlap. There actually is an outcome measure that does overlap, 1540, it's post-op stroke and death

in asymptomatic patients undergoing carotid endarterectomies that was approved in 2012. And so it's an actual outcome measure that kind of overlaps in some ways with what this was looking at.

MR. LYZENGA: We wouldn't consider that competing though or --- but yes.

DR. FLEISHER: As we said in the beginning, the question will be, as it gets more robust, whether this process measure will survive in the face of the robust outcome measure. But it sounds like we just endorsed it to go to the next level for approval.

MR. LYZENGA: And it sounds like we're comfortable with these two measures existing together.

So next measure. Thank you. Appreciate it. Now we will go to the AUGS measures. The first one is 2038. I don't know if our developers are on the phone.

DR. GUNNAR: So this is measure 2038, performing vaginal apical suspension at the time of hysterectomy to address pelvic organ prolapse. And just to remind us, when this was considered in the

past I'm reminded. It's a new measure, this application, but it had been considered and turned down in the past.

MR. LYZENGA: Yes.

DR. GUNNAR: Do we know what the recommendations were at that time that we brought it up?

MR. LYZENGA: So this was actually discussed in our previous -- by this committee the last time around. And I believe we actually may have deferred it rather than -- or maybe, no, it went down. I'm sorry. It did, it was voted down. And they have resubmitted it now after -- sorry, I'm remembering now.

This was one where we have the issue of testing where there was one site in particular out of I believe four sites that had had a weird issue with their coding. And it's sort of a systematic problem that had given -- really skewed the results. So we asked them to come back with some -- to give us some further explanation and data review testing results I believe. So, yes, that was, that was the issue in the

last round of evaluation.

DR. GUNNAR: So I'm seeing Barbara giving me a head -- So before we --

MR. LYZENGA: Maybe we could have our developers.

DR. GUNNAR: --- have the developers address. So who is on the phone?

DR. PULLIAM: Yes, I am representing the developers. My name is Samantha Pulliam and I am representing the American Urogynecology Society.

MR. LYZENGA: Thank you. And if you could give just a quick overview of the measure, an introduction?

DR. PULLIAM: Okay. So this measure, as you've mentioned, is performing vaginal apical suspension at the time of hysterectomy to address pelvic organ prolapse. A brief description of the measure just is that the percentage of patients undergoing hysterectomy for the indication of pelvic organ prolapse, in which a concomitant vaginal apical suspension is performed. We believe this is an important measure and have persisted with it after our

last experience, mostly because so many women undergo surgery, over 200,000 surgeries a year, for pelvic organ prolapse. And yet up to 34 percent of them don't undergo a current, a concurrent colpopexy or apical suspension procedure.

The statistics essentially show that the risk of re-operation within ten years is significantly elevated if one doesn't have an apical suspension at the time of the hysterectomy.

With regard to the changes that have been made in this submission from last year, I think specifically we addressed the concerns regarding the testing. Our testing had been reporting based only on billing codes. And we have reevaluated the testing based on electronic and paper chart review. So that's the main difference in terms of the submission from last year.

MR. LYZENGA: Great. Thank you.

And I think we have whoever --

DR. LEVY: So we had a fairly robust discussion about this the last time. It is a process measure but if you look at what the outcome measure

would be, it's something you would measure three years later, four years later, and it's not really feasible to look at an outcome measure that you could assess reasonably.

So from the standpoint of a process measure that's directly related to an outcome, I think they've demonstrated that it is. And given the robust discussion we had last time, I'm not sure how much more discussion you want to have. We, I believe, passed this last time, based on the evidence.

DR. FLEISHER: So the evidence last time?

MR. LYZENGA: It did pass on the evidence last time I believe.

DR. LEVY: Yes.

MR. LYZENGA: It failed on reliability testing is my understanding, so --

DR. LEVY: Correct.

MR. LYZENGA: I don't know if we want to go through the votes again or if we, do we have a motion to --

DR. TEMPLE: If I'm not mistaken, wasn't this one of the ones that was the pilot in surgery

where the evidence had been approved even before we reviewed it last time?

MR. LYZENGA: That is correct. Although I think maybe the time frame now has sort of, the window has been exceeded from the point where we would want to review evidence again.

DR. LEVY: Yes, I think the evidence is as robust as they can give us given the retrospective nature of the data that they can provide. But clearly there are, you know, of the 200,000 procedures being done annually for pelvic organ prolapse, about 78,000 of those are hysterectomies. And their evidence is pretty clear that only 35 percent of them are having apical suspensions performed at the same time.

There is reasonably good evidence, certainly at ten years, that doing an apical suspension at the time will reduce the number of re-operations for apical prolapse. And that's really what we want that married to that outcome. And I think the evidence is reasonable that that's the case.

DR. FLEISHER: Dr. Erekson.

DR. EREKSON: So I just wanted to comment

on the NQF worksheet beforehand. The NQF pulled out a Cochrane Review that showed that there is likely to maybe change some recommendations. And I think it's really important to realize that this Cochrane Review is looking at the different types of apical suspensions, not apical suspension versus vaginal hysterectomy.

And so the measure developers are not choosing that you have to do one type of apical suspension over the other, they are just saying please do an apical suspension because that's actually addressing the problem which is the prolapse that's being done. So I just wanted to make sure that the committee was aware of that distinction.

DR. GUNNAR: Is this risen to essentially the standard of care -- I'm just naive -- is this a standard of care issue and those who don't really aren't meeting the standard of care? Or is this just -- because it doesn't make any sense to my why you would --

DR. LEVY: The evidence in the ACOG committee opinion is, you know, level B to C evidence.

And in that regard I think we would like to see it rise to a standard of care. And one way to do that is to create a measure that's holding people's feet to the fire.

MR. LYZENGA: Dr. Moss?

DR. MOSS: Just a question. Not having content knowledge in this area, if the evidence is so compelling that this is advantageous, why do 34 percent of board certified surgeons not do it?

DR. LEVY: Actually it's only 35 percent are doing it. Yes, so go ahead, Liz.

DR. EREKSON: So I think the other thing to realize about this specialty is that it's been an evolving specialty over a long period of time. And the actual subspecialty of female public medicine -- and reconstructive surgery just was approved and had its first board certified surgeons in 2013. So the field has been evolving. But then you have all of this other, all of these other surgeons out there. And I'm not saying that those surgeons can't perform these procedures, but I think the evidence, it's a new field, and the evidence is pointing towards this

direction.

DR. MOSS: What does it add to the operation time-wise?

DR. LEVY: It doubles or triples the time of the operation if you do it right.

DR. KO: Wow, really?

DR. LEVY: Yes. It can. I mean it depends on, you know, if the uterus is falling out and it's small then it takes 20 minutes maybe to get the uterus out. And it might take you an hour-plus to find the uterosacrals, high up, protect the ureter, do the suspension, do the cystoscopy to make sure you haven't kinked the ureters, and then close.

DR. EREKSON: In addition, it's more technically difficult, which also speaks towards this evolving specialty and more training, because it's certainly a more difficult procedure than just doing a hysterectomy.

MR. LYZENGA: Dr. Moss.

DR. MOSS: So again just a question, is there a potential unintended consequence of driving folks who aren't, either aren't adequately trained or

experienced to be pushed into doing something they might not do well?

DR. EREKSON: Well --

DR. PULLIAM: I guess I could answer that if I should.

DR. FLEISHER: Go right ahead, absolutely.

DR. PULLIAM: I suppose in theory there is that consequence. But I think a more likely consequence would be that patients would be cared for by physicians who are proficient in these types of procedures. I think this is a change that's afoot as new trainees become gynecologists. And I think it will continue to be true that people who are able to do these procedures will become a source for referring to these procedures.

MR. LYZENGA: Larissa?

DR. TEMPLE: Just in terms of consequences, when you do the suspension can you do it vaginally or do you have to do it transabdominally? And the second, when you do the suspensions is there a higher risk of ureteric injury?

DR. PULLIAM: So it can be done vaginally,

abdominally, retroperitoneally. There are a bunch of different suspensions, some with mesh, some without. And as Liz said, we are not, this measure doesn't, doesn't capture that.

Yes, there is a higher incidence of ureteric kinking or injury. Cystoscopy is basically required when you are doing this procedure to ensure that the ureters are intact.

DR. GUNNAR: I guess the next question specifically is in the literature that's provided to support this, when you perform this procedure do you tack on complications? Even though you are protecting a long-term outcome benefit, do you add complications?

DR. LEVY: So short term when performed by those who are doing this properly, which includes cystoscopy, the answer is no, because you identify the kinked ureter, you change your sutures and you re-do it. So it's an interoperative recognition of kinking of the ureter when you do a high suspension.

DR. EREKSON: So the cystoscopy at the time of apical suspension is actually an approved measure that we approved last standing committee to

make sure that we don't have unrecognized ureteral injury. When you look at the data of what is the actual morbidity of prolapse procedures, and especially perioperative morbidity and 30-day outcomes, I have published on some of this data and it's incredibly low.

Older woman, woman over the age of 80, undergo these procedures and do very, very well. And so it's in the wrong hands, yes, you could have more ureteral injury but we've already protected against that because we have a measure that says you should do a cystoscopy. And a lot of older women undergo these procedures and do very well with them.

DR. GUNNAR: I guess not to beat this up, but then shouldn't those two be combined? We've had this discussion earlier. That is, if you do the suspension you should, you must do the cystoscopy, or are they separate or are they connected? I know just from the NQF point of view I would bet these are tied.

DR. EREKSON: I think they are two process measures. And I would love to hear from the measure developers. But I kind of see this as when you have

process measures then you can have composite process measures too. And they are measuring two separate things. And, yes, we want our surgeons doing both. But --- and I think it can move towards that.

DR. PULLIAM: Right. So as a measure offer I think we had sort of two concerns: one is that neither of these have been a measure yet, so I think to combine them we would need to really address a couple of issues. One is that the exclusion criteria for the measure we are talking about now are different than those for the cystoscopy measure.

And then the second thing just is that the goals of the two are separate. This is as closely tied to an outcome measure as we have. And the other is primarily a safety procedure. So whether both measures that involve an apical suspension, I think the intent is different than the applicability specifically of a cystoscopy is more broad.

DR. KO: I just have a quick question. I didn't realize the board came in 2013. And so I assumed that the people who are boarded in this are able to do the primary procedures, the suspension and

then do the cystoscopy. Is cystoscopy part of the training of everyone in gynecology and so that's not a --

DR. LEVY: Yes, it is.

DR. KO: -- resource issue?

DR. LEVY: Yes, it is. It's within our CREOG guidelines. And we have modules and measurement on that.

DR. CIMA: So just --- I am not sure that this fits right in with this segment or it's just this is evidence, but the problem is does the evidence apply to all people doing this procedure? I mean, I was also looking at this and the next one about where you are going to get this is from a registry that's as yet to be defined, and is it only a urogynecologist that's going to be applying that or will all these gynecologists have to participate in this registry in order to get this data processed? And so, you know --

DR. LEVY: So this is easy data to get from claims or from records. But there are different codes for the suspension and for the hysterectomy itself. So you don't have to participate in the

registry in order to be able to report this. And it would apply to all gynecologists who perform hysterectomies for prolapse.

MR. LYZENGA: Collette?

MS. PITZEN: Yes, just a comment, I'm not a content expert, but I want to talk a little bit about process measures and the expectation that if you have a process that you are actually requiring everyone who is in the denominator to have that process performed, I'm wondering if the evidence is strong enough in that area, and I'm going to use to strengthen the body and the quantity and the overall? What I am hearing is the evidence is grade B and C.

So just kind of throwing that out there.

MR. LYZENGA: Lee?

DR. FLEISHER: So one of my questions -- I apologize but it was the time to get my match list -- one of the things I'm wondering about is does everyone perform the suspensions because is this a full employment act for the urogynecologists?

DR. LEVY: So it should not be. I mean these are, these are bread and butter procedures when

they're performed for prolapse. There are certainly different degrees of prolapse. Mild to moderate prolapse is something easily managed by a general gynecologic surgeon. If honestly it's down to her knees and she's got huge prolapse, those cases are generally being sent to the subspecialists, because they need much more robust and multi-compartment management.

DR. FLEISHER: So not that you could speak on behalf of the College, but would general gynecologists support a measure developed by the urogynecologists?

DR. LEVY: Yes, I think we would.

DR. SAIGAL: I'd like to make a comment too about I think Collette's point's great about the evidence level. I think the thing we're to think about is whether every process measure like this is going to have a randomized trial to support it? Probably the answer is no.

So I think if a society really does feel we're making surgery better, I think for cystoscopy it makes perfect sense to me that you want to make sure

you're not hurting the patient --- it's a quick thing to do. And even if it wasn't compensated people would do it because they don't want to have -- they're not as bundled perhaps. So it might be, you know, you just don't want the woman to have a blocked ureter. So I think it makes good sense and we should probably support it even without randomized data.

DR. LEVY: I would just say that the outcomes we are looking at are so far down the road that to generate controlled trials and level one evidence is going to be nigh on to impossible.

MR. LYZENGA: Barry?

MR. MARKMAN: Yes, that was my question: how long has this procedure been performed?

And the second part of that is, is it -- I mean, you know, there's a systemic review but is there any level one or level two studies within that systemic, you know, review to support this?

DR. EREKSON: So these operations have been around for quite some time. Sacral colpopexy has been around since 1962, the uterosacral ligament suspension, in some form or another, from before

that. So these are operations that have been around quite a long time.

I think one of the things that you have to remember is that uterine prolapse does not happen because the uterus is diseased, uterine prolapse happens because it's a hernia. And unless you address that hernia in your repair your procedure is going to fail.

And so if you look at the large, multi-centered research networks for pelvic floor disorders they are not going to look at a randomized controlled trial of vaginal hysterectomies for prolapse and not do anything for apical support. What they are looking at is randomized controlled trials between the different surgical procedures to try to figure out which is the right procedure to do. And so that's where the data of the field is at.

And in terms of does this have to be a urogynecologist versus a gynecologist, if you read the residency requirements for graduation as well as board certification for general gynecology, these are procedures that board certified general gynecologists

can be trained in and perform. But what we want, I think the measure is looking at please do something. Please do right by these women and please do something.

DR. GUNNAR: So we are going to go back and vote on the evidence; is that right, Andrew?

MR. LYZENGA: Yes.

DR. GUNNAR: Okay. So can we ---I don't know what we're saying here. We're saying --

MR. LYZENGA: No, no, we need our process. There we go.

DR. GUNNAR: Are you ready? Go ahead.

MR. LYZENGA: I think we can tally them up. Twenty-one percent high, 71 percent moderate, 8 percent low, zero insufficient. So that measure passes importance to measure -- I'm sorry, evidence.

Now we will go to opportunity for improvement, performance gap.

DR. LEVY: So again we've already addressed that, but 35 percent of the time it's being done, which means 65 percent of the time it's not being done. And I think that's evidence in and of

itself that there is quite a large gap.

MR. LYZENGA: Any additional comments or can we vote? Let's vote. You ready, Alexandra? Go ahead and vote.

Seventy-nine percent high, 21 percent moderate. The measure passes performance gap.

So we can go to reliability. And again this I think is where we had the issue last time, so --

DR. LEVY: I don't know if the developer wants to --

DR. PULLIAM: Sure.

MR. LYZENGA: Yes, maybe let's hear from the developer and hear what they've done to address the committee's concerns from the last time.

DR. PULLIAM: All right. So in the last round we used billing codes both for the identification of the hysterectomy procedure and the identification of the apical suspension procedure. And I think our difficulty there was that we found an institution that didn't do what other institutions do in terms of those codes.

So instead of that we reevaluated based on chart review. So the reliability evaluation or calculations in this submission are based on the identification of a hysterectomy based on ICD-9, ICD-10 or CPT codes for hysterectomy supported by diagnosis of prolapse, and then chart review to confirm the presence or absence of an apical suspension procedure.

