NATIONAL QUALITY FORUM

+ + + + +

SURGERY STEERING COMMITTEE

+ + + + +

WEDNESDAY, MAY 28, 2014

The Steering 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.

PRESENT:
LEE FLEISHER, MD, Co-Chair
WILLIAM GUNNAR, MD, JD, Co-Chair
ANTHONY ASHER, MD, FAANS, FACS, Carolina Neurosurgery & Spine Associates
ROBERT CIMA, MD, MA, Mayo Clinic
RICHARD DUTTON, MD, MBA, Anesthesia Quality Institute
ELISABETH EREKSON, MD, MPH, Dartmouth Hitchcock Medical Center
FREDERICK GROVER, MD, University of Colorado School of Medicine
JOHN HANDY, MD, American College of Chest Physicians
MARK JARRETT, MD, MBA, North Shore-LIJ Health System*
CLIFFORD KO, MD, MS, MSHS, FACS, American College of Surgeons, UCLA School of Medicine
BARBARA LEVY, MD, FACOG, FACS, American College of Obstetricians and Gynecologists
BARRY MARKMAN, Aetna
KELSEY McCARTY, MS, MBA, Massachusetts General Hospital

LAWRENCE MOSS, MD, Nationwide Children=s Hospital
AMY MOYER, The Alliance
KEITH OLSEN, PharmD, FCCP, FCCM, University of Nebraska Medical Center, American Society of Health-System Pharmacists
COLLETTE PITZEN, RN, BSN, CPHQ, MN Community Measurement
LYNN REEDE, DNP, MBA, CRNA, American Association of Nurse Anesthetists
GARY ROTH, DO, FACOS, FCCM, FACS, MHA Keystone Center
CHRISTOPHER SAIGAL, MD, MPH, UCLA
ROBERT SARWIN, MD, MS, Seattle Children=s Hospital, Organization of Children=s Hospital Surgeons-in-Chief
ALLAN SIPERSTEIN, MD, Cleveland Clinic
AMBER SLICHTA, RN, BS, MS, Health Foundation for Western and Central New York
LARISSA TEMPLE, MD, Memorial Sloan-Kettering Cancer Center
A.J. YATES, MD, University of Pittsburgh Medical Center

NQF STAFF:
HELEN BURSTIN
ANN HAMMERSMITH
VY LUONG
ANDREW LYZENGA
AMARU SANCHEZ
REVA WINKLER

ALSO PRESENT:
MAUREEN AMOS
DEBORAH DEITZ*
ROGER DMOCHOWSKI
JEFF JACOBS
WANDA JACKSON
TONI KAYE
DAN MORGAN*
SEAN O=BRIEN
SUZANNE POPE

MATTHEW POPOVICH
SAMANTHA PULLIAM
SAM TIERNEY
* present by teleconference

TABLE OF CONTENTS

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

8:30 a.m.

MR. LYZENGA: All right. Welcome, everybody, to the in-person meeting of the Surgery Steering Committee. I'm Andrew Lyzenga. I'm the senior project manager on this project.

I'll just run through a few quick housekeeping items here before we start. So just to start out, again, a few housekeeping things. Restrooms are right out the door this way and to the right once you pass by the desk here. We'll have a few breaks during the day. Depending on how the measure reviews go, sometimes we end up having to do a little bit of work through lunch, but we'll see how things go.

Everybody should be able to log into the wi-fi network. The user name and password is right up there. If you haven't been able to access it, let us know and we'll help to

troubleshoot that. We would ask that you do mute your cell phones during the meeting so you don't interrupt the discussion.

The project staff here is Wunmi, who is out of town right now. Amaru Sanchez is our project analyst. I'm Andrew Lyzenga. And Dr. Reva Winkler is our senior director.

Let's see. And we also have our general counsel here, Ann Hammersmith. She's going to say a few words about disclosure of interest, and we'll actually walk around the room and introduce ourselves, and we'll ask Ann to explain, say a few words about disclosing any interests or potential conflicts you have. Ann?

MS. HAMMERSMITH: Thanks, Andrew. As Andrew said, we're going to go around the room and combine introductions with disclosures of interest. Those of you who have been on our committees before are very familiar with this process. I'm going to just say a few

words, give you a few reminders, and then we'll go around the room.