So that's been the new way to identify this which basically eliminates the problem that we had with the prior submission.

DR. LEVY: And so with that chart review, it certainly appears to meet reliability criteria.

DR. GUNNAR: Any other discussion? Collette?

MS. PITZEN: Just a comment. For a relatively high-volume procedure this is still a small number of cases that are being used to talk about the performance of the measure and the reliability.

DR. CIMA: Yes. And to go to the point earlier, Barbara says, you know, these are all procedural codes that are there that we can use so now

it becomes a feasibility issue if you actually had to. So how reliable are we going to be going forward if in order to do this, half of the time, or half of the procedure needs to be chart extracted?

DR. LEVY: Well that, it's actually a bigger discussion than that. So maybe why don't we talk about that during feasibility.

DR. GUNNAR: So with regard to this, the reliability evidence as they present or the data as you're collecting it -- is it reliable? That's what we are voting on right now. So is it measuring what it says it measures?

DR. CIMA: Well, it does when you go back and look at the charts. So that means --

DR. GUNNAR: And let's -- and that -- yes, correct. Any other discussion?

(No response.)

DR. GUNNAR: All right, let's vote.

MR. LYZENGA: Calculating. We have 32 percent high, 52 moderate, 16 percent low. So the measure passes on reliability. We can go to validity.

DR. GUNNAR: So any further discussion on

validity?

DR. LEVY: I just had a question for the measure developers about the risk adjustment here and why there is a risk adjustment for degree of prolapse? It would seem to add a lot of complexity that seemed unnecessary because any hysterectomy for prolapse, it seems like, ought to have a suspension. And I didn't understand the reason for a risk adjustment.

DR. PULLIAM: Right. So I think you're correct, largely. I think the concern we had was just that there might be some confounding between the surgeon volume and the degree of prolapse as in were all of the high-volume surgeons doing the more advanced prolapse cases and so they ended up having more apical suspensions. And so I think what we wanted to prove with our evaluation there was just that there was a gap, regardless of whether you were looking at high-volume surgeons or low-volume surgeons, and regardless of whether you were looking at degree of prolapse.

I think our intent wasn't so much to provide any more complexity to the evaluation except

to bolster our evidence that the gap was there and that this was not just an anomaly of surgeon volume.

DR. LEVY: But so when you are specifying the measure are you risk adjusting for Baden Walker? Because that's going to require chart review.

DR. PULLIAM: Right.

DR. LEVY: You're not going to get that from --

DR. PULLIAM: No, we are not. We are not.

DR. SAIGAL: Can I ask a question? So what you are saying is that you are using administrative codes for this. And the last time the problem was is that one of these facilities used those codes differently?

DR. PULLIAM: That's correct.

DR. SAIGAL: So how does this address that? How do these data address that?

DR. PULLIAM: Right. So the administrative codes that were used differently were the apical suspension codes. So instead of using those codes to calculate the numerator, we instead used the chart review to calculate the numerator.

DR. SAIGAL: Okay, I get that. But so if people will be reporting on the measure using codes you'll still have a problem capturing what you want to capture; right?

DR. PULLIAM: So the codes will be used to identify the hysterectomies for apical suspensions. And of 638 patients who underwent chart review, 600 of them were accurate with hysterectomy and codes from all institutions.

DR. SAIGAL: So that one institution just codes differently and it will not be --

DR. PULLIAM: That one institution --

DR. SAIGAL: -- so it wouldn't be valid in that hospital still? We still think that's true?

DR. PULLIAM: No, no. The one institution codes apical suspensions differently. The hysterectomy codes and the billing codes for hysterectomy were accurate.

DR. SAIGAL: I get that. But if the measure is looking at apical suspension and that hospital does it differently, then this measure won't function at that hospital and presumably other

hospitals that use the same billing pattern?

DR. PULLIAM: No. The hospital would use the identification of hysterectomy with a diagnosis of prolapse for the identification of cases. So the denominator -- the numerator would be calculated from that cohort based on chart review. So that the billing codes for apical suspension which were erroneous at that institution would not come into play.

DR. SAIGAL: Okay. So the measure is specified by chart review now.

DR. PULLIAM: That's correct.

DR. SAIGAL: Okay. Thank you.

DR. EREKSON: So just when I was going through the validity testing, I read through it, I think that there was a little bit of confusion that has slightly been cleared up with the NQF worksheets. Can you clarify specifically, this measure is not risk adjusted by either prolapse grade or by surgeon volume; correct?

DR. PULLIAM: That's correct.

DR. EREKSON: And why you looked at

prolapse grade and surgeon volume was to add to the face validity of this measure that women with larger prolapses and surgeons doing more cases are actually performing apical suspension more often; correct?

DR. PULLIAM: That's correct.

DR. EREKSON: Okay.

DR. GUNNAR: Yes, Fred, did you have a --

DR. GROVER: It is good to see you bring this back and that you have done all this work to try to bring a answer to our questions from a year ago.

DR. PULLIAM: Thank you.

DR. GUNNAR: All right. I think -- okay, Collette?

MS. PITZEN: Okay. Maybe I am stating the obvious but a measure that is using procedure codes to identify the numerator and denominator was not successful, and frequently identifying the numerator I have really strong concerns about the threat to the measure.

DR. CIMA: When you either have four hospitals and one of them, so that's a quarter, and you look at 200,000 hysterectomies being performed,

90,000 being performed for prolapse, or greater numbers than that, this measure would then force a lot of people, a quarter of those people to have to set up a system to do chart abstraction for this procedure. So that, in other words, this is one of the issues.

DR. GUNNAR: So I think that goes to feasibility.

DR. CIMA: That goes to feasibility. But, you know, if you pass this then you pass and we have to go through everything. But taking into the picture the whole big thing, it's valid to do it but is it really valid that we are going to be able to do it? Because a quarter of them are not going to have accurate data until they do the chart review.

DR. PULLIAM: That's -- I'm sorry to interrupt -- but that's correct. I think we are specifying this as a chart review data, but everyone will be required to do the chart review data. This is not, this is not something that allows both to be done. We've only submitted it for a chart review.

So if you were talking about -- I mean I think there are many things that will require a chart

review but the full evaluation will require a chart review of everyone.

DR. EREKSON: And, Collette, just to -- and maybe this is the measure developer, the way I read the measure, the denominator is billing codes and the numerator is chart review; correct?

DR. PULLIAM: That's correct.

DR. EREKSON: Okay.

DR. PULLIAM: That's correct.

MR. LYZENGA: So we may need a clarification in the submission form for that as well. I think it appears right now that it -- to be specified for use with administrative data only. I don't think it lists chart review. So we'll just coordinate with you on that.

DR. PULLIAM: Okay.

DR. KO: Just a quick question of clarification. The one out of the four hospitals that coded it differently, was it different and wrong or just different and that's just the way they code it? And will it be kind of like the pulmonary measure where they'll learn how to code it?

DR. PULLIAM: Right. So we did some work on that and what essentially we learned was that the place from which we obtained data was not within the system, the place where the ultimate billing codes are dispersed. And so there was some error locally. It sort of defies my understanding. But in the end apparently if we had obtained our information from a different department we would have had the correct codes.

So they are essentially insisting that the codes that they eventually bill for are correct. And we didn't have that information.

DR. KO: Because making it, if it's something that's correctable, making it a chart abstraction measure increases the burden multi-fold.

DR. PULLIAM: Absolutely.

DR. GUNNAR: All right, so any other discussion on validity?

(No response.)

DR. GUNNAR: All right, let's vote.

So we have 8 percent high, 67 percent moderate and 25 percent low. It passes.

Keep going. Feasibility. Dr. Levy.

DR. LEVY: I think we all have a problem with feasibility around the table. First of all, chart abstraction is a problem.

The second issue with this measure is that if we were to look at a way to do this with administrative data, the correct coding initiative has, in their wisdom, decided that all apical -- all suspension procedures will be bundled into hysterectomy. We've been able to actually bypass that but it's going to require a modifier. And so we don't have any testing to say whether using administrative claims we would be able to capture things appropriately.

Given that they have specified this as a chart review, that means that particular problem is gone for the moment but certainly the feasibility of collecting this using chart review is going to be terrifically burdensome, as Cliff said.

DR. GUNNAR: Amy?

MS. MOYER: I am curious what the difference in burden is -- and maybe it's that the

data are conveniently captured somewhere -- between taking something from a chart and putting it in the registry versus taking something from a chart for a measure. Are they similar? Are they different? I mean, should I think about them similarly?

DR. LEVY: I think that people who participate in the registry commit themselves to a great deal of abstraction. This is much less burden than participating in a registry. You know, my hope was that these could be specified using claims, and that would have been much better for all. You know, I don't have a lot of confidence that there will be a lot of people participating in the registry. And that leaves people with the only opportunity to do things to do chart review.

You know, could you do ongoing chart review when you come back from the OR and keep an ongoing record? You could do that. And there are ways to do it that are not horribly burdensome, that don't require you to pull every chart you did last year, but just do it on an ongoing basis.

But there is a big difference.

DR. GUNNAR: Yes, I mean the surgical, you know, improvement projects get measures where those are pulled by chart review. I mean this isn't, this isn't a unique point of view from NQF's perspective.

DR. LEVY: I think the only difference there is that the SCIP measures might be facility measures or individual practitioner measures. These are individual practitioner or group practice measures, they're not hospital measures. So we're not going to have hospital abstractors spending time in doing this work, it will be done by the individual provider.

DR. CIMA: And the SCIP chart polls were almost universally electronic, for almost all institutions, after a fashion.

DR. GROVER: Can I ask, how many data elements are there -- will be required to be pulled by chart review?

DR. MORGAN: My name is Dan Morgan, and I'm one of the developers. Dr. Pulliam had to step off the call.

The number of elements is actually

fairly small. It's really a matter of looking at the operative note and looking for the description that the colpopexy was done.

DR. GROVER: So it's at one spot in the chart, and it's -- is it five data elements, ten, or three, or -- ?

DR. MORGAN: I would -- I would imagine that we -- when we did the chart abstraction for this, we did a more extensive look. I really think that it could be down to one or two -- excuse me, one or two elements identifying which colpopexy was done, that it satisfies it, and that -- that would be it.

DR. GROVER: Okay, thank you.

DR. LEVY: Fred, I think we'd need to know that the hysterectomy was done for prolapse, so we would need to know the primary ICD-10 code for -- for the reason for the hysterectomy, and then just the other data element would be was a suspension performed?

DR. GROVER: But that -- that would be -- what you -- the coding could be identified

administratively, so you've honed it down to just looking then at those patients, and you're picking up two or three data points --

DR. LEVY: Yes.

DR. GROVER: -- in the op note?

DR. LEVY: Yes.

DR. GROVER: I mean, that sounds pretty doable to me, but I'm not -- I'm not paying for it.

(Laughter.)

DR. GROVER: But on the other hand, to put this in perspective, we got a lot of pushback on -- I mean, the best data is clinical data, and how much of the health care expenditure in this country is spent toward quality and safety?

I don't know the exact answer to that. Is it two percent at that, or one percent? It's very little, and I think we tend to overemphasize the cost of collecting good data sometimes at the expense of our quality of our care.

DR. GUNNAR: Dr. Yates.

DR. YATES: Yeah, just a question I

got from our GYN colleagues that are here and for the developers.

I am not familiar with that part of the CPT book, but, you know, they defined some of our CPTs by the diagnosis going into it depending on what type of hip fracture it is, whether or not you get paid one way or the other. They don't distinguish prolapsed versus non-prolapsed for a hysterectomy in the CPT.

DR. LEVY: That's correct.

DR. YATES: And they have no intention of expanding the CPT so that it would --

DR. LEVY: No.

DR. YATES: -- include, with or without apical --

DR. LEVY: Well --

DR. YATES: -- with --

DR. LEVY: -- so I'll put on my ACOG hat for a minute and tell you that because of the CCI edits and our inability then to collect data appropriately, we probably will be putting in code change proposals to put in bundled codes

that have vaginal hysterectomy with the different suspensions and the different repair codes.

You know, we're looking at at least two years to two and a half or three years down the road before that can get through the system, but clearly, in order to be able to use administrative data, we've got to be able to capture it in a way that's not going to require three modifiers.

DR. GUNNER: Any further -- oh, Dr. Moss.

DR. MOSS: Just a quick question for the NQF staff: is there experience with other measures that requires abstraction of a small number of elements from an op note, and what has been the feasibility experience with those?

MR. LYZENGA: I don't know that we've collected any data or anything like that, or really gotten much feedback. I mean, in general, measures that require chart review are sort of considered to be lower feasibility than those that use administrative data, but you know,

that's kind of up to your judgment as experts here.

DR. EREKSON: So our Surgical Standing Committee, last time that we were in session, approved 2052 and 2063. Both were measures that were identified by denominators, and then the cystoscopy, which is that one data element of the procedure, was identified by chart review.

So our committee last time did approve those.

DR. DUTTON: I am sensitive to burden of reporting, but one of the things I think we've learned over 20 years of doing this is if we have a good measure and put it on the table, people will figure out how to measure it and how to get the data.

DR. GUNNAR: All right. Let's vote on feasibility.

MR. LYZENGA: So voting on whether this measure has high, moderate, low, or insufficient feasibility.

(Pause.)

MR. LYZENGA: So we have zero for high, 67 percent moderate, 33 percent low, zero insufficient. It passes on feasibility.

And we move on to usability and use.

Any comments from the Committee on usability or use of the measure?

(No response.)

I believe we have, just looking at the form here, it hasn't been used yet. Is that correct, I'd ask the developer?

DR. LEVY: Yes, it hasn't been developed -- it hasn't been used.

DR. MORGAN: Yes, it has not been used yet.

MS. REEDE: So pelvic floor registry is a goal, not --

DR. LEVY: Correct. It -- it just came to fruition this year, and just enrolling practices right now.

DR. GUNNAR: Amy?

MS. MOYER: Listening -- I mean, in listening to you all talk about this, this

certainly sounds like something that I'd want to know if I were having this kind of surgery, and as a purchaser, although we don't purchase a lot of hysterectomies, they are very expensive when they happen, and avoiding that re-operation and complications down the line would -- would be of interest to us, and getting it done right the first time, so -- .

MR. LYZENGA: All right. Shall we vote?

Go ahead and cast your vote for usability and use. Your options are high, moderate, low, or insufficient information.

(Pause.)

MR. LYZENGA: We have 13 percent high, 71 percent moderate, 17 percent low, zero insufficient. It passes usability and use.

So now we will move to overall suitability for endorsement.

Any additional comments?

(No response.)

MR. LYZENGA: Seeing none, let's go

ahead and vote on overall suitability for endorsement.

Your options are yes or no.

(Pause.)

Oh, we lost it again.

All right. I think we're going to have to do another re-vote -- redo on this one. Let us know when you're ready.

(Pause.)

All right, here we go. Let's re-vote on overall suitability for endorsement.

(Pause.)

Okay, we have 92 percent yes, eight percent no. The measure passes.

I believe the next measure is an AUGS measure as well. This is 2677. If the developer wants to give a just very brief introduction of this one as well?

DR. GUNNAR: So before we go there --

MR. LYZENGA: Sorry, one moment.

DR. GUNNAR: -- I think that question that still remains is now that we have two

measures that really are connected, how should they -- are they tied together? Should they be tied together?

They didn't propose them together, but really, they are intricately entwined.

MR. LYZENGA: And so this again is the measure that we discussed a little bit earlier that was passed in the previous cycle, cystoscopy. So I guess the question is whether we think that should be combined or in any other way -- harmonized in any other way, as I think we had a bit of discussion about this earlier, and there were some -- the developer noted that there were some differences in the exclusions for the two measures, and maybe they could give us -- just reiterate what the differences are between these two measures, or where there may be opportunities for harmonization, or where harmonization is not feasible.

DR. GUNNAR: Does the -- yeah, does the developer fully understand what we're asking?

DR. MORGAN: I think that ostensibly,

down the line, we could look at that. I think for right now, these both being new measures that have not yet been collected, that we would like to gain some experience with these to start, and that they do entail populations that are similar, so we'd like to keep that open for the future, but for right now, get some experience --

DR. SAIGAL: I think that the -- the idea that these are two different -- they have two different outcomes they affect, and the quality improvement initiative may be directed in two different directions: we'd indicate that for the start at least, let them start out with these measures, and then see what needs to be done to improve the compliance, and then, if the processes are in place, then merge them.

DR. LEVY: Yeah. The other point is that the denominator for these two measures is quite different. So I was just looking at the definition again for what we passed last time, and it's "performing cystoscopy at the time of pelvic organ prolapse to determine lower urinary

tract injury," so we may or may not be capturing the same -- I -- we just need some experience with this.

DR. GUNNAR: So I think for the record, a future evaluation should address whether or not they are connected, should be harmonized or combined.

MR. LYZENGA: And I think the recommendation of this Committee is that they should go forward as individual measures at this time and that we need to explore, as we gain more experience with the measure, whether combination with another measure or harmonization would be appropriate. Is that -- is that fair?

DR. GUNNAR: Does anybody disagree with that position?

(No response.)

DR. GUNNAR: Seeing none, we will carry on with the next measure.

DR. PRESTON: This is Mark Preston here. Can you hear me?

MR. LYZENGA: Yes.

DR. GUNNAR: Yes.

DR. PRESTON: Okay. My name is Mark Preston, I'm a urogynecologist and member of AUGS, and on behalf of AUGS, I want to thank you for giving us the opportunity to present our measure 2677, entitled Pre-Operative Evaluation for Stress: Urinary Incontinence Prior to Hysterectomy for Pelvic Organ Prolapse.

This is a process measure, as mentioned before, which is briefly described as the percentage of women undergoing hysterectomy for pelvic organ prolapse who have pre-operative evaluation for stress urinary incontinence.

And what is our rationale for this measure and why do we feel it's important? You know, as was mentioned with the last measure, pelvic organ prolapse is a very common problem affecting women, and it is the primary indication for approximately 200,000 surgeries annually in the U.S.

Many women in the above group undergoing hysterectomy for pelvic organ prolapse

also have stress urinary incontinence. However, the presence of the stress incontinence is not always volunteered by the patient, and what is more, in cases of severe prolapse, underlying stress incontinence may be masked by kinking of the urethra due to the prolapse, which is often referred to as occult stress urinary incontinence.

When stress incontinence is not treated at the time of prolapse repair, the patient will often suffer from stress urinary incontinence following the prolapse repair, necessitating either an additional surgery with its associated risks and costs, or the patient having to live with her incontinence.

Implementation of this measure will improve quality by increasing the pre-operative diagnosis of stress incontinence in women undergoing hysterectomy for pelvic organ prolapse, allowing for both better pre-operative patient counseling as well as greater likelihood of appropriate treatment of the stress

incontinence at the time of surgery, thus decreasing the incidence of post-operative stress urinary incontinence.

So thanks again for considering the measure.

DR. FLEISHER: Liz, comment or evidence?

DR. EREKSON: So when you're looking at the evidence the authors present, I think the most compelling is the up-to-date flow diagram, where you look at -- it's just assess for stress urinary incontinence and have a conversation about these symptoms before surgery, and not all patients are going to choose to actually undergo a concomitant anti-incontinence procedure, but this will significantly -- doing this assessment, having this conversation with the patients, can reduce your re-operation rates by quite a substantial amount.

DR. FLEISHER: And --

PARTICIPANT: The numbers are small, but I have no reason to doubt them in the way

they're presented. It's just hard to divorce myself from how this is going to be operationalized, but I will try. So the evidence is successful.

DR. FLEISHER: So I have a comment. I was a little confused because it actually talked about preferences, and it actually talked about evidence saying you should go forward, but it also talked about that the evidence of a negative test is not necessarily good enough, if I remember correctly, to say you shouldn't go forward with doing something together.

So one of my questions is how tightly linked is a pre-operative evaluation with the correct decision in surgery?

DR. EREKSON: So I think the best data, when you look at this evidence, is some randomized control trials, and they can give you a number needed to treat, so if you have a negative cost stress test prior to prolapse surgery and you don't undergo a sling, and the measure developers can correct me, there -- one

out of five women will need a sling in the next year.

And so there's pretty substantial numbers needed to treat, but it looks to me like the measure, and what this measure is trying to get at, is do something, do an assessment, have a conversation with the patient. I don't -- maybe I am mistaking?

Yeah, yeah. But in the patients who have this assessment and then undergo concomitant sling procedure, they're much less likely to need surgery then in the future.

DR. FLEISHER: But my question for the developer or for you Liz is but if it's negative, there are still a number of women who undergo this procedure, so the question is, is this the right process measure? Is the evidence linked closely enough? Or is this really should be a PROM, and it should be was there a good discussion, as opposed to was there a pre-operative test?

DR. PRESTON: Well, so I mean you do

raise a good point. You know, the predictive value of doing this is not 100 percent, and it's not -- the negative predictive value is definitely not anywhere near 100 percent, and if you look at the flow chart that Liz was referring to, you know, in women who have no -- in women with a negative occult stress test, or even in women who have -- in women who have a negative occult stress test, 26 percent still end up having some urinary incontinence afterwards.

But if they have a positive occult stress test, there's an over 50 percent chance, you know, about a 50 percent chance. So there's a -- you know, there's -- even though the predictive value is not perfect and there's a substantial difference in doing that evaluation and letting it guide not only the -- well, mostly the conversation because obviously you have to have the conversation, and the patient gets to participate, in discussion, the decision-making process, but doing the evaluation allows you to have a much more informed discussion with the

patient as to what the likelihood is of having this problem following surgery than if you don't have, you know, if you don't -- if you don't do the evaluation.

DR. FLEISHER: So --

DR. PRESTON: If you have a -- but, you know, the positive stress test picks up about two-thirds of those people, so -- that are going to have post-operative stress incontinence.

DR. FLEISHER: Thank you. I would argue that a negative predictive value as low as you talk about, that my question is do you have sufficient evidence that doing this test leads to a discussion which leads to better outcomes, as opposed to all your -- you have done is your process is doing a test, that's all your process is? So -- .

DR. CIMA: That's all -- the real -- the issue is the measure is basically saying did you do a test? And not did you act appropriately upon that test, did you have the discussion and make a decision?

And that was the -- I mean, the data basically on about 800 patients, two different trials looked at it, and basically, they said yes, you don't -- if you do the test, you're going to identify patients, but the real question -- and you know, so we said earlier that the reason to do the suspension is to avoid surgeries down the road, but this test, this is different. This is saying did you have the discussion, and that a certain number are going to need to have it anyway, and how are you going to define who that group is and who is accountable for it? You really just are saying who is -- who has the test?

So the evidence suggests that asking the question does have an -- provides information, but it doesn't impact an outcome.

DR. FLEISHER: So recognize that's the question, does the relationship with the health -- no, this is not the right --

DR. PRESTON: And, you know, it really depends on if you were to -- you can never answer

that question the way you're phrasing it because what you're saying then is that you would 100 percent of the time -- I mean, if you 100 percent of the time said if you have a positive stress test you should do the incontinence procedure, okay, but that takes then -- then you're eliminating any patient interaction and patient participation in that discussion.

So, you know, if you were -- if we were to phrase it and say if you have a positive cough stress test, you should do a sling, and you're going to -- you know, if you look at it that way, yes, then you -- then we reduce -- you change the outcome significantly. If you have a positive cough stress test and you do the surgery, you have a 16 percent chance at having incontinence afterwards, versus not doing the surgery, where you have a 50 percent chance of having incontinence after the surgery. So there is a big difference in the outcomes.

But -- what -- the step that is missing there is the conversation, but you can --

no test would pass that criterion because you're -- you can't eliminate that step, right?

I mean, it's just like saying if you're in the 90s percent blockage of your coronary artery --

DR. FLEISHER: No, I understand. I think --

DR. PRESTON: -- you understand what I am saying, right?

DR. FLEISHER: Yeah, no, I understand what you're saying. Do I have other comments from any Committee members? Because I understand what your argument is.

DR. SIPERSTEIN: Yeah, I guess, you know, the evaluation -- I mean, requiring the evaluation, I think it's -- it moves towards impacting the outcome. It's just like we just approved a measure saying it's okay to write a prescription for antiplatelets at discharge. We're not proving that they filled it or they took their pills, but we feel that that action is moving towards a positive outcome. So it's a

surrogate of moving in the right direction.

DR. FLEISHER: And people get to vote on that.

DR. SIPERSTEIN: We're not -- we're not demanding perfection.

DR. FLEISHER: Right, people get to vote.

Fred, did you have a comment? No? Okay.

A.J.?

DR. YATES: I hear where you're coming from. The issue is what is the healthcare state that is affected by doing the screening? And in this particular case, you would have to say that the healthcare state is the -- this is going to sound ephemeral, but it's the ability to do an effective shared decision-making with the patient in terms of how to go ahead, and that would be -- without the stress test, you're not giving the patient a fair, full assessment that impacts their shared decision-making as to whether they want to agree to go ahead to have a sling versus

not.

It's not -- it's -- it may not be necessary for them to have that sling, but it does enter into the shared decision-making that the patient deserves to have with the surgeon.

I would argue that this is -- this is moving ahead to usability and everything, but in terms of this evidence, this comes down to an office visit discussion that I don't know why this is a hospital measure as opposed to a PQRS-type measure that would be aimed at the surgeon's processes as opposed to surgical outcome.

DR. FLEISHER: Liz?

DR. EREKSON: I also think if you're really going to the nuts-and-bolts, that I am going to check a box and say I did a cough stress test versus I am going to check a box and say I talked to the patient about her options, I think I'd want that cough stress test there. You know, I would want that testing there to back up the fact that I had that conversation, yes.

DR. FLEISHER: So recognize that CSAC,

we would like to not have the physician do the checkmark, it would actually be the patient who does the checkmark that they have an adequate -- now, that's not there yet, but that's really a PROM as opposed to a physician instructional member.

Other comments? And I don't want to -- Collette.

MS. PITZEN: I am going to try my -- maybe I won't make sense.

But again, to me, this feels like a standard of care, an assessment that should be happening that perhaps maybe is not happening, versus a process measure that is going to be used for whatever use, quality improvement or accountability.

Again, you want it to have that direct relationship with the outcome, and again, I think that brings it into question a little bit.

DR. FLEISHER: Other comments before we vote solely on evidence that this process is linked to -- that performing this process is

linked -- and actually, I would like NQF staff to say it correctly because I don't want to bias it.

MS. JOHNSON: I am going to hand it over to Andrew just because my mind was still a little bit in the next frame, sorry about that.

MR. LYZENGA: Well, we're -- what we're voting on is the -- the level -- the degree to which you feel sufficient evidence has been provided to support this process as a performance measure, again, suitable for public reporting or accountability or quality improvement.

We have some sort of guidance, and there's an algorithm which we -- it should be in your -- oh, it's not in here. It should be in your packet of information. It's these little colored things here, which walks through evidence, or your decision tree with respect to evidence, and there are some sort of standards that we like to see, including a systematic review of the evidence; a statement on the quality, quantity, and consistency of the evidence; we would like it to be based on a

guideline. So there are, again, some sort of, you know, sub-criteria here that we would like you to look at when you're making this decision.

But ultimately, the question is how strong is the evidence supporting this measure, this process measure?

DR. CIMA: So the one thing about practice variation -- now that's unpublished data from a survey from a specialty society, so I'm just -- I mean, it's not really published data, and all this refers to published data.

MR. LYZENGA: I think -- I mean, I think we allow, Karen correct me if I am wrong, for the developer to sort of do their own review of the evidence, and that will -- we can allow that to speak to the evidence, but I will turn it over to Karen, maybe, for some additional remarks.

MS. JOHNSON: Yeah, so the best thing to have would be a systematic review that somebody else has already done all that work and graded the evidence, but in -- if that didn't

happen, then it would be up to the developer to summarize all the evidence for you and tell you about its quantity, quality, and consistency. So is that what you're referring to? I see --

DR. CIMA: No --

MS. JOHNSON: -- a different page up --

DR. CIMA: -- because this is them providing their own evidence, that we don't know where it came from.

MS. JOHNSON: Is this under 1(b) or 1(a), I am sorry?

DR. CIMA: 1(b).

MS. JOHNSON: Okay. So this is the gap information.

MR. LYZENGA: Yeah, we are 1(a).

MS. JOHNSON: So --

DR. CIMA: I thought 1 was all done together.

MS. JOHNSON: No.

MR. LYZENGA: No, we do those separately.

MS. JOHNSON: No, they suggest evidence, and then we'll come back to gap, but I'll go ahead and answer your question. It is okay for the developer to give you kind of proprietary information if that's what they're giving you for gap.

DR. FLEISHER: Collette, and then we --

MS. PITZEN: Hi again.

We had a measure for our diabetes patients, we went from an intermediate clinical outcome of LDL less than 100 to appropriate use of statin, so moving from an outcome measure to a process measure, and that process measure needed really strong evidence that I expect every diabetic patient according to the ACC/AHA guidelines to be on a statin.

I think where this perhaps maybe is a little bit lacking is -- or how it could be stronger is if the patient did demonstrate urinary incontinence, then you would expect that kind of procedure to be happening.

DR. MORGAN: Can -- this is Dan Morgan, I am one of the developers.

And this may speak to some of these concerns. The -- when a stress test is done, and that is through our own proprietary data, and this -- we have evidence that the -- an anti-incontinence procedure is much more likely to be performed, so I think that that -- that while this is a process measure and it may be something that the physician is saying yes, I have done or documented, it does -- we do have evidence that it leads to a change in their care that we know is a highly effective treatment, so that I hope that you can see that it's -- that yes, that it is a test that's performed, but it also leads to definitely a change in treatment plan that leads to a chance in outcome.

DR. FLEISHER: Okay. Why don't we vote?

MR. LYZENGA: So voting on evidence. Your options are high, moderate, low, and insufficient.

DR. FLEISHER: Okay.

MR. LYZENGA: We have 67 percent -- zero percent for high, 67 percent for moderate, 24 percent for low, and ten percent insufficient, so the measure passes on evidence.

And we can go ahead to performance gap.

DR. FLEISHER: Comments? Liz?

DR. EREKSON: So the information that the measure developers present on performance gap is proprietary information that they did from a chart review. They essentially went back and looked at -- from a sample across four sites, they then selected high-, medium-, and low-volume surgeons, and then looked back in the chart to see if this cough stress was performed.

And low-volume surgeons were much less likely to do the cough stress, or at least document it in their records, than high-volume surgeons.

And so that does show a variation in care.

DR. FLEISHER: Any other comments?

(No response.)

MR. LYZENGA: All right, let's go ahead and vote on performance gap. High, moderate, low, or insufficient.

(Pause.)

MR. LYZENGA: So we have 23 percent high, 68 percent moderate, nine percent low, zero insufficient. The measure passes performance gap, and we'll go ahead to reliability.

DR. EREKSON: So this is a measure that's collected from chart review. The denominator is identified, and please correct me if I'm wrong, by ICD-9 and CPT codes, the numerator is identified by chart review.

And the measure developers went back and looked at if they could find if it was documented in the chart, and they had a separate person then review the charts, and their kappa was 0.83.

DR. FLEISHER: Other comments?

(No response.)

DR. FLEISHER: Okay, vote? Do you have a comment?

DR. TEMPLE: I guess I am just trying to understand if the reliability of the reporting reflects the reliability that they'll get when they use this measure in real life.

So is it the intention of the developer that there will be constant chart review, or is it the plans of the developer that the physician will tick off stress tests done, and that will be how the measure is done?

And I am just -- it's the same -- I mean, I am just curious if the plan is always chart abstraction versus physician self-report with some subsequent audit.