First, I want to remind you that you sit as individuals on this committee. You don't represent your employer. You don't represent anyone who may have nominated you for service on this committee.

Also, I want to remind you that, for our purposes, because of the unique nature of the work that we do, you may need to disclose things that are not financial. People often say I have no financial conflict of interest, which is great. But you can have something else that we would look for you to disclose. For example, if you served on a committee where the work of the committee was relevant to what you'll be doing here, we would look for you to disclose that.

Just because you disclose does not mean that you have a conflict of interest. Part of the idea of this exercise is to be as

open and transparent as possible with each other and with the public. So just because you reveal something, it doesn't mean you're conflicted.

We do have some committee members who are conflicted for particular measures. I think each of you have a short memo with a chart that indicates Reva's show-and-tell that indicates what the measures are and who is conflicted. So when we go around the table, I'm not looking for you to recite right now what measures you have to step away from. You should do that at the time the measure comes up because we want that in the record that you have recused yourself, and staff will prompt you or give you any advice that you might need on that.

We are particularly interested in your disclosure of grants, research, or consulting, but only if it is relevant to the work of the committee today and tomorrow. So in other words, please don't recount your

resume. But if you have some activity that you think is relevant, please do disclose it.

Then, lastly, we don't want you to sit there in silence if you think that there's a conflict. If you think you may have a conflict, if you think that a fellow committee member may have a conflict, if you think that someone is acting in a biased manner and is not being a good committee member, we want you to speak up. A conflict of interest regimen only works if other people involved do their part, so we're really relying on you, as committee members, to speak up if you think there's a conflict or if you think something untoward is going on.

You can do that by raising it openly in the meeting. You can go to your co-chairs, who will go to NQF staff, or you can go to NQF staff. But we ask you to do it in realtime, not wait two weeks and say, well, you know, something seemed kind of odd to me or I think

that I may have had a conflict.

So with that, let's go around the table. Tell us who you are, who you're with, and if you have anything that you want to disclose. And we'll start with the chairs. I always make them go first.

CO-CHAIR FLEISHER: I'm Lee Fleisher.

MR. LYZENGA: And just to note, you know, when everybody is speaking, please hit the microphone when you speak so that our transcriber can catch what you're saying.

CO-CHAIR FLEISHER: So welcome and thank you all for joining us. I'm Lee Fleisher. I'm co-chair. I'm a professor and chair of anesthesiology at the University of Pennsylvania. My potential conflicts of interest is that I'm a member of the Committee for Performance and Outcome Measures at the American Society of Anesthesia that does develop some of these measures, but no other

leadership role within the ASA.

I do have a grant on risk-adjusting process measures from AHRQ, on which I'm a co-investigator with Jeffrey Silber for that work. I'm also a member of the CSAC, the Consensus Standards Advisory Committee of the NQF.

CO-CHAIR GUNNAR: I'm Bill Gunnar, National Director of Surgery for the Department of Veterans Affairs. I live in town, so I have great sympathy for any of you who had to fly in after 6:00 last night. Good to see you all here and nice to meet you and be a part of this.

I don't have any disclosures, conflicts of interest that I am aware of. And my bias is probably to the fact that I'm a cardiac surgeon and leave it at that. So, next.

MEMBER MOYER: I'm Amy Moyer. I'm the Manager of Value Measurement with the Alliance, and we're, in a nutshell, a

not-for-profit healthcare purchasing cooperative.

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

MEMBER SAIGAL: Chris Saigal. I'm a urologist at UCLA. I sit on several AUA committees, including the Quality Improvement and Patient Safety Committee, and I'm a co-founder of a company called WiserCare, which I don't think is relevant to this.

MEMBER PITZEN: Collette Pitzen. I'm a measure developer with Minnesota Community Measurement and nurse by background with a quality improvement and reporting and measurement design background. I have no conflicts to declare. Although being a measure developer, we do not have any measures in the general surgery area. Thank you.

MEMBER MCCARTY: Hi, my name is Kelsey McCarty. I'm the Quality and Safety

Program Manager for the Department of Anesthesia at Massachusetts General Hospital. I am a member of the American Society of Anesthesiologists, but I do not have any involvement in the measures submitted for today and I'm not a member of any committees within that organization. No other conflicts.