DR. PRESTON: You know, I think for now, chart review -- we're kind of looking at chart review as sort of really kind of having to be essential, and from my point of view, the problem with, you know -- we all love tick boxes, I would love to have a tick box, but I think we all know that the reliability -- it's a lot

easier to check a box whether you did something or not than it is, you know, making sure you've documented it in the chart, somebody went back to the patient and asked.

We can't say it would verify it at least most of the time --

DR. FLEISHER: In the specs, what does it say? Because we are voting on the measure as is, and I --

DR. PRESTON: Yeah, it's -- well, as is is chart review.

DR. FLEISHER: It is chart review in the specs, and therefore that is what we're approving or not approving.

DR. TEMPLE: Right. I just wanted to make sure from a reliability perspective that that's what it was.

DR. FLEISHER: Thank you.

Any other comments?

DR. YATES: Yes.

DR. FLEISHER: A.J.?

DR. YATES: I just -- my comment is

that there's a lot of things in a busy clinic that you comment on because they're pertinent positives, but it doesn't mean that you comment on every pertinent negative.

Now, this would make that pertinent negative more important to comment on, if they're going to adjudge your quality by whether or not you're doing this part of the exam, but I would -- I would say that this may be under-reporting what is actually done because it may be done by more surgeons than the chart review shows, but they just chose not to comment on it in the body of their dictation.

DR. FLEISHER: Thank you.

Other comments?

(No response.)

DR. FLEISHER: Vote?

MR. LYZENGA: Voting on reliability. Your options are high, moderate, low, or insufficient.

(Pause.)

MR. LYZENGA: All right. We have nine

percent high, 83 percent moderate, nine percent low, zero insufficient, so the measure passes on reliability.

We can go ahead to validity.

DR. EREKSON: So again, in the validity, this is the proprietary information that the measure developer is presenting, and they divide surgeons into high-volume, low-volume, and medium-volume.

Oh, I am sorry. So one of the face validity testing is that the high-volume, low-volume, and medium-volume surgeons perform the cough stress test more often. The other is that when a cough stress test is performed, that the anti-incontinence procedure is more likely to be performed at the same time.

DR. FLEISHER: Amy?

MS. MOYER: So a concern I have if I am reading this correctly is there is no exceptions to the measure, so if a patient comes in and says yeah, this is an issue, they still have to have the tests to confirm, is what I am

reading, which seems like it could lead to over-testing or over-utilization.

DR. EREKSON: So the --

DR. FLEISHER: Comments from the developer -- oh, Liz?

DR. EREKSON: So the measure is described -- and I really read through this very, very critically because there's very invasive, very expensive testing called urodynamic testing, and then there is a reduction cough stress test that does not even necessarily require catheterization.

And so as it is written, they are not talking about the very invasive, you have to get catheterized, you have to come in for an hour to an hour-and-a-half-long procedure, it is during your pelvic exam when we're assessing for surgery, was the cough stress test performed, so the pelvic exam is performed with a full bladder.

So I don't think it necessarily produces a huge burden or increased risks to patients the way it is written.

DR. PRESTON: I would agree with what Liz has said, I think that was you, Liz.

This -- you know, for those of us who are -- who do this as our primary clinical work, this is a part of pretty much every evaluation. It takes ten seconds, and there is no -- there is no resource used other than, you know, my breath that says cough.

DR. CIMA: Is this something that is done by all members of the specialty?

DR. LEVY: Yes.

DR. PRESTON: Yeah, I would say so.

DR. FLEISHER: Other comments?

(No response.)

DR. FLEISHER: Voting? Yeah, let's vote.

MR. LYZENGA: Voting on validity: high, moderate, low, or insufficient.

(Pause.)

MR. LYZENGA: We have four percent high, 87 percent moderate, nine percent low, zero insufficient.

The measure passes validity, so we can move on to feasibility.

DR. EREKSON: So again, for feasibility, this measure identifies the denominator by CPT codes, and -- CPT and ICD-9 billing codes, but the numerator is a chart review.

DR. CIMA: I have a question for the developer on that because I didn't quite -- so since the procedure is done in a hospital, that's where the ICD-9 codes are going to reside. This is an office-based procedure, so you're going to have to go back to your record. How are you going to merge those? How are you going to operationalize that?

DR. PRESTON: That's a really good question.

DR. LEVY: From a practical standpoint, you also code in the office, when you get back to the office, you code for your procedure using ICD-9 or -10 and a procedure code, so your clinic note will identify an ICD-9

or -10 code that you're evaluating the patient for prolapse.

DR. CIMA: Well, I'm just trying to figure out if it's linked -- you're going to identify groups that actually had the surgery --

DR. LEVY: Correct.

DR. CIMA: -- but then you may identify in your clinic patients whom you're evaluating for pelvic organ prolapse who then you have to -- you don't -- so how -- you know, some are going to go to surgery, some are not, so I am just wondering how you --

DR. LEVY: So you'll pull --

DR. CIMA: -- because you may have two different populations.

DR. LEVY: You'll use your administrative data in your office to pull those charts of patients who have had a hysterectomy for prolapse, and then you'll do the chart review to see was the cough stress test done.

DR. CIMA: Glad you'll be the one selling this to your group.

(Laughter.)

DR. PRESTON: No, but it -- from a practical standpoint too, this is something that should -- you know, most of us -- I will speak for myself, I would put in my admission note when the patient is coming for surgery as part of the evaluation under my physical exam.

And I don't think -- you know, it's not a big burden to include that in your note, and I think if people know they're being measured on it, they'll make sure they put it in.

DR. MOSS: So since the criteria can be satisfied by a ten-second interaction in the physical exam, will that mean that the abstracters will read the whole physical exam, and what level of sophistication will be required to figure out whether they did it or not, and how much variability is there in the way people might document that in their physical exam?

DR. FLEISHER: Any comments?

DR. PRESTON: I would think that it would actually be -- what we looked for when we

did this in our institution was for a description either of Valsalva cough -- Valsalva stress test or a cough stress test, and usually, there is really standard language around it that stress incontinence was or was not the module.

DR. FLEISHER: A.J. and then Amy?

DR. YATES: Yeah, I mean, if you are going to try to harmonize the outpatient record with the inpatient data, it just sort of strikes me that this is headed towards a reporting requirement for SCIP sort of thing where we all had -- as surgeons, we all had to make our attestation that there was antibiotics given and antibiotics expected to be given within four hours after surgery and completed within 24 hours after surgery and DVT prophylaxis was to start tomorrow.

And it may be that when someone is doing a hysterectomy, that they're going to now have to have their attestation as part of their operative note that a urinary stress test was done in the office before this hysterectomy and

the results were, and did it influence the outcome of this procedure.

That might be where you collect your data, but that's --

DR. FLEISHER: Thank you.

DR. YATES: -- but I -- but that may -- but that may be what evolves to make this happen, because we don't know where the data is going to be put.

DR. FLEISHER: Correct, but that is actually not our -- that is not what we're voting on now. It's how they will implement it.

DR. YATES: Well, it's not --

DR. ADAM: Hello there, this is Rony Adam. I am another one of the measure developers, and I got on a little late, I apologize.

I also wanted to mention to everybody that this is also intended to be a registry measure, which would take care of a lot of the difficulty in an outpatient versus an inpatient because it's already going to be loaded onto the

registry.

DR. CIMA: So where is the registry right now?

DR. ADAM: AUGS.

DR. CIMA: And you're -- it's in AUGS?

DR. ADAM: Well, it's in -- it is fairly advanced implementation. We have already started the registry now, we're getting --

DR. CIMA: So the registry is for every gynecologist, or is it a specific group that's going to use it? Because this becomes the issue. We're going to be --

DR. ADAM: It's for anybody, anybody who does pelvic floor disorder surgery, whether a gynecologist, urologist, or a urogynecologist.

DR. CIMA: So that's the other issue, though, that I was going to bring up on the feasibility, is there are other people that do this other than gynecologists, correct? Urologists will also be doing this, so -- and they're not on the development team here, so if they're going to be asked to be measured, has

there been a consultation with them?

DR. PRESTON: Well, they're not going to be doing hysterectomies for prolapse.

DR. SAIGAL: Well, I -- I mean, as a urologist, at our center at UCLA, we do do them. I mean, I don't do them personally, but we have a pretty active program, but maybe it's not nationally common, I don't know.

MS. PITZEN: I just wanted to share, technically, it is possible, when you're collecting data retrospectively, to use the CPT procedure codes that a surgeon is billing actually from his office for the procedure as a way to pull in your denominator, and then you're collecting your numerator information based on that denominator.

DR. FLEISHER: Thank you. Liz, did you have a -- ?

DR. EREKSON: I just wanted to comment on my understanding of the registry, and I know there's measure developers on line as well.

The registry was created as a multi-

stakeholder panel between the FDA mesh companies and AUGS, and what it is, it's a very feasible registry that's an online interface, so it's very, very easy to enter data into, and could you talk to the cost of participation in the registry? And there is no requirements of membership in any society to participate, correct?

DR. ADAM: Correct. It is not designed for one specific subspecialty.

I don't know the cost offhand, but it is -- if anybody is on line that recalls the cost particularly, please chime in, but it is not onerous.

DR. FLEISHER: Hello?

MS. HUGHES: Hi, this is Colleen Hughes. It is going to be $195 for each participant, and that's a yearly fee, so it's very affordable.

DR. FLEISHER: Thank you, John?

DR. ADAM: That's what I was going to say.

DR. FLEISHER: John?

DR. HANDY: Yeah, I just wanted to mention the pre-meeting comments, the American Urological Association endorsed this measure.

DR. FLEISHER: Thank you.

So our focus should be on how the measure specified is what we vote on. It is great that we can give them suggestions of how to implement it, but that's separate and outside of this.

So if we can vote on feasibility.

MR. LYZENGA: Feasibility now: high, moderate, low, or insufficient.

(Pause.)

MR. LYZENGA: We have five percent high, 68 percent moderate, 23 percent low, and five percent insufficient, so it does pass feasibility.

So we'll go to usability and use.

DR. FLEISHER: Any comments?

(No response.)

DR. FLEISHER: Why don't we vote?

MR. LYZENGA: All right, voting on usability and use: high, moderate, low, or insufficient information.

(Pause.)

MR. LYZENGA: Five percent high, 64 percent moderate, 32 percent low, zero insufficient, so the measure passes on usability and use, and we'll move to overall suitability for endorsement.

We can go ahead and vote unless anybody has any comments.

(No response.)

MR. LYZENGA: All right, voting.

(Pause.)

MR. LYZENGA: Seventy-three percent yes, 27 percent no. The measure passes.

So that means we will move to the ASA measure.

Ah, break, you're right. We should do a break.

DR. FLEISHER: We have a 15 minute break, and then we'll try to start up quick. I

think we've gone through a lot of the gaps, and I think it will only take about five minutes to assign people gaps to do it afterwards, so the goal is still to get home -- by 5:15 at the latest, go to public comment.

(Whereupon, the hearing went off the record at 3:59 p.m. and resumed at 4:13 p.m.)

DR. GUNNAR: So, this is Measure 2681, Perioperative Temperature Management. It is a new measure and the developers are the American Society of Anesthesiologists who we have here in the room.

And so, if you will present, sort of tee it up for us, we'll go from there.

MR. POPOVICH: Sure, thank you, Dr. Gunnar.

Can I just ask if Dr. Jim Moore is on the line? Okay, he was on the line before, can I just check to make sure that he's on?

DR. MOORE: Hello, this is Jim Moore on the open line phoning in on behalf of the ASA.

MR. POPOVICH: Great, thank you, Dr.

Moore.

Okay. Good afternoon, my name is Matt Popovich. I'm the Director of Quality and Regulatory Affairs at the American Society of Anesthesiologists. I'm joined here with Tom Miller, Director of Health Policy Research at ASA and joining us on the phone is Dr. James Moore who chairs one of our ASA committees that develops, reviews and provides expertise on measures aimed at improving the care of anesthesiologists provided to patients -- that anesthesiologist provided to patients to each day.

Dr. Moore is a Clinical Professor and Administrative Director of the Preoperative Evaluation Planning Center in the Department of Anesthesiology and Perioperative Medicine at UCLA.

The measure before you is a Perioperative Temperature Management Outcome measure. The measure was developed in 2010 as a revision to a measure this committee is very well

aware, NQF 0454.

That measure in particular allowed either a process of care, the active warming of a patient or a patient meeting a temperature threshold at the end of anesthesia time as counting towards successful performance.

That was changed by the -- or the ASA felt that more emphasis should be placed on the outcome, that a temperature of 35.5 degrees rather than a process of care would prove better quality given to the patient.

ASA maintains that this is an improvement over the previous measure, as mentioned in our submission as it focuses on a pure outcome of a vital sign rather than a process of care.

We are introducing this measure, as Dr. Gunnar noted, as a new NQF measure.

As part of this initial presentation, I ask that Dr. Moore provide a few introductory remarks on how this measure can drive quality improvement at the local level and the patients

under care that this measure provides to patients and patient outcomes.

Dr. Moore.

DR. MOORE: Thanks, Matt. Thanks, everyone, I appreciate the opportunity to address your committee.

This revised measure evaluates the percentage of procedural patients who undergo general or neuraxial anesthesia of at least one hour duration and for whom the temperature of at least 35.5 degrees Celsius was recorded within 30 minutes before or 15 minutes after anesthesia end time.

As Dr. Popovich mentioned, it is intended to replace Measure 0454 which is potentially a set of a process of the applying active warming for the patient.

Anesthesiologists have struggled to design rigorous outcome measures that reflect meaningfully the quality of the care we provide.

One reason for this is that many of the important clinical outcomes for our patients

don't depend only on what we do.

In preventing hypothermia, we potentially prevent a variety of clinically important complications, include morbid myocardial outcomes, surgical wound infections, coagulopathy, increased transfusion, delayed wound healing and delayed recovery from anesthesia.

And beyond these sequelae, hypothermic patients also experience increased shivering, and discomfort. Temperature management is something anesthesia providers do have direct control over once the patient arrives in the operating room.

We believe holding anesthesia providers accountable for preventing hypothermia will improve patient care and reduce complications through preservation of normothermy.

Allowing a process which is active warming to fulfill the current measure the current measure without demonstrating that the warming effort would be effective is less

desirable.

So, while adverse outcomes associated with hypothermia occur on a continuum of temperature, and there's no one temperature that can be used to draw clear lines for all patients for all operations. The selection of 35.5 degrees as the target was the best opinion of the experts at the time and remains the selection best supported by the current clinical evidence.

And the ASA appreciates your consideration of this measure.

DR. GUNNAR: Dr. Grover?

DR. GROVER: Yes, let me start out. That was very well stated and I won't repeat all of that.

Before we go down the various avenues here that we go down, though, I would like clarification on one area and that is the way that this is stated, it's either a pre-op temperature or an early post-anesthesia temperature.

And it seem to me like you have to

have the post-anesthesia temperature or an intraoperative pressure. And just with the way -- I mean if it said pre-op and post-anesthesia, that would be fine, but or, you could just do a pre-op and you aren't going to know whether the patient developed hypothermia.

DR. MOORE: That would be correct, sir. And the intention and as it's currently worded that the temperature would be within 30 minutes before the anesthesia end time or 15 minutes after the anesthesia end times.

And as a practical consideration, although the time periods may really be somewhat arbitrary, they're at least based in part on the notion that near the end of the operation, and this is an intention to capture a temperature near the end of anesthesia care, it's a logical accommodation to differing systems of care and systems of data recording and capture.

Clearly, it's at the end of care, within 30 minutes of the end of anesthesia time but also potentially within 15 minutes after.

The temperature does not change quickly near the end of anesthesia care, but we do want it to be reflecting that general period of time. And near the end care, a variety of things can happen that can dislodge the probe unintentionally or as there's preparation to leave the operating room, it's intentionally withdrawn.