MEMBER MOSS: Hi, I'm Larry Moss. I'm a pediatric surgeon at Nationwide Children's Hospital in Ohio State in Columbus. I'm on the steering committee for the pediatric NSQIP and the Measures and Standards Committee for Children's Hospital Association.

MEMBER SAWIN: I'm Bob Sawin from Children's Hospital in Seattle and surgeon-in-chief, and I'm representing the Organization of Children's Hospital Surgeon-in-Chiefs, and I have no conflicts.

MS. HAMMERSMITH: Just a general reminder, you sit as an individual, so you're not representing an organization. Thank you.

MEMBER YATES: I'm A.J. Yates, Jr. I'm from UPMC in Pittsburgh. I serve as the chairman of the Evidence-Based Medicine Committee for the American Association of Hip and Knee Surgeons and have and actively sit on Appropriate Use Criteria Work and Clinical Practice Guidelines for the American Academy of Orthopedic Surgeons, but none of these have been involved in performance measures that have been submitted to the NQF.

In the federal sphere, I serve on a technical expert panel for physiciancompare.gov, and I also serve on the technical expert panel for the Yale CORE group and CMS in terms of the measure for cost of the total hip and total knee replacement and also serve on MEDCAC and the FDA.

MEMBER REEDE: Lynn Reede. I'm a certified registered nurse anesthetist. I work for the American Association of Nurse Anesthetists as Director of Practice. I have

no conflicts.

MEMBER DUTTON: I'm Rick Dutton. I'm an anesthesiologist. I work clinically at the University of Chicago. I am the Director of the Anesthesia Quality Institute, which is ASA's national anesthesia registry program. And I'm Chief Quality Officer of the ASA and involved in multiple ASA committees, including the Performance and Outcomes Measure.

MEMBER LEVY: I'm Barbara Levy. I'm an obstetrician/gynecologist and Vice President for Health Policy at the American College of OB/GYN. I also serve on the PCPI Executive Committee, and I chair the AMA RBRVS Update Committee.

MEMBER OLSEN: Yes, I'm Keith Olsen, professor and chair of pharmacy practice, University of Nebraska Medical Center. Probably not conflicts of interest, but I am a member of the Board of Regents for the American College of Critical Care Medicine

and head the Guidelines Committee for that organization and also hold funding from NIH.

MEMBER TEMPLE: I'm Larissa Temple. I'm a colorectal surgeon at Memorial Sloan-Kettering Cancer Center. I'm an associate professor at Cornell. I sit on, I co-chair our Quality and Safety Committee, but I have no conflicts and no measures being submitted.

MEMBER GROVER: I'm Fred Grover. I'm past chair of the Department of Surgery at the University of Colorado. Currently, I'm on the faculty there in cardiothoracic surgery. I'm the past president of the STS, and I think I get the prize today for the most conflicts. I'll be lucky if I can say much, but they're listed here.

I am listed as a SCIP committee member, too, but none of these measures came up. It was all a technical infectious disease panel, and I don't have any conflicts there.

Thank you.

MEMBER ROTH: I'm Gary Roth. I'm a cardiothoracic surgeon at a community hospital in Lansing, Michigan. And I'm also the medical director for the Michigan Health Hospital Association Keystone Center for Quality and Safety.

I don't have any conflicts. Just as far as disclosure, I'm actively involved in the Quality Committee at the Michigan Society of Thoracic and Cardiovascular Surgery, and much of the work that we have been doing has been instrumental in many of the STS initiatives and measures.

MEMBER SIPERSTEIN: Allan Siperstein. I chair the Endocrine Surgery Department at the Cleveland Clinic and also serve as the NSQIP physician champion for the institution.

MEMBER CIMA: I'm Bob Cima. I'm a colorectal surgeon at Mayo Clinic, professor of

surgery. I also am the vice chair for the surgical practice at Mayo Clinic, and I lead most or sit on the committees that respond to most of these measures, as opposed to develop them. So if you're a responder, I don't know if that's a conflict. But other than that, no conflict.

MEMBER KO: Good morning. My name is Clifford Ko. I'm a professor of surgery at UCLA. I'm a colorectal surgeon. I also work at the American College of Surgeons, and I'm the Director of the Division of Research and Optimal Patient Care that houses all the quality programs at the college, the trauma, the cancer, bariatric program, and I'm also the Director of NSQIP.

There are some disclosures or conflicts with a few of the measures where we have a bariatric program that we partner with the Society of Bariatric Surgery, and they submitted a few measures. So that's one area.

Another potential area is that I was told by our D.C. staff that we had taken on some of the upkeep of some of the perioperative measures from the PCPI. I'm not exactly sure which ones they are, but she said they're going to be discussed later this afternoon. So maybe during the break, I can verify which ones those are. Thank you.

MEMBER MARKMAN: Good morning. My name is Barry Markman. I'm a retired plastic surgeon. I currently work for Aetna Medicaid in their division. I don't believe there's any conflicts, but I do chair and participate in many of the surgical quality programs for Aetna. And I believe I don't have any conflicts at this point.

MEMBER ASHER: I'm Tony Asher. I'm a professor of neurological surgery at University of North Carolina at Chapel Hill and practice in Charlotte, North Carolina. I don't have any obvious conflicts. As a matter

of disclosure, I am the director of most of organized neuro surgeries national quality programs, including our large registry program.

MS. HAMMERSMITH: Okay. Thank you. And I understand we have one committee member on the phone, Mark Jarrett.

MEMBER JARRETT: Yes. I'm on the phone, and I apologize to everybody I couldn't be there but we have a pleasant visit from Joint Commission, so I have to be here back in New York. I'm the Chief Quality Officer of North Shore-LIJ Health System. I am not a surgeon. Actually, I'm a rheumatologist by trade, but I live in the quality world with all my surgical compatriots. And I have no conflicts. I do serve on a PCPI committee, but it does not actually do any measures. And as well, I sit on the Musculoskeletal Committee for NQF. But, again, no conflict. MS. HAMMERSMITH: Okay. Thank you, everyone, for

those disclosures. Based on the disclosures this morning, do you have anything that you want to raise, anything you want to discuss with each other, or any questions of me or the staff?

(No response.)

MS. HAMMERSMITH: Okay, thank you.

MR. LYZENGA: All right. Thanks again, everyone. And welcome again. It's nice to see everybody and put faces to the names and voices, at least for most of you. I'm just going to say a few words about just a list of the standing committee members. You also should have a list of the, a roster in front of you and printed out on your desk.

I'll just say a few words about sort of the role and process of the standing committee. If any of you have previously served on NQF committees, in the past we've reseated a new committee each time we've done a project, done a new call for nominations and so forth. We've tried to transition to more of

a standing committee system for a number of reasons, among those just sort of gaining some efficiencies in terms of project startup, also trying to gain some consistency in decision-making across time in a particular topic area, but, as well, to try to allow you, as committee members, to sort of take a little bit more ownership of the portfolio of surgery measures and manage that portfolio over time. And Reva will say a few more words about that in a few minutes.

As a standing committee member, as Ann Hammersmith just mentioned, you're acting as an individual representative for the NQF multi-stakeholder membership. You'll be serving either a two-year term or a three-year term, and, at some point during this meeting, we're actually going to go around and have you draw from a little cup or something and see whether you're going to have a two- or three-year term. That will be done randomly.

You'll be working with NQF staff to work through the project. We'll be drafting up a report after this meeting to summarize your decisions and sort of give, you know, an introduction and background on the project and everything.

We're expecting you to review all of the measures that are submitted under this project and to evaluate each of those measures against each of the evaluation criteria. I think you have in front of you, we've printed out a couple of tools to help you with that: the algorithm, this nice colored thing here, as well as a script of sorts for the lead discussants which sort of structures the discussion a little bit.

In terms of process, we'll be first discussing evidence, and this goes for each of the criteria. We'll have some discussion around the evidence topic, and then we'll vote on the evidence criterion, and then we'll move

on to scientific acceptability and vote on that criterion and so forth.

You'll be making recommendations through this meeting and some follow-up work. The report that we summarize, as staff, and we'll send you a draft of that report to get some input from you and everything. And that will be posted for comment, sorry, public comment period for 30 days, and then will be released for an NQF member vote, and will go to the CSAC, the Consensus Standards Advisory Committee.

Again, as a standing committee, you'll also be sort of managing and overseeing the portfolio of surgery measures over time, which, again, Reva will say a little bit about in a few minutes.

We have 29 measures under review today and tomorrow. You can't really read that probably, but you have an agenda, as well, and have those measures listed.