Also, the timely measuring of temperature in the recovery room is not always consistent. So, if we capture a temperature within 15 minutes of the end time in the PACU, that does suffice. We do expect to need to have the times recorded during anesthesia care which is the 30 minutes before as well.

And as an aside, the current measure 0454 specifies 30 minutes before or 30 minutes after the anesthesia end time.

Does that address your concern, sir?

DR. GROVER: Well, as long as it -- I don't like the or part. I mean you've got to have -- you can do it before, but there's got to

be during and after.

And I think you said you've changed that and that was my main criticism of that. But otherwise, I think excellent measure that we can go into.

DR. MOORE: Thank you.

DR. SAIGAL: Can I ask a question?

DR. GUNNAR: Yes.

DR. SAIGAL: So, the way you're measuring this here with the one measurement, that is consistent with how the evidence was developed that hypothermia leads to arrhythmias so forth.

Like, in those studies they looked at one temperature measurement or how do they -- if you measure this, we'd be capturing the relevant, you know, predictive factor? That's my question.

DR. MOORE: Well, the intent is to reflect the quality of anesthesia care provided in warming the patient or maintaining normothermia by virtue of the temperature as it exists towards the end of case.

There is quite a lot of variability in patients with respect to the start of the case and how they often come in hypothermic.

It's also a practical consideration as to how to design a measure that's reportable and consistently so. Which is not to say that the temperature should not be measured throughout anesthesia care and anytime we apply active warming, it is the intention that, for appropriate practices, that those things will apply.

As far as the time point with respect to the measurements, there have been studies that looked at a variety of time points and some that don't suggest a threshold in the comparative groups but rather find that patients who are actively warmed versus those that aren't have quite different temperatures.

So, there wouldn't be one way of describing all of the various papers that we cited with respect to that particular consideration.

DR. GROVER: The premise here as I understand it, and is that the -- I mean the evidence is really Class 1A in terms of hypothermia and the adverse periprocedural hypothermia and the adverse outcomes.

So, switching over to outcomes only are showing that you're preventing hypothermia, you would assume then that this would be correlated with less adverse events.

And I don't want more questions on this or move into the evidence, Bill, or you all?

DR. GUNNAR: I think we have one more colleague.

MR. LYZENGA: Yes, I was going to ask Kelsey if she wanted to comment here, backup discussion.

MS. MCCARTY: Sure, just I mean I do think that the wording of the measure is confusing. I had the same thought that Dr. Grover did when I first read it that it was before the start of the anesthesia and the last 15 minutes as opposed to a contiguous 45 minute

period.

So, I do think that that part could be improved. But, generally, I think, you know, there's a lot of literature to support that looking at that postoperative hypothermia does -- is a bad outcome and has associated downstream bad outcomes and that it is a good thing to measure it and try and reduce it. I think there's strong evidence for that.

DR. GUNNAR: Thank you.

Collette?

MS. PITZEN: Just a technical question for staff.

I guess I view this as an intermediate outcome measure and not a pure outcome measure.

MR. LYZENGA: That was our assessment as well, I think, in our preliminary assessment here. We did identify it as an intermediate clinical outcome and really the difference there is the level of evidence required and they have provided the level of evidence required for an intermediate clinical outcome.

So, we can -- I don't know if the developer is also okay with classifying it as a intermediate clinical outcome, but --

MR. POPOVICH: We're okay with that.

MR. LYZENGA: You're okay with that? So, maybe we just change it in the follow-up to this meeting.

MR. MILLER: Yes, my one question is, you have to show the -- or the preface of the question is, you have to show 35 degrees sonogram. Sometime within that 45 minute period, how do you handle other measurements if someone was 34 degrees on three measurements up to the 35 degree five minutes before the end of the case and then hit the recovery room at 33?

Do they still get credit for the 35 one time if they were able to warm them up when they saw they weren't quite warm enough? Is it enough to have the one time 35 degree?

DR. MOORE: Well, generally, we do believe that it is if you have a temperature of at least 35.5 degrees. The temperature does not

change quickly if the measurement is accurate.

We do have problems, at times, where when there can be significant variability in the probe reading, in which case, it may not be reliable. Normally, that's in the direction of artifactually lower temperatures and not artifactually higher.

However, when patients do start hypothermic, it is more difficult to warm them. Even so, if they have a low temperature at any point or at the start of a case and can be warmed to 35.5 or higher, we do believe that that reflects good care with respect to warming measures and should help to prevent significant sequelae as well.

DR. GUNNAR: So, I was just getting my grounding evidence here.

I mean, for me, from a surgeons perspective, it's easy for me to think, okay, end of, you know, incision, drapes removed, patient extubated, patient out of the room, patient entering PACU.

Is it just me that it might lead to some -- that it's too subjective to say end of anesthesia end time? Because I don't know what anesthesia end time is.

DR. MOORE: And, sir, as a surgeon, you probably wouldn't see it routinely. It typically happens when the care of the patient is transferred from the anesthesia provider to the PACU nurse.

So, typically, it's happening in the PACU. And that commonly may occur 10 to 15 minutes after leaving the operating room once the report is given and vital signs are taken.

So, the idea of being 30 minutes prior to that anesthesia end time or 15 minutes after is to represent the end of the end of the case, generally speaking.

It's also the same wording as in Measure 0454 which also may be a little confusing in that except that the time has changed from 30 minutes immediately after in the current 0454 to 15 minutes immediately after.

I do appreciate the potential confusion generated by saying immediately before or 30 minutes immediately after instead of being more clear about what that is supposed to mean.

DR. GUNNAR: But, let's be practical. Say you have an esophageal temperature probe in. You take that out at the time the patient's extubated. That temperature, that last recorded temperature actually would be your most useful temperature if the rest proceeded as planned, which was patient was, you know, placed on the cart, moved to PACU.

You wouldn't have to rely on the PACU temperature as the first temperature, it would be the one that you used in the operating room at the time of the esophageal temp was removed. Is that -- did my ear catch that correctly from a process point of view?

DR. MOORE: That's correct, sir. Prior to the extubation and the last temperature according to the esophageal temperature probe in someone who was intubated would likely be the one

that would pertain to the reporting of this measure.

MS. MCCARTY: I would just add that not all cases have intraoperative temperature management. And so there are some cases, I know we do some at our institution, where that's one vital sign for certain cases we don't document.

And, so the first temperature reading you get you one pre-op and you get one post-op and those are your temp readings and you really do rely on the nursing teams to capture those measurements. And you don't have the anesthesia team capture it.

So, not all work flows as you described.

DR. GUNNAR: And, for that reason, I'm trying to ground it to a point in time. Maybe it's just me, I think end of anesthesia is -- or anesthesia end time is --

MS. MCCARTY: Well, so for billing purposes, on the anesthesia side, that's actually a requirement for billing.

DR. GUNNAR: Sure.

MS. MCCARTY: And so --

DR. GUNNAR: No, understood.

MS. MCCARTY: Oh, okay.

DR. GUNNAR: Understood, understood.

I understand that it's a point in time that's measured in the chart, I guess.

In relationship to this temperature, the intent of, you know, identifying the temperature upon which you're going to measure quality is anesthesia end time. That can go quite far into the PACU experience, right?

DR. MOORE: It is possible, sir. I would say that typically it may go about 10 minutes at the most in a routine case after the entry into PACU.

And just as a practical matter, if we were wanting to describe a measure by one consistent event, it's just it seems like the most likely to fulfill all intended circumstances, although you could go by, say the end of surgery, there is conflict in determining,

is that when the last suture is done? Is it when the dressing is done? Is it when the anesthesiologist now has control again of the patient?

Whereas, the anesthesia end time is recorded for every case for which we bill.

DR. GUNNAR: No, I'm just saying what -- I was trying to figure why it wouldn't be within 30 minutes immediately before or 15 minutes immediately after the patient leaves the operating room.

DR. MOORE: That would be a pretty close approximation as well. There are cases where when we transfer to an ICU it can take longer than that and we are still caring for the patient.

And I do agree that leaving the anesthetizing location which is usually the operating room, sometimes other places, is another thing that is often reported. Although, at the same time, it is not necessarily commonly recorded in the anesthesia record in all places.

Whereas, the anesthesia end time is. So, maybe in part a practical consideration.

DR. GUNNAR: And again, I'm not trying to be argumentative, but if I take the patient to the ICU with anesthesia, the first thing they do if the patient's still anesthetized under really the effects of anesthesia, the first thing they do is transfer, you know, they end their relationship, transfer to the ICU nursing care and head back to the operating room.

So, actually, the likelihood that, in most cases, that the anesthesiologist is going to be connected to the patient in the ICU is less than if they were in the PACU.

DR. MOORE: Well, I still would argue with a potentially long transfer, it's still incumbent upon us to try to maintain appropriate temperature or warming measures if it's feasible.

And it may be more likely that that temperature after the end time would be the one we would have to look at if we end up with a prolonged situation which isn't likely, I agree,

to go beyond 30 minutes before, but it can happen.

But I appreciate you point. It's some of these choices -- they seem arbitrary, they are also designed to be consistently reportable by mechanisms that are already in place for 0454 as well.

DR. GROVER: I guess I had one other comment and I think I know what the answer is going to be to this.

But, ideally, it'd be nice to know what the lowest temperature was during the operation, but I assume then that somebody could be looking away or that might not be picked up.

So, this would be a more reliable -- what you're describing here is a more reliable way to do it to measure at certain times. Am I correct on that?

DR. MOORE: Yes, sir. I think it's more consistently reportable across practices. Places that have an electronic record system and the abilities from that to determine the lowest

temperature could make that feasible.

Although even with doing things like looking at a lowest heart rate as a contraindication for a beta blocker intraoperatively, often those data are manually abstracted. So, I think you're right.

MS. MCCARTY: I don't know if there's a good place to talk about unintended consequences and this might be a dumb question, so I apologize.

But, is hyperthermia at all a risk and is there any danger of once this is a metric and people are pushed towards doing more warming of going too far in the other direction? Is that a concern at all?

DR. MOORE: I think hyperthermia is a concern any time that we actively warm patients and, even in some cases, without substantial loss of heat during surgery and with patients covered by a lot of draping and blankets, it can occur.

It's incumbent on us to always measure the temperature anytime we're applying active

warming and it is a concern. I would say yes.

It doesn't happen very often to such an extent that it is considered a clinically important sequelae, but I think that there's evidence that the benefits of preventing hypothermia are real that in balance, we, in providing appropriate care for monitoring closely the temperature when it starts to rise, we can take measures to prevent it from rising too high.

MS. MCCARTY: Thank you.

DR. GUNNAR: Any other discussion? Are we ready to vote on evidence? I think so.

MR. LYZENGA: Alexandra, I think we're actually going to call this an intermediate clinical outcome, so we'll go with the next slide.

So, you're options are here, high, moderate, low or insufficient evidence.

DR. GUNNAR: You should have 22, right?

MR. LYZENGA: We have 62 percent high, 33 percent moderate, zero low, zero for insufficient evidence, one with insufficient

evidence with exception.

The measure passes on evidence. So, we can move to performance gap.

DR. GUNNAR: Dr. Grover?

DR. GROVER: You want me to take that?

Well, the measure developers supplied considerable data beginning in 2010 going back from when they had the 0454 measure showing increase in compliance with the measure of the course of years, but still less than a 100 percent when looking at the 10, 20 and 30 percentiles, leaving room for improvement, as they noted, with approval of the processes of care to encourage attention to keeping patients at normothermia.

In 2013, the mean performance score was 95.3 percent. But, again, they didn't -- I'm not sure about the post-anesthesia temperature management in that situation and the author published that 5.8 percent of patient who actually appeared to pass that measure actually did develop hypothermia.

So, I guess my point is, it looks like there is a gap and there's still room for improvement in the lower 30 percentile quartile of practitioners.

DR. MOORE: We would agree with your reasoning.

MS. MCCARTY: Maybe I was wrong, I thought that I read in the measure submission that the 5.8 percent of patients that still had postoperative hypothermia were under the previous definition where it was also possible just to use warming technique.

So, I don't know if it was split out between of the 5.8 percent which only had intra-op warming but not a postoperative measurement or if it was all combined?

DR. GROVER: I wasn't sure, either. But it was related to the previous measure.

DR. MOORE: So, yes, the paper I believe you're referring to from 2014 showed that about six percent overall of patients who met that current measure arrived in PACU with a

temperature of less than 36 degrees. And some of the surgical subpopulations had an even higher incidence than that.

DR. GUNNAR: Any other discussion? So, let's vote on gap.

MR. LYZENGA: Voting on performance gap. Your options are high, moderate, low or insufficient.

I think we can call it. Yes, ten percent high, 90 percent moderate, zero low, zero insufficient.

So, the measure passes on performance gap. And we can move to reliability.

DR. GUNNAR: Dr. Grover?

DR. GROVER: Well, again, from what I read, the reliability appears to be high. With the measure itself, I have it progressing from .733 to .975 from 2010 to 2013, the previous one.

And the reliability for providers improved from .523 to .644 from 2010 to 2013.

So, taking that previous measure and assuming this measure's actually tighter, I think

the reliability looks very promising, realizing this is a new measure, so we don't know for sure.

MS. MCCARTY: We had a slightly different interpretation on the reliability.

I think earlier we had talked about the threshold being .7 and it looks like all of the reliability data is below that.

In addition, I think the metric measure developers, I thought it was in here that they said it's admittedly low reliability based on those and especially gets even worse when you start to take out some of the exclusion criteria.

And that the thought was that as more data comes into the NACORE database that that might be improved over time.

DR. GUNNAR: Developers want to address that?

MR. MILLER: Sure, thank you. I'll start with that.

So, yes, the latter measures that you were talking about were from the NACORE data set versus the claims data which were the previous

measure for NQF. And we were looking at about two percent of the cases and two to three percent of the physicians and providers. So, a much smaller data set, as you can tell, very small, very new.

I took the old measure and I randomly sampled, this is not scientific, but I randomly sampled small sample sizes like we have now and that really kicked down the reliability measures quite a bit.

So, we're pretty confident that most of the low reliabilities could have meant that we have only a couple hundred physicians and only three facilities is due to the low sample size.

So, we have seen increases here but we think that as the data gets more and more populated, we'll see those reliability measures go up to what we saw before, hopefully, in the prior measures since the change is pretty small.

MS. MCCARTY: And this metric proposes to only take the data from the NACORE database or will it be going through the claims data as well?

MR. POPOVICH: Sure, we have a CPT code that is applied to this, so using the previous measure which also had a CPT code, we would expect that more data would be available more than NACORE.

MS. MCCARTY: Thank you.

DR. GUNNAR: Collette?

MS. PITZEN: A question for the developer.

Did you undo some data element reliability testing? So, in terms of the numerator, was the temperature that was provided through the database, was that actually the right temperature within that time frame?

MR. MILLER: Right, no, we did not. To my knowledge, we did not do any, though, I believe in that sense.

MS. MCCARTY: I'm sometimes not sure which -- if this comment is good for reliability or validity, but since Collette mentioned it, along those lines, I'm wondering to what extent equipment kind of plays into this?

So, temperature probes are known to be much more reliable methods of capturing these versus the forehead stickers or sometimes even the wands can give really inaccurate readings.

And we've had cases where, you know, those types of equipment that are used in more out of main OR areas have been used in, you know, due to clinical judgment, things didn't look right so someone pulled the temperature probe and got much different readings and it turned patients were having issues.

So, to what extent can we really trust that people aren't having hypothermic issues if we don't know what their processes for obtaining those measurements are?

MR. POPOVICH: Sure, Dr. Moore, can you speak to that, please?

DR. MOORE: Yes. Well, yes, certainly the site of temperature monitoring is important and ideally, we give most credence to true measurement of core temperature monitoring which can be done with an esophageal temperature probe

and some other means especially when someone is under general anesthesia and intubated, that's quite feasible.