And with that, I will turn it over

to Reva to talk a little bit about the portfolio and give you sort of a sense of the scope and breadth of that portfolio and characteristics of it. Reva?

MS. WINKLER: Great. Thanks, Andrew. Before we launch into the portfolio, I'd like to introduce the Senior Vice President for Performance Measures here at NQF, and that's Helen Burstin. She wants to say a few words.

DR. BURSTIN: Good morning, everybody. Thank you so much for coming. Lots of familiar faces. Thank you for coming back. I guess that's a positive sign for those of you returning, and thanks to those of you who are willing to join us in this journey. It's actually exciting to have standing committees. I think you're now the fourth or fifth I think we've convened, and it actually has made a big difference. I think there is sort of a sense of your ownership over the portfolio. As Reva

launches into this, you're obviously not going to make decisions right now about what should be in or out or kind of get a sense of the gaps. But it really does lend itself towards the idea that, over time, you're the steward of all measures related to surgery, perioperative care, surgical care, really making sure that we try to bring in the measures that actually matter, trying to increasingly remove measures that are no longer adding value.

To be perfectly honest, it's really important that we recognize there are costs of measurement beyond just the burden of collecting data but also the opportunity costs of people continuously focusing on measures that perhaps we should declare success and move on to some harder measures perhaps. So we really will look towards you for that role, and we've really found this to be important.

And I would also like you to keep an eye as you're going through this to say I'm

evaluating this measure but not so much during this discussion but kind of keep in mind where there are clear gaps in the surgical field of where measures should be, where they aren't.

Part of what we'd also like to have you do as standing committee members is actually help us identify where there may be measures in use in a health system or in a registry or something along those lines where we should actually be prospecting, going out and trying to pull those measures in. It's oftentimes not very satisfying to sort of sit passively waiting for measures and then bringing them to committees and then sometimes having committees go, some of these are great but some of these don't really meet the bar. So I think we're really going to want to enlist you in that effort.

So with that, I'll have Reva lead you through the portfolio. As you can see on your desk, it's quite substantial in surgery.

Thanks.

MS. WINKLER: I'd ask each of you to take a look at one of the documents on your desk that is the portfolio of NQF-endorsed measures related to surgery. This is one of NQF's largest portfolios. And when we look across the entire portfolio of NQF measures, which numbers somewhere around 700 right now, a substantial number, around 130 of them, do relate to surgery in some way or another. And so we really need to be aware of all of the plethora of measures and perhaps have some thoughts about, you know, do we have too many measures, not the right measures, do we have the right combinations of measures.

And so I think one of the advantages of standing committees taking ownership of the portfolio is looking at it from this perspective. In the past, steering committees really didn't do that.

So we've got a large number of

measures related to surgery. Now, not all of these measures are assigned to this committee for ownership. But, nonetheless, you need to be aware of measures that exist within NQF's portfolio that may be assigned to other committees. And they're assigned to other committees for a variety of reasons, and I will fully admit that, at times, they may be arbitrary. So we just had to assign them some place, and we picked one.

So as you look through this document, we've tried to organize the measures in some sort of logical fashion. I am completely open to any suggestions on if there's a better way to organize these measures. Please, help us out. I'm definitely open to that.

You're going to find that the measures that you're evaluating today, there are nine new ones, new submissions for 20 maintenance measures that have been endorsed by

NQF, some of them for quite a while now, some of our earliest measures that have been endorsed for almost a decade, as well as measures that were initially endorsed maybe four or five years ago and are a little overdue for review.

Just realize that in this list, those that have the asterisk with the number are not part of the actual surgery portfolio that you are overseeing but do belong to other topic area committees but, nonetheless, relate to the work that's done in surgery. And you should be aware that they exist. We've had many conversations with committees that say we really, you really should have a measure about X, and, in fact, we do, but it's often shepherded by a different committee.

So this is an attempt to help you understand the breadth of NQF's portfolio of endorsed measures related to surgery. So you can see that many of the adverse outcome

measures are in our safety project. Eye surgery measures are combined with other measures for eye care professionals in another portfolio. Similarly, oncology are grouped together, care coordination, and some perioperative stress testing is in cardiovascular.

So, again, as I said, some of these assignments are arbitrary. But it is going to be helpful in your overview of the measures you are responsible for to understand the other measures that are particularly pertinent to the area of surgery.