When you mentioned the skin temperature monitoring, that actually does correlate fairly closely with core temperature. Although it's not a true reflection of core temperature, it is pretty close, although, generally, it tends to run about two degrees less than the corresponding core temperature.

So, when the forehead skin monitoring is used, you will get a lower measurement even though that may reflect, say, a temperature of 34 degrees on a skin temperature on the forehead may reflect a core temperature of 36 degrees.

The measure does not specify the site of measurement and it's certainly a valid point.

And I apologize if I didn't quite hear the whole question, I have been on the road and left a conference room to get into a cab and I just left the cab. So, if I didn't address everything you asked, please let me know.

MS. MCCARTY: I think mostly addresses it, but based on that, I do have a follow-up question.

I'm just curious where that data comes from on the stickers being consistently two degrees low. Is that sort of -- is that consensus or is there any kind of data showing that that's the case?

DR. MOORE: I do have papers and citations consistent with that which I can pull up and try to give you as we're speaking. I need to pull over from where I am a bit.

But, it's pretty well established that specifically a forehead skin temperature reflects fairly well a core temperature.

However, generally, skin temperature does not and there can be limitations, especially when this measure applies to both neuraxial anesthesia as well as general anesthesia without intubation.

There can be limitations as to whether we can reliably get a core body temperature which

can also be obtained by femoral artery catheter probe, nasopharyngeal probe which isn't always that easy to place correctly to get a true core and some tympanic probes, although those can also be not necessarily widely available in the operating room and also difficult to place that.

DR. GUNNAR: So, I think that to frame this, I think the question is, is it true that esophageal temp, rectal temp and a blood temp are reliable or are equally reliable and that skin temperature is less reliable than the first three, is that a reasonable statement? Or would you say they are all equally reliable?

DR. MOORE: I would revise it to say that, in the interest of reflecting a true core body temperature, that esophageal, pulmonary artery and when placed correctly, nasopharyngeal can well reflect core temperature.

Rectal probes, not as much. Rectal probes do not as reliably reflect the core temperature and neither do generally skin surface temperature monitoring.

There are other technologies that can be used, but even when we do the esophageal temperature probe, you must place them distally enough to get an accurate reading or it may itself not reflect the core appropriately.

And the skin surface temperatures are generally considered lower than core and I can provide some references specifically on that correlation.

But there are special cases of skin temperature monitoring such as temporal artery thermometers which are a different kind of measuring a region of the temporal artery near core temperature and that the supervening skin temperature should approximate core temperature in those. That is also not widely used in the operating room.

MS. MCCARTY: One of the themes we've been talking about today is sort of the workarounds that might come from making some of these measures.

And just to follow on Dr. Yates'

earlier question about how you might have an anomaly of a reading that might be high whereas the pattern has been that the patient is consistently hypothermic and you kind of addressed that part of it.

But, in general, I'm just curious from the developers if you have any concerns about behavior changes that might be unwanted as a result of making this a metric and workarounds that people might develop that would get away from the spirit of what you're trying to accomplish?

DR. MOORE: Well, one thing I'd expect is that people may go more away from that type of forehead skin temperature measure that you mentioned and that we commonly used for a wide variety of cases because the number won't suffice even though we think it reflects by virtue of the relationship of the core temperature what may be adequate normothermia, though it may lead to more expense for probes that would replace such a use.

And I think practically speaking, it's

still going to be difficult in cases of neuraxial anesthesia to get a valid core temperature because none of the modalities we commonly use are very easy to apply in someone who's not intubated and especially with neuraxial anesthesia.

DR. GUNNAR: All right, we're going to move along. If there are -- thank you for those comments. Unless there's any other new position, we'll vote on reliability.

MR. LYZENGA: Go ahead and vote.

We have 20 percent high, 65 percent moderate, 15 percent low, zero insufficient.

So, the measure passes on reliability and we'll go to validity.

DR. GUNNAR: Dr. Grover?

DR. GROVER: Well, on validity, they use face validity for one with 23 physicians with a five level rating scale, five being the highest and with maximal agreement, and the mean rating there was 3.78 out of five with four disagreeing, three being neutral and 16 agreeing or strongly

agreeing.

They also investigated the effect of adding the exclusions, as I think you mentioned earlier, which was cardiopulmonary bypass and what regional blocks, I think. And they found if they added those in, that actually decreased the validity which went along -- which confirmed the fact that it was ideal to exclude those people.

So, I think, by and large, the measure met the validity testing.

MS. MCCARTY: I'm just curious if you're able to share, if you know, what were the reasons of the four people on your committee that weren't in favor of it?

MR. POPOVICH: Well, the process, we didn't solicit them for actual, you know, why did you vote this way. We have a diverse amount of opinions on the measure expert panel and I think it was four that you mentioned, but we didn't go into any detail about why they may be not fully agree or strongly agree.

DR. GUNNAR: Any other comments?

Dr. Yates?

DR. YATES: My one comment is basically circling back to my original question but it's more relevant to validity at this point. And that is, is the validity of this as good as it could be if it's just one measurement out of several in that same 45 minute period?

And would it not be a more valid measure if you at least had two contiguous points in time that were 35.5 or higher? And then you might be able to at two points make a line and a line makes a trend, it's at least graphical for me in terms of optics and I can see that being more valid.

But, and I can't change how they have written the measure, and I can see that, but it would be interesting to see whether or not the reliability and validity improved if they were forced to use two contiguous points in time as opposed to one.

But, that's my only comment and it's not meant to be a suggestion, it's an

observation.

MS. MCCARTY: If I could comment on that just to play devil's --

So, I agree with you and I kind of made the same comment earlier and that does concern me. But, I also, in thinking about the work flows, I know that there's many situations where you might only have one reading and that might be with the sticker, other types of measurement tools where I can see a major increased burden or increased cost to having people repeat it.

And I'm a little bit more concerned about that impact of having multiple measurements than the ideal of only having one measurement as a starting place.

So, just to throw that counter argument out there for the record.

DR. YATES: And to make it a conversation, I would argue that the -- I find it hard to believe that within 30 minutes of the end of the surgery or the anesthesia time there

wouldn't be at least one last temperature in the operating room and certainly the PACU nurse gets one when the patient hits the door.

So, I'd be surprised if those two aren't out there somewhere. And that would --

DR. MOORE: I mean so, that can pertain to how the values are reported. When traditional paper records are used, it's probably most common that the temperature values are recorded more frequently than every 15 minutes.

With electronic records, documentation you may have it every single minute.

DR. CIMA: But you do get into issues of technique. We noticed that when we were doing SCIP. You go from esophageal probe under general anesthetic in the OR and then you go to a tympanic probe in the PACU.

And what I didn't know then is that tympanic probes, there's, you know, there's an art to doing it right. And if you actually throw it in -- if you just throw it in at the wrong angle, you can get a three, four degree

temperature difference, you know.

And so, if you want to -- if you start with probes and trends, that's going to be a very hard case and I think that even counts in the OR as we do more and more local regional blocks, I mean our orthopedic practice hardly ever gets a general anesthetic now. So, you're going to be using different technologies and comparing technologies is a big hassle.

So, I don't know how the developers would handle that.

DR. YATES: I don't mean to be all sentimental and go back in time, but, boy, I can remember when you always asked, when you had that low temperature, you always asked for the mercury thermometer.

DR. MOORE: I agree, tympanic measurements is tricky to do correctly. When it's done well, it does reflect the core temperature. It's often not done perfectly.

DR. GUNNAR: All right, I think we can -- any other discussion on validity?

Collette?

MS. PITZEN: I appreciate the face validity that was submitted. However, with this piece of clinical information that's coming from hospital systems, I really would have liked to see some data element validity testing.

You're asking for a range of a time period of when that temperature can be provided and that you're providing the right temperature during that time frame.

So, I do have concerns about the validity testing.

DR. GUNNAR: Any other comments?

Shall we vote?

MR. LYZENGA: Let's vote. Voting on validity, high, moderate, low or insufficient.

We have 15 percent high, 65 percent moderate, 20 percent low and zero insufficient.

So, the measure passes on validity. And we'll go on to feasibility.

DR. GUNNAR: Fred?

DR. GROVER: Well, this is, I mean you

guys in anesthesia can correct me, but this is pretty much routine practice. I mean measurements, you're on the soup in terms of practicing evidence-based medicine.

So, I think this is quite feasible. You're trying to keep it relatively simple by collecting a few. You know, those of us that are, you know, in the OR might say, well, why don't you pick on in 30, too? But again, you begin to start decreasing maybe the reliability of your data and we understand that.

So, I think overall this is quite feasible.

DR. GUNNAR: Kelsey?

MS. MCCARTY: I agree.

DR. GUNNAR: Any other discussion?

Collette?

All right, no problem. Let's vote.

MS. MURPHY: May I speak

DR. GUNNAR: Yes, yes.

MS. MURPHY: It speaks to abstraction of paper records and it's not specified for that,

so could you speak to that question?

MR. POPOVICH: Sure. Collecting this information, it's part of a process of care that would typically take place within 24 hours after, you know, for coding, for CPT coding.

So, in that regard, some abstraction would have to take place, but it would immediate as a process of care. The data is readily available as a vital sign.

Dr. Moore, did you want to add anything to that?

DR. MOORE: You're correct and the data should be readily available, although there may be cases where abstraction need to occur. The mechanisms are in place.

MR. LYZENGA: All right, if no other comments, let's vote on feasibility.

Your options are high, moderate, low and insufficient.

We've got 58 percent high, 37 percent moderate, 5 percent low, zero insufficient.

The measure passes feasibility. And

we'll move to usability.

Any comments from the committee on usability or use?

DR. GUNNAR: Fred?

DR. GROVER: Well, I think, I mean I assumed reading this that this would be communicated to each provider and institutions so they know what their results are.

And then I see that you're also planning to allow professionals to report this measure to PQRS and the QCDR reporting mechanisms beginning in 2015.

So, I would think there are no unintended consequences, I would think, that amount to anything, so I would think this is very usable.

MR. LYZENGA: Any additional comments? Seeing none, let's vote on usability and use.

And 53 percent high, 42 percent moderate, 5 percent low, zero insufficient.

The measure passes on usability and use and we'll go to overall suitability for

endorsement.

Any discussion before we vote? Appears not, let's go ahead and vote on overall suitability for endorsement.

Unanimous yes, 100 percent.

Thanks to our developers, we appreciate your coming here.

MR. POPOVICH: Thank you.

MR. LYZENGA: So, we owe our representative from Yale, Elizabeth can join us.

To continue, we will actually do the gap analysis probably by email or on the phone call. And perhaps you can send some stuff out.

So, this is a measure of Admission After Outpatient Surgery and our discussants are Dr. Cima and Dr. Grover.

So, Dr. Drye, you want to introduce yourself and give us a brief overview?

DR. DRYE: Hi, my name's Elizabeth Drye. I'm a Director at the Center for Outcomes Research and Evaluation at Yale and I directed the development of this measure along with some

very talented staff, including our lead who is in Australia and couldn't be here today.

I'm trained as a pediatrician, not as a surgeon, so you are really the experts on this topic and it involves all of you because it's a broadly defined measure.

I was just going to quickly walk through the process for development, the rationale, a couple key challenges and our learning to date and I'll try to do that in three minutes.

So, we developed the measure under contract to CMS. It's their measure and we used their typical transparent process that included literature review, a national expert panel Dr. Dutton served on, so I think he's recusing himself and we held a public comment period as well, a national public comment period.

In our view, the measure fills an important gap and can help advance quality improvement. As you all know, about 70 percent of surgeries are done in the outpatient setting.

The outcome for the measure is hospital visits within seven days of outpatient surgery, their direct admission or after discharge, an unplanned admission, an observation stay or an ED visit.

And we're focused on that outcome for several reasons. First, in the cohort of patients we're looking at which is Medicare 65 and older, the rate is relatively high at 10.5 percent within seven days.

Second, the causes are often not always often preventable and they include things, as you know, like nausea, vomiting, uncontrolled pain, urinary retention, wound infection, bleeding or more serious complications.

And third, and this is a key point, often the surgical team is really not aware of these outcomes because the patient circles back to the hospital or an ED and the loop is not closed to inform the providers of the patient's outcome.

And fourth, we see variation across

hospitals in the rate of hospital visits after adjusting for the differences in their patient mix and the wide differences in the types of surgeries that they do.

So, the intent of the measure is to really make those visits transparent to providers and patients and support quality improvement.

There are many challenges to this measure, I'm sure you'll flag some of them for me. But, two that I wanted to just highlight was that we wanted to really identify a cohort of same-day surgeries so the patients expected to go home.

And to do that, this is a claims-based measure. We took a conservative approach and we used a list of surgeries that Medicare has approved for ambulatory surgery centers.

So, this measure is designed to profile hospital outpatient departments only, not ASCs. But we used the ASC list because it helps us stay focused on same-day surgeries and we further narrowed it in a couple of ways we can

talk about.

And then, second, we had to risk adjust as I alluded to before, not just for patient comorbidities but for the differences in the types of surgeries across hospitals because the outcome rate really follows from the type of surgery done.

And so we adjust not just for 24, I think it is, patient comorbidities but also for the relative value unit that reflects the complexity of the specific surgery and for the body system operated on, the anatomical body system using a body system classification that AHRQ developed.

And that approach served us well, but that's a little more complex than our typical risk adjusted measure.

And then, finally, I just wanted to note, we've already had some learning to date, even though this measure is not in use. It has a related measure which is on the agenda with a little bullet on harmonization below this

measure.

We developed for CMS also a measure of hospital visits within seven days following outpatient colonoscopy. And that measure right now is we're running, not we but another contractor, is running a national dry run for the measure and they're going to report to both ASCs and HOPDs their measure scores.

And there's a lot of learning taking place during that dry run and some of it applies to the surgery measure because the data processing and the outcome are very similar.

So, we've already learned then a need to handle a particular merged statement in our input files a little differently than we did for the analysis that went into this application.

We rerun all our analyses and things look pretty much the same, so we didn't resubmit the application. But I did want to note that the outcome rate for the population is 10.5 percent, not 10 percent as stated in the application we submitted. It just -- we had missed a few

surgeries and outcomes.

And as that dry run continues over the spring, the scores will be sent to facilities in July for the colonoscopy measure. We may do some additional learning and we may need to jiggle our results here a little. We don't anticipate anything major, but I just want to let you know that we will come back to the committee if we had to change the measure in any way or any of the results change.

So, I look forward to your questions and your comments.

DR. SAIGAL: Quick question. Is just for hospital outpatient or also ASCs?

DR. DRYE: So, this measure is just for hospital outpatient departments in contrast to the colonoscopy measure.

And the quick reason for that is that ACSs, as you know, are really, really highly varied in the kinds of surgeries they do and we just, we would have loved to include them but we couldn't find a way to include them. Some are

also very small, so we stuck with hospital outpatient departments.

DR. FLEISHER: Can we go -- first get the comments from Robert and Fred and then we'll go the floor?

DR. CIMA: As the developer went through, this is to look specifically at hospital outpatient practices, not ASCs, which is a big distinction. And it's looking at returns, unplanned returns, to emergency rooms, admission to the observation status, admission to hospital or other interventions that require hospitalization in Medicare patients.

The rationale behind this really is to see what -- and just sort of putting words from other documents here, to see if providers who are doing the outpatient procedures can improve processes to avoid these events.

It's unclear from the data provided if there's anything they can do about, but I don't think we have enough data to actually say one way or the other.

But that's the rationale behind the measure.

The numerator and the denominator is pretty clear. The denominator is basically, there's a large list of procedures in the Appendix that goes through what I would consider, you know, sort of routine hospital-based outpatient procedures.

There are some exclusions for cataract, eye things are excluded.

And then the numerator is basically anyone that falls into that category showing up within seven days.

And so, it's pretty as to what they're trying to do with the groups. They're well defined and it's based on administrative outcomes.

So, I mean the outset says ten percent who are doing this will show up for some type of need and that's when we sort of get into the details of, well, are you really going to be able to alter that? And that's a different question.

DR. GROVER: Yes, I think Elizabeth did a great job of describing this. It's a pretty complex metric, really.

And I think the importance of risk adjustment like you're doing for the procedure as well as the comorbidities is helpful.

But, one area I think that you really get into in detail that's very, very important in the protocol is how you make the distinction between an unplanned visit versus a planned visit. And you might want to get into a little more detail on that for the group because that's key.

DR. DRYE: Sure, yes, sure.

So, we do, if post-discharge of the patient comes back and is admitted, it could be for a planned procedure for treatment of a condition whether it's something was identified during the first surgery and it's follow-up more definitive care.

And we used an algorithm that we developed for the readmission measures that Yale

developed for CMS that classifies admissions. It was actually focusing when we developed it on admissions, not readmissions as planned.

If they are not for an acute diagnosis like sepsis and they're for something you would typically schedule as an inpatient procedure.

So, there aren't a lot of those in this outcome but we don't count planned admissions. And we've actually pretty much, all our outcome measures now that are looking at readmissions or admissions follow that approach.

DR. FLEISHER: A.J.?

DR. YATES: Yes, a point of clarification, the number ten percent came up. If I'm looking under the worksheet, the range, the risk adjusted range is 0.5 to 2.5, correct?

DR. DRYE: Yes, so I couldn't squeeze everything into my three minutes, but I'm really glad you asked that question.

So, this measure score, we would recommend reporting as a ratio rather than a rate. It's calculated like our readmission

measures. On the denominator, is the -- and like a lot of the STS and other risk adjusted measures -- the denominator is the expected number of hospital visits.

The expected rate is not going to be zero for anyone. It's always going to be some expected admissions. We don't really know what the -- or visits -- we don't know what the really best number we could get to is at this point.

And then the numerator it's like an observed count, it's not an observed because we use hierarchical modeling to account for the sample size variation and clusterings. But it takes into account the observed amount of hospital visits.

So, it's like a smoothed estimate and I think a lot of you on this committee have seen similar models before.

So, the score we recommend reporting as the ratio of -- you can call the numerator different things, but we call it a predicted over expected. And the typical hospital, that would

be one.

And so, if the ratio is less than one, it would be, you know, 0.9, 0.8. They would be having fewer hospital visits than expected. And if it was greater than one, they would be doing, in this case, having more visits than expected, it would be doing worse.

In our other measures and some other measures like all our readmission measures, we multiply that ratio by the national accrued rate of the outcome.

So, in this case, we could have said, well, let's just -- we'll multiply it by ten, so instead of reporting one, we'll say, you know, the average hospital is 10.5. And if you're at, you know, 0.8, you know, it'll be eight.

And we would report that rate instead of ratio because consumers have an easier time with that than thinking about a ratio of observed over expected or predicted over expected.

But, we don't recommend doing it here because hospitals vary so much in what they do.

So, if they're mainly doing orthopedics, you know they're going to have a really low rate than if they're urology or some of the other -- focused on some of the other specialties.

And so, that number then, it just isn't as good of an indicator as it is when we're taking, for example, like a cohort of heart failure patients and we're reporting a readmission rate only for the heart failure patients across hospitals.

DR. YATES: Well, let me rephrase my question because you've corrected me, that's not a percentage, that's the ratio.

But, if I look at the raw numbers of hospital visits out of 212,000 surgeries, there was 4,000. So, you're really talking about a two percent incidence, correct? I mean if I read those numbers right, it would be about two percent.

DR. DRYE: Yes.

DR. YATES: And that two percent incidence, in terms of determining performance

gap, can you, in that small percentage of patients that are coming back, I have two questions.

In that small percentage, can you assure the committee that you're model for the planned readmissions, your protocol for planned readmission capture doesn't have enough noise that it doesn't rattle around inside that two percent which is such a small number?

And then the second question I have is, does that two percent represent a higher number than those patients that were done as outpatients under Medicare in other environments such as an ASC since this is hospital based?

My concern being that the patients that might be higher risk are done in a hospital based ambulatory surgery center or outpatient center out of concern that they might have a complication and might need to be kept.

So, those two questions, are you sure that with that small rate, you're really going to eliminate the planned admissions?

And the second question is, how does this compare in terms of the numbers of people being seen in hospitals after they've been to a surgery center, for instance, that maybe only deals with less ill patients?

DR. FLEISHER: Before you do that excellent questions, let's vote on evidence because this is --

DR. YATES: It's about performance gap. I'm asking is this the performance.

UNKNOWN PARTICIPANT: I would say it's a lot about performance gap.

DR. FLEISHER: Right, so, but we're only --

DR. YATES: I thought that was a gap question because I don't know that I'm seeing two percent of people coming back.

DR. FLEISHER: But this is a health outcome. The simple question we have, is there evidence to say this is --

DR. YATES: Oh, I'm sorry, yes.

DR. SAIGAL: Can I ask a question on

evidence?

DR. YATES: I thought we got to gap.

DR. FLEISHER: Yes.

DR. SAIGAL: About the evidence before we vote on it?

DR. FLEISHER: Yes, the evidence. But this is a valid outcome.

DR. SAIGAL: The question we have here is where there's evidence at the process that they suggest, like patient education and pain medication management, will impact this outcome. Is there evidence that that's true?

DR. FLEISHER: It's actually not even evidence, it's just a rationale really.

DR. SAIGAL: So, can we --

DR. FLEISHER: It's just sort of plausible rationale.

DR. SAIGAL: That's what it seemed like, thank you.

So, can we use that as justification for the measure if there's no evidence and it's just a thought process?

DR. FLEISHER: Yes.

DR. SAIGAL: We can?

DR. FLEISHER: If you find it, you know, adequate.

DR. SAIGAL: Should we?

DR. FLEISHER: Right.

DR. CIMA: That's a different statement, a very important question and I think it'll get answered when we go to the performance gap.

DR. FLEISHER: The question, do you think there's sufficient evidence that this could be an outcome that could be modified?

DR. SAIGAL: Well, there is no evidence. So, basically, the question is really, do you agree with the thought process and does anyone know about any studies that have -- like where they've looked at how do you better educate a patient or jump on their pain management and the discharge planning, if that reduces readmissions?

DR. FLEISHER: There are studies

related to pain management as well as nausea and vomiting measures.

DR. SAIGAL: That reduce readmissions?

DR. CIMA: This is more of a holistic approach because if you look at the top reasons that a provider in the Appendices for why people are coming back, it's not those.

It's urinary retention, urinary symptoms, renal papilla, I mean, these are all measured in percents of things.

DR. FLEISHER: Yes, actually, one of the pieces of evidence is actually a paper I wrote on 750,000 cases for Medicare data and it's --

DR. CIMA: Yes, but nausea and vomiting is in there, but what I'm just saying, when you look at total numbers, it's small.

DR. YATES: And for those on the phone call, for the record, performance gap is on the big screen. So, I'm sorry if I jumped ahead.

DR. FLEISHER: Thank you for putting it out.

Fred?

DR. GROVER: Maybe I'm missing something here, but I had down on the evidence here that you reviewed a number of manuscripts and the rates vary from 0.5 to nine percent, depending on the type of surgery with ones varying from 1.3 percent to 13.6 percent of outpatient surgeries in HOPDs.

DR. DRYE: Wait, the outcome rate --

DR. GROVER: Did I misinterpret that?

DR. DRYE: That's correct. And the outcome rate here is 10.5 percent, but that's the percent of patients that nationally across all hospitals that have a follow-up visit within seven days.

The score is different, it's a ratio. And so, it's low number, one is the average and less than one is better. But you can think about it in your head, you're going to multiply that by ten. The average outcome rate for hospitals here is ten percent and some will be 2.5 times that, that's our max. So, you know, 25 per hundred and

some would be, you know, lower than that.

So, we leave it as a ratio so as not to suggest that any particular hospital's rate is around ten because the surgeries vary so much that the hospital rates vary a lot.

You know, we're estimating the expected and then how well the hospital is doing compared to its peers with similar peer hospitals doing similar surgeries.

Does that answer your question? So, on average it is ten percent and there are studies which we talk about in here about addressing, you know, nausea, vomiting, pain, some of these are straight complications.

And so, again, this measure, like our other risk standardized measures is a peer comparison. So, the fact that you see variation shows that some hospitals are doing a lot better in avoiding hospital visits with these patients than others.

DR. FLEISHER: Other comments on evidence? Okay.

DR. PARZYNSKI: Can you hear me?

DR. FLEISHER: Yes, go ahead.

DR. PARZYNSKI: Hi, this is Craig Parzynski at Yale.

DR. FLEISHER: Yes?

DR. PARZYNSKI: I'm the statistician on this project. So, I just wanted to add that on Table 2 in the Appendix that there are actually 21,000 outcomes and so the rate is ten percent. So, I just wanted to point to the correct table in the Appendix. But, I'm not sure what you all have, I'm just looking at what we submitted.

DR. FLEISHER: Please vote.

MR. LYZENGA: We have lost a couple, so think we'll have a lower number. I think we can go ahead and call it.

DR. FLEISHER: Yes, call it.

MR. LYZENGA: Eighty-three percent yes, 17 percent no.

And the measure passes on evidence. And now, we'll move to performance gap,

opportunity for improvement.

DR. FLEISHER: So, if you can address Dr. Yates' comments.

DR. DRYE: Sure, and I know Dr. Fleisher's done research on this exact issue of this outcome, or almost this exact outcome between hospital outpatient apartments and ASCs and other settings.

Yes, there is a difference, and in our data, there was a difference looking at HOPDs versus ambulatory surgery centers. And the expected rates and the actual rates are higher in HOPDs, as you would expect because presumably, the more in the more complex high risk patients are there.

And that's one of the reasons we didn't want to report this measure simultaneously in the same data set for HOPDs and ASCs. So, we stuck with HOPDs and we tried to just narrow the cohort of surgeries, the group of surgeries in the measure to those that are explicitly identified as same-day surgeries and that they

could be conducted at ASCs.

We don't look at a lot of surgeries that are done in hospital outpatient departments. If they're not the ASC list, they're not in here and if they're more complex or inpatient surgeries, they're not in here.

DR. CIMA: It's on the gap issue and excuse me if I don't understand how this variation was generated. But the hospital variation, that's the big issue we're getting to.

So, we're saying it's ten percent and it's different depending on the surgery, so we can't really use that. We don't have that data.

But then when they go through in the Appendices, it basically says that they found in this part here, it's 211 hospitals that were at 24 percent had fewer than expected and 95 percent -- there were 34 percent that were better.

But then when it goes through when you did an interval estimate, so almost like putting the confidence interval around it, it said that out of those, what was it, 4,200-plus hospitals,

80 hospitals were worse than expected, 4,119, there was no variation and 35 were better. So that works to like one percent was bad and a little bit more than one percent was bad and a little bit less than one percent was bad.

So, I'm wondering about the performance gap here. I mean it's like 98 point something percent are doing as expected. So, we're going to go through all of this and I'm just wondering, that, to me, was the biggest striking thing.

And I don't understand the statistics behind it. So, maybe you can say, but when we're talking about one percent performance gap --

DR. FLEISHER: That was the model.

DR. DRYE: Yes, so I apologize that this measure's coming at the end of the day. It's complex, so please ask your questions.

So, there's two things that apply here. The first is that we only had a 20 percent sample of these surgeries. So, we have one-fifth of the volume at each hospital.

So, when we run this nationally, CMS reports is nationally, they'll have five times as many cases. So, instead of about 50 cases per hospital, they'll have 250 cases. So, we have small sample size.

And the second thing which Dr. Fleisher was alluding to is we use a modeling approach which takes into account sample size and if there are very few cases, it assumes the hospital is a typical performer.

So, it kind of pulls -- it makes those outliers hard to find. It pulls estimates toward the middle because it weighs that we have very few observations and so we're less certain.

So, what we expect to see with this, you have one percent and two percent outliers on each side, like you said, is we'll have five times the cases. We expect to a lot of outliers. We'll have a million cases nationally which is great.

So, as outcome measure developers, we're just plagued when we don't have enough

cases. But for this measure, we expect to see a lot more outliers. We don't have the ability to run that now because we don't have enough data.

DR. CIMA: But our problem here is we've approved a lot and disapproved a lot on a lot less data and we have to go with what's being submitted. And you have 200,000 surgeries and so, I'm just wondering with what we have, as we've said before, in the submission, there is no performance gap.

We're making an estimation that there will be when we get the data. But I'm saying, right now, in front of us -- and that's what was I confused by because we always say we go with what were submitted. I mean, is that not right, Andrew?

DR. FLEISHER: Yes. One of my questions to staff, because we've had this debate at CSAC a lot, of what's considered in and outside the measure as far as quality rating precludes.

DR. DRYE: Yes, I just want to differentiate, there is a big range of

performance when you look at the point estimate. And then we use a very strict 95 percent confidence interval to identify outliers.

I mean it's because what CMS does when they publically report, they use 95 percent intervals.

But you could decide that we could identify, you know, that's just what we ran for this. If you thought, well, if there are 80 percent confident that they're truly different, you're going to see a ton more outliers.

But the key point, no, it's important because as policy people, we're not -- we're trying to see is there really variation there.

When we have five times as many cases, you will see that variation even with the 95 percent interval estimate. We know that. I mean we're going to move from 50 cases per hospital to 250 on average. So, you'll see many more outliers.

I think others who are familiar with this modeling approach, you know, can chime in

on.

DR. FLEISHER: So, let me ask, because the issue is cut points Yale set in the document. Because if CMS says a different cut point for confidence intervals that will change, correct?

DR. DRYE: Yes, but CMS would never -- I mean I can't -- CMS is on the phone, they would never report this with a 20 percent sample.

So, there's two ways to see more outliers, use 100 percent sample which they would do if they use it or if you just wanted to see it for the committee's understanding, we could use a less strict confidence interval and show you how many outliers we get.

DR. FLEISHER: So, my question for NQF staff is, as we rate the measure, they've chosen what's an outlier which is, Robert, your question.

DR. CIMA: Yes, but, so reliability testing, this is part of this also, is they scored it as a 0.5. So, I'm just wondering if this is being treated specially here because

we've said reliability testing has to be 0.7 and what's written is what's written.

And so, I just don't want to get the sense that we're treating this differently than what we normally say.

MR. LYZENGA: And this is something, again, that we've been wrestling with a bit, the question of how the measure score is, you know, is reported.

So, Karen, I don't know if you were listening to this earlier, but we're discussing that if we report these, we've got the observed to expected ratio and we see variation in that score, I think. But when we sort of apply a 95 percent confidence interval, we get -- and, you know, say if you fall -- if you're going above, you know, outside of that 95 percent interval or below, then you're getting a very small number where above average or below average, something like one percent above and one percent below.

But this is sort of the question of how you're, you know, applying the measure to

some degree.

Now, you could configure it in different ways and report it in different ways and get different results.

I don't know if you have any thoughts on that and whether sort of to what extent this committee should be addressing that? Or what is in the purview of our discussion here?

MS. JOHNSON: Well, when I think about gap, I think about variability and I don't think as much about the confidence interval, part of it. To me, that's a little bit more reliability.

So, I think those are the questions that you'd ask for reliability. Can you really distinguish one provider from another?

So, I think the question here is, are you getting variability between the scores? I don't think I'm helping you guys much on this one.

DR. FLEISHER: I guess the way to ask this question is, if they didn't do the modeling the way they're doing it, there'd probably be

more variability. The modeling is creating the lack of variability which is a reliability.

DR. YATES: I have a question on that.

DR. FLEISHER: Yes, A.J.?

DR. YATES: I apologize about the two percent, I'm going by what I interpreted as events out of surgeries and that's the centers.

But out of the ten -- in your Appendix, 65 percent of the visits ended up in readmission. Did you run the proportion difference, the proportional difference of -- I mean that's five percent of the patients of 6.5 percent -- or what and I saying -- 6.5 percent of all the patients ended up admitted.