Next slide, please. So the way that we've organized it, and, as I said, it was somewhat arbitrary but just trying to get, you know, a handle on the beast, was to look at the groups of measures that are, one, on the more general side that I've termed perioperative care. There are 12 measures. VTE prophylaxis is a big subset of perioperative care, so I

pulled that out. Similarly with antibiotic prophylaxis, a big group of measures. Care coordination. Adverse outcomes, yes, lots of those. And measures specific to ambulatory surgery centers.

So they aren't operation-specific or surgical procedure-specific. They tend to be more broad spectrum.

However, we do have a substantial number of measures that are applicable to different types of procedures and different types of surgical sub-specialties. And so you'll see that we've got, you know, measures for abdominal and colorectal surgery. We have three new measures for bariatric surgery. We do have breast surgery. Cardiac surgery, a large number, one of our biggest subgroups. Eye surgery, GYN and GU surgery, orthopedic surgery, pediatric surgery, thoracic surgery, and vascular surgery. So, you know, the portfolio really does have a lot of breadth and

depth in measures for surgery.

And another way of slicing and dicing the measures is by type of measures. This is one of the portfolios that has a large number of outcome measures. Surgery has generally been a leader in submitting and having NQF endorse outcome measures. There are also process measures, but, again, we've got a large number of outcome measures. There are a few efficiency measures, a couple composite measures, and measures related to cost and resource use.

So, again, also we can slice it and dice it another way and look at it by care setting. And so we really can see measures that are applicable in a wide variety of care settings where surgical care intersects all of those care settings.

And then the so what of measures is a question frequently asked is how are these measures used, and our best assessment of use

is actually with our federal partners, understanding the measures that are used in federal programs. And so we do have, a goodly number of them are used in federal programs. We do, however, know that there are many of these measures that are used in private programs and other public reporting entities outside the federal government.

So this is the portfolio, this is our portfolio of measures. And, again, we're going to be giving you sort of the oversight responsibility of this portfolio. While we're not in a position today, as you're just getting started, to really, you know, grapple with what's in or what's out and what we've got and what we don't have, please keep in mind as we go through the work not only today but going forward how everything we discuss fits into this large portfolio.

When people are searching for measures to use for their various programs,

they often come to NQF and search through our database looking for, you know, appropriate measures to meet their needs. These measures are tagged as associated related to surgery. So sometimes, you know, they may get confused if there are multiple measures that seem to be doing the same things, like what do I do with this, how do I sort this through?

So, again, I think, as good stewards of the portfolio, we're going to be asking you to continually think about the measures that are in the portfolio. And as that portfolio evolves over time to meet the needs in the marketplace to continuously drive quality improvement, it will be necessary to take in new measures and retire old measures. And that is just a natural life span of measures in the portfolio. And so we're really looking to you to help us be sure that NQF's portfolio of measures really reflects the most usable and important measures related to surgery.

So we have a couple of minutes if anybody had any questions or comments or reaction about the portfolio. Go ahead.

MEMBER YATES: Just looking at that slide, the fourth categorization that might be useful to people who are looking for help or for gaps in quality measurements is where does the data come from, and it would be great if those categories were, if you had a list of how many of your measures are based on administrative data sets and how many are based on registry data sets and how many are based on some alternative form of data collection because it would give you a snapshot of where the universe of data is coming from, given the fact that there are discrepancies between the administrative data sets from CMS and registry data, which may be more carefully, not so much carefully but it's gathered in a different way. And I think that would be a good way to categorize your portfolio.

MR. LYZENGA: And we do have that information. We can actually put together a document with that, you know, with the data source for each of the measures, as well as if you're interested in which measures are in which federal programs and other kinds of information.

MEMBER YATES: It would be good to have a snapshot on this slide showing what the cumulative effect is. And you can then track that over time because it's my suspicion, as we move forward with the impact of performance measures not being limited to simple public reporting, that you're going to see an enriched environment of registry data that's going to be important to track and see what trends in that regard.

MS. WINKLER: Thank you.

MEMBER DUTTON: Thank you. It seems from the instructions like the purpose of the NQF and the Committee is to produce the best

possible measures at the cutting-edge of science, discriminatory, relevant. I want to make sure I understand that that is the specific mission because I think we're going to hear from a lot of the stewards push to keep topped-out measures certified, and that's because of the obvious economic need of all the societies, all the professionals to have nine performance measures in order to avoid payment hits going forward.