Did you run that separately and look at admissions? And the reason I ask is because the readmission for total joint replacement, for instance, the readmission model only looks at admissions not at hospital visits or ER visits or ob status.

And since we're dealing with a population of patients that are 65 and older, is

it necessarily a performance gap that out of an abundance of caution, someone calls up and they come to the ER and they end up not having anything or they are put in obs for overnight because they complained of some dizziness and they're okay. Is that as important as the fact that they literally had to be admitted?

And it would be interesting -- I mean do you have that data?

DR. DRYE: I think you might be asking for -- were you asking about the reasons for admission for the admission?

DR. YATES: No, just solely admission.

DR. DRYE: Yes.

DR. YATES: Just in terms of harmonization, the readmission model for the readmission measure that's out there --

DR. DRYE: Yes, yes, I know that one well.

DR. YATES: -- only counts hospital readmissions?

DR. DRYE: Yes, yes.

DR. YATES: It doesn't count ER visits, it doesn't count obs visits. And so, I'm asking you, I guess, one, why did you add those on there?

DR. DRYE: Yes, sure.

DR. YATES: And was there a significant performance gap for just hospital admissions, which still represent 6.5 percent of all the patients that have the outpatient surgery?

That, to me, seems like the bigger question.

DR. DRYE: Yes, yes. So, that's a good question. I don't think that we looked at the performance gap by admissions only during development, but I can check and, Craig, you're there, do you remember doing that? I don't think we did.

But, the reason, while you're thinking about I'm going to go ahead and answer the second part.

Why did we include emergency

department visits and observation stays?

It really goes to the measure concept. These are patients who expect to go home and not have to go back to the emergency department or to go and be observed in the hospital or get admitted because we've limited the surgeries in this measure to surgeries where that's what's expected.

And yet, a lot of these patients go back and I think you -- this is really just think about it from the patient's perspective, this is a patient centered measure design that's saying, hey, we should make these outcomes visible because a lot of people are going back with urinary retention, as you note, with nausea, with vomiting, with pain.

Those are things that happen post-op, but, you know, can they be better managed? Can they be addressed with a phone call or a visit in the outpatient setting or prevented with better planning.

And they're just not visible now,

either the patients or providers. So, we wanted to include those EDFs stays even though they're less severe than the admissions, most likely, and we didn't have a way to sort of -- we could have tried to score them, you know, like as less or more, I mean as less than admissions but we didn't have a good way to do that.

DR. YATES: Yes. Just as a countervailing argument, I mean as for in terms of patient centered expectations, when a patient's discharged from the hospital, they don't expect to come back to the hospital and be readmitted or be seen in obs or in the ER within a few days either.

So, I'm not sure that that follows that it's something special about being in outpatient surgery. I mean the patients discharged from the hospital hope to stay home for the next 30 days. Yet, we don't use these other parameters.

And so, I'm just wondering whether or not those parameters were added to make the

performance gap bigger or you just didn't look for the different.

DR. DRYE: No, they were totally added because we felt that from the patient perspective, those were important outcomes.

DR. YATES: Okay.

DR. FLEISHER: So, Liz, is this about performance gap? Because -- great.

DR. EREKSON: So, when I look at the Appendix that was provided, there's Table 2 and Table 2 shows a wide variation, depending on organ system on whether or not these patients fall into these categories.

And so, male genitalia and female genitalia surgeries in that body are somewhere over 20 percent.

And then when I look at Table 5 and 6 where you're using the model, you're having that 2.3 percent in the low-liers and the 26.7 percent in the high-liers. So, and those are deciles, I believe is what the table says.

So, it does seem like there's a wide

variation in practice and I think that speaks to a performance gap.

DR. CIMA: Yes, but they risk adjust that. I mean that's the whole point and so, you're going to expect higher -- for certain procedures you expect higher but that doesn't -- when you mean the whole body of those procedures, it's still not a huge difference. I mean, I don't know if I'm getting that.

DR. FLEISHER: John?

SR. HANDY: Well, if you do the numbers in the Section 2(b)(5.2), so of the 4,000 providers that you had, 80 were under performers. They were doing worse.

And so, if you extrapolate that to the 200,000 patients, it's 4,000 patients. So, it's a lot of people that are being affected by that even though the confidence intervals drive them out to be small numbers with regard to the providers.

DR. FLEISHER: Thank you.

Why don't we vote? Or, Liz, do you

have -- no, you've already commented.

Why don't we vote on performance gap?

I think we're done.

MR. LYZENGA: We have 25 percent high, 65 percent moderate, 10 percent low, zero insufficient.

The measure passes on performance gap. So, we'll move to reliability.

DR. CIMA: Well, the developers provide that with the model, if we want to use the model as a thing, the reliability score is 0.5. Now, I'm not sure if that's -- I mean that's just because of the modeling.

DR. DRYE: So, the way we're doing reliability here is we're looking at two samples from each hospital of patients to -- we split the sample in half and then we look and see whether the score is the same using one group of patients and the other group of patients so that we're getting at an underlying quality signal. It should be similar.

And we use the interclass correlation

coefficient to look at the comparability of the score.

And 0.5 in the world of risk standardized outcome measures is pretty strong. And here, we're also still not using a full sample, we combined three years to get more cases, but we had to split the sample, so we're using about 30 percent of what each hospital would have. And as you get bigger sample sizes, reliability goes way up.

So, we're happy with 0.5 as it is even with 30 percent compared to other risk standardized measures. We know it will be better when we have a full sample.

It's a different scale, I think, than what you guys are talking about for process measures and the reliability in that context.

DR. FLEISHER: Comments? No? Please vote.

MR. LYZENGA: Voting on reliability.

Is that it?

Twelve percent high, 88 percent

moderate, zero low, zero insufficient.

So, the measure passes reliability. We'll go to validity.

DR. FLEISHER: Okay, Robert, comments on validity?

DR. CIMA: It just depends on if you look at how in the Appendices whether or not you think this really represent things that can be modified. We have our opinions those who actually who do surgery versus those who may not.

But, you know, face validity was on a panel of 13 or 14 people and most people seem to agree that there was some valid, four people were high, most were somewhat agreed.

And so, I mean it's a measure, I think, could it be better? Should it be inpatient versus ED visits? Those are discussions, but I think, you know, people don't expect to go back to need care.

The only question, you know, sometimes, are we making an assumption that all these visits are related to the surgery? We

don't know that, so it could be other reasons. People go to the emergency room for a lot of different reasons.

But those are the three issues that I would think we need to be -- someone would have to process in their mind.

DR. FLEISHER: Fred, any comments?

DR. GROVER: Face validity is pretty much as described. I thought the validity overall was probably pretty good.

I mean, do you have further comments, Elizabeth, to refine that?

DR. DRYE: No, I mean within our TEP which is mostly surgeons and also providers and patients, payers, there was a lot of discussion.

There was some discussion around, you know, combining ED, obs and admissions. But, in the end, I think it was broad agreement that there was valid and we received public comments to that effect as well.

DR. FLEISHER: My one question is, what starts an, quote, outpatient procedure?

Because that can vary greatly in a hospital base because a lot of times, they'll start in the outpatient setting and be admitted.

DR. DRYE: Yes, so that's a great question.

So, you have to bill, and Craig, you're on the phone, I know, and you can --

We only include, again, procedures that are billed as outpatient and that are expected to be same-day procedures. But even that might vary across hospitals.

We wish there was a perfect like code that said, this starts as an outpatient procedure but there isn't that code yet. And so, we tried use very careful claims processing algorithms to make sure we're focused only on outpatient claims.

We, like other measures that you have looked at, we used physician, you know, claims, we look at those settings, we take into account issues like the three-day payment window and make sure we're capturing those. There's no 100

percent clean way to do it.

Some of the learning that we'll do with the colonoscopy measure will help us here as well. And when, you know, when providers -- the way that CMS reports this type of measure, they provide patient level data to facilities and facilities can look at it and, you know, dispute it.

So, I think there's still some learning to do to make sure 100 percent we are identifying outpatient. But even if you use the claims perfectly, as you know, transitions can take place mid-course.

Patients that stay overnight are not counted in the outcome. So, they're not counted as admit patients. We don't look at look at what happens until they're discharged home. So, there is that flexibility in the outcome that if you're just kept overnight, we're still calling you an outpatient and we're not saying that there was something that happened that wasn't supposed to happen.

DR. FLEISHER: So, 23 hour stays of outpatient?

DR. DRYE: It's, I mean one night because CMS uses -- most hospitals, if it's a second night, they'll have to bill as an inpatient.

DR. FLEISHER: That's billed on whose code on the surgeons, anesthesiologists or hospitals?

DR. DRYE: It's the -- Craig, you're on the phone, right? It's the surgeon, it's the operating surgeon that we looked at. Craig, I want you to confirm, please.

DR. PARZYNSKI: I'll jump in if you say anything incorrect.

DR. DRYE: Okay.

DR. FLEISHER: Other questions? Sorry, I -- excuse me.

DR. GUNNAR: Yes, I wanted to just -- so the observation bed rules have changed to just under a minute less than 48 hours.

So, and then there's also this intent

or unintended consequence potentially of, you know, laparoscopic cholecystectomy is an outpatient procedure now with the good feeling that you've got this two days of observation bed if you need it for that particular patient.

If this measure goes into endorsement, does that conflict with that otherwise usable mechanism for a safety net?

DR. DRYE: I think they're aligned and, honestly, we talked with the CMS payment people to make sure we weren't going to, you know, we weren't creating some strange misalignment with that.

Because, here, we really want unanticipated returns to the hospital or unanticipated, you know, admissions which the threshold, as you were saying, if you're really admitted for two days post one of these surgeries, that's really -- or more -- that's really unexpected.

So, here, you have your obs stay and you're not going to be counted in the outcome of

this measure. So, both things give you flexibility to -- I'm not sure that's what you're saying to keep the patient.

DR. GUNNAR: Well, there are actually interqual criteria around, you know, what is an acceptable use of an observation bed. And if the mere use of the observation bed is nonperformance or less than excellent performance, then I'm sure -- I envision there's a potential conflict in that.

DR. YATES: Well, one of the unintended consequences of the readmission penalty is that the use of observation status has gone through the roof. I mean it's gone up dramatically over the last, you know, few years enough that it, you know, the 48-hour -- 47-hour rule is being debated again.

DR. FLEISHER: Other comments?

DR. DRYE: I'm just not sure I'm understanding -- I'm communicating well because, in this measure, you would not be counted as in the outcome. It would not be an adverse outcome

if you stayed in observation status, for example, overnight.

So, we're only counting the obs if they go home and come back.

DR. GUNNAR: Correct. All I'm trying to say is that if you have excellent backup to your same-day surgery environment, there are cases that you would use that environment for and the patient would benefit from going home.

So, I guess what I'm asking is, is that could you subselect or think about subselecting outpatient procedures that really there's no expectation of an inpatient or an observation stay associated with that?

That's a different subselection than what's happening in the community or across the country which is, let's push our outpatient environment. If we've got good backup and we have an observation bed readily available and comprehensive inpatient care.

So, this is where I'm struggling.

DR. YATES: And I know where you're

coming from because there's people trying to do same-day 23 hour total hip replacements.

But along those -- the contrary -- the flip side of that is, is that those patients that you fully expect to go home that day in arthroscopy of the knee, for instance, if they have now -- I mean you may drive behavior to keeping those patients under obs if they're having a little bit of trouble peeing or they're having a little trouble with the food they're trying to put down. And you just have this low threshold of using obs beds which, you know, does have --

DR. GUNNAR: Which is good quality care.

DR. YATES: Right. Well, it just changes the behavior. But that's neither here nor there with validity. I'm following on your conversation.

But, you know, I think they showed -- in defense of the measure, the validity, I think, is really good when you look at their numbers.

DR. FLEISHER: So, let me ask something of Elizabeth because the colonoscopy measure, just to be aware, was actually identified as sort of pilot measure, but there was going to be further testing and that was actually just debated again in CSAC.

How come you've chosen to put this out as a fully formed measure or is there?

DR. DRYE: Yes, I mean these outpatient measures, this is where the big measure gaps are and they are new and they do need testing.

The colonoscopy measure, the reason that it makes sense to lead with that national testing that's going on right now in all facilities, ASCs and HOPDs will see their scores in July, is that actually it is going to help us learn about a lot of the things that you're talking about now, how observations define, you know, like how much variation there is, whether what we're seeing differs at all than what hospitals are seeing in their own data because

they'll all get their data.

So, we're sequencing a little, this is like lean management, right, we're sequencing so that we learn from the first one. But this one, I mean we tested it with 20 percent sample. We think it's really strong. We know policy around, you know, observation stay is going to be a little bit dynamic.

And you all are raising something that, you know, CMS is always saying and I think we need to see as much of this as possible that they will track for unintended consequences and follow it.

All these issues you're raising are really important to be looking at further. But we don't think it's not a reason to move forward at this point.

DR. FLEISHER: Okay, any other comments?

DR. GROVER: Well, just that the -- I think for the validation, too, the risk adjustment's critical, you know, that you've done

that.

DR. DRYE: We copied some other good people on that one.

DR. FLEISHER: Okay, let's vote validity.

MR. LYZENGA: We're voting on validity, just in case if anybody's on the phone still who didn't hear it.

Sixteen percent high, 84 percent moderate, zero low, zero insufficient.

So, the measure passes on validity. We'll go to feasibility.

DR. FLEISHER: Comments?

DR. CIMA: It's all coded in administrative data. It's hard to hide somebody showing backup in their emergency room.

But, how about -- what about -- it'll track someone going having hospital A and showing up in emergency room C somewhere else down the road, right?

DR. FLEISHER: Fred?

DR. GROVER: I think it's very

feasible for the reasons just mentioned.

DR. FLEISHER: Great. No other comments, please vote.

MR. LYZENGA: We have 84 percent high, 16 percent moderate, zero low, zero insufficient.

So, the measure passes feasibility. We'll go to usability and use.

DR. FLEISHER: Comments?

DR. GROVER: This is what we had -- the conversation we had previously.

DR. FLEISHER: Right. I would hope a lot of these comments and A.J. and Robert and others, particularly since I suspect by the time this is actually rolled out, they'll be up for reendorsement. So, these are going to be really critical to get into it.

DR. GROVER: I know it's late.

DR. FLEISHER: No, I meant -- I didn't mean it that way.

DR. GROVER: Okay.

DR. FLEISHER: I meant the fact that it's going to -- they have to come back in three

years. Well, you have to come back every year for endorsement so let's make sure a lot of those comments are in the final report because they're really critical for Yale to address.

So, I keep saying that, but I --

DR. GUNNAR: Then I won't say anything else.

DR. FLEISHER: No, it's really important work.

Okay, let's vote.

MR. LYZENGA: Voting on usability and use, high, moderate, low or insufficient information.

We have 33 percent high, 61 percent moderate, six percent low, zero insufficient.

So, the measure passes on usability and use. And we will take overall suitability for endorsement next.

Are there any comments before we vote? Does not appear so, let's go ahead and vote.

Ninety-five percent yes, five percent no.

The measure passes.

Thanks everybody, it's been a marathon session.

DR. FLEISHER: We need to open for public --

MR. LYZENGA: Oh, yes, public comment.

Operator, can you open the lines for public comment?

OPERATOR: Yes, sir. At this time, if you'd like to make a comment, please press star then the number one.

At this time, there are no public comments.

MR. LYZENGA: All right, thank you.

DR. FLEISHER: Thank you all. We will see you in 30 minutes for those joining us for dinner. And outstanding, in my perspective, comments today.

MR. LYZENGA: Thanks everybody, good work.

(Whereupon, the above-entitled matter went off the record at 5:55 p.m.)

F

I

J

M

N

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 Standing Committee

Before: NQF

Date: 03-19-15

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