So is that something we're not supposed to consider? So all eight topped-out antibiotic measures can just go away right now, or how should we process that?

DR. BURSTIN: The expected question, of course. May as well get to it early. No, it's absolutely the right question, Rick. And one of the things we've talked a lot about in the last year or so is this question of whether NQF should move away from this idea of binary endorsement of yes/no and

really endorse more fit for purpose, so it's intended for QIs, or is there a different bar, for example.

To date, we don't have that. So for right now, we need to have you act on what is in the criteria. And one of the must pass criteria we do have is that there is a performance gap or a variation across providers. Now, that can change, depending on level of analysis, which I think is something to consider, as well, provider versus, you know, hospital versus clinician, for example.

We do have a category called reserve status. It's really intended to be an exception, and the idea here is that measurement science doesn't yet give us great confidence that, when you take your eye off of a particular measurement, over time will we start to see a, you know, declining performance without sort of the laser-like attention to it.

I think it was put there

intentionally, initially, several years ago because of exactly these concerns. I think the world has changed a bit in that I think we're already seeing, for example, CMS and Medicare, in fact, pulling measures that are topped out, recognizing, again, this issue that it's not just you're measuring them but there's actually people chasing patients around with clipboards to make sure every single one of those patients gets in to meet a financial penalty.

So I think that you can continue to use reserve status. The idea would be only those measures that you think are exceptional measures, they meet every criteria, except they are topped out. And the idea there would be that they would be in this reserve status not intended to be used as measures of first choice but more so measures in reserve that can be pulled up as needed to see if there's, in fact, evidence or sort of periodic surveillance to make sure we haven't seen decrease in

performance.

Frankly, I think it's something, and Lee sits on the CSAC, we're going to potentially go back to CSAC this coming summer and see if, in fact, it's outlived its usefulness. We've not seen people use the measures in reserve status. I think people are increasingly getting comfortable that some measures that are topped out are topped out because they've been completely built into systems such that it's hard to imagine -- for example, a measure we recently, I think, did not recommend for continued endorsement in the cardiovascular project was aspirin in emergency departments for chest pain. I mean, it's hard to imagine walking into an ED right now and not having somebody hand you an aspirin almost irregardless, I think, of your chief complaint I would fear.

I think part of your thinking should also be, as you're looking at some of those

measures, is some of that high performance reflecting the fact that it's something that's in, you know, such a laser-like focus in terms of pay-for-performance or pay-for-reporting programs, or is it really that, from your clinical perspective and your perspective as end-users of these measures, they've just been so built into the process of care that measuring them has really outlived its usefulness?

So I think we're really going to look to you to offer us that guidance. But I think, you know, our sense would be reserve status should be something you use as an exception. And I think, over time, as we explore this bigger question of whether some of those measures might just be put in QI buckets, but even people in the QI field would not particularly want to continue to use measures that are also topped out. So we'd welcome your thoughts on this because it's going to come up, obviously, quite soon.

MS. WINKLER: I'm going to add on to that and to address Rick's question a little bit more focused for today's work. And what we're asking you to do is use the NQF criteria. That is the common language that we use. Measure developers are aware of it. We use it with all of the committees. So that really is what standardizes the work that we do in assessing measures.

So we really are looking to you to apply those criteria. And you will find, you know, situations where evidence is maybe not as good as everybody assumes it to be, or we're talking about no opportunity for further improvement. And we really are asking you to use the criteria because that is the common platform for everyone that's working in this space: developers, end users, you know, the NQF endorsement process, and all the various steps through it. That's what we're using, so we'll ask you to look to that, as well.

CO-CHAIR GUNNAR: Go ahead.

CO-CHAIR FLEISHER: So, Reva, just to be clear, if we actually start having this debate around some of the antibiotic measures and it's performance gap that's the major issue, how would you like us to send the signal? I mean, some people could vote no, some people could vote yes, and it may all be around that one question.

And the second question following that up is also around harmonizing measures in which we see two similar measures.

MS. WINKLER: Right. A couple of things. Essentially, it will be a little bit easier when we have one in front of us to walk through, but I will ask you to stick to the criteria and vote with the criteria. The criteria and the way we evaluate them do provide you a couple of opportunities to make exception when you feel strongly about something.

Also, clearly, you know, if you

determine that there is no opportunity for improvement, we can ask the follow-up question, okay, is this measure one that might be suitable for reserve status? If so, we need to do the complete and full evaluation through all the criteria, and it has to meet it very, very well. That's the criteria. So we'll do it as we go through.

Harmonization is another important thing. We've been having these conversations with all the committees and certainly with all the developers for many years now. This was a major topic of conversation when the surgery measures were reviewed three years ago. So this is not news.

So harmonization means alignment of the specifications for similar and related measures to reduce the burden and the chaos and the cacophony of measurement out there for the end-users. So this is not a new thing. And if harmonization really hasn't happened yet,

perhaps it's time that we need to signal that that's, that more attention needs to be paid to that.

So it's an important part of the criteria that I expect that we need to discuss. It tends to be the last one after we look through the whole measure because if the measure is not going to meet the criteria and not be recommended, we don't have to worry about harmonization. It's just a moot question.

So it is the fifth criteria we will be addressing on the measures. And it will be an important one for the measures on the table today. Go ahead.

CO-CHAIR FLEISHER: The easiest way to do it is put your name tag up to signal you have a question.

MEMBER ASHER: Thank you. In a related issue, I noticed in some of our conversations that we're looking at harmonization, it's normally thought of in the

context of the existing measures. But in situations where a measure is being reviewed and one in particular that was being applied to a certain, in this situation, registry effort, I immediately thought was if the intent was to encourage participation in that type of format. Is it the case that we should encourage a broader application of a particular measure, and does that achieve some of the same objectives?

MS. WINKLER: Sure. Remember that, in terms of recommendations from the committee, we don't own the measure. Someone else does. It's simply guidance and recommendations you can make that you would like to see them consider in future iterations of the measure. But today we are asking you to evaluate the measures as submitted, as written on the submission forms given to you.

Your discussion, naturally, always goes into, you know, questions of why didn't you

do this or why not do this? Great. That's advisory to the measure developers. They certainly can take that under advisement and hopefully perhaps, in their next iteration of the measure, you know, respond to it. But there's no guarantee of that. So we are asking you to look at the measures as submitted with the information they've provided today.

Okay. Any other questions? Again, I hope, over time, it's a little too much work to do today, but, particularly since this is a standing committee, over time, reflecting on the kinds of measures that exist related to surgery, that's your world. And these are the measures that are, you know, used to evaluate how well you do your job.

And so I really would encourage you to think in a way that we've never asked steering committees to think before, and that's think big and broad and ask how do we, as nation and as a specialty of medicine, evaluate how

good a job we're doing on behalf of patients? So, hopefully, we'll have the opportunity, as this group meets in future venues and future activities, we can have this conversation. But, again, we're just starting out, so, please, when you have a few minutes, give it some thought.

MEMBER REEDE: Thanks, Reva. Yes, you just said medicine, and I imagine -- well, thank you. I just want to clarify we are looking at this from the healthcare, the whole team, the whole care of the patient. Thank you.

MS. WINKLER: Yes. This committee is deliberately, as are all NQF committees and the NQF membership, deliberately brings together multi-stakeholders and people from all different aspects of the care delivery team but also the end-users of measures, the people who are involved in, you know, patient care, families, patients, purchasers, I mean the

whole world. That is always NQF's perspective, so I did not mean to limit it at all.

MEMBER TEMPLE: Reva, one of the questions I have, and I know it specifically comes up with the GU measures, GU and GYN measures, about some measures that were developed in pilot programs and decisions were made in pilot programs not just on importance but also on, you know, splitting measures versus combining measures. And I would think that, as we review, that we would want to sort of follow the recommendations of those previous committees, and I'm sure it's going to come up more than just today.

MS. WINKLER: Yes, I was planning on talking about that beforehand, and we're going to look a lot to Chris Saigal because he was the co-chair of that effort and provides the continuity between those two to help us do that. Chris?

DR. BURSTIN: One more comment I would make is I think the other thing you'll see and Reva did point out how many outcomes are actually in this portfolio, which is great. But NQF does have a hierarchical preference for outcomes, as you've seen as part of our evidence discussion. So I think that one of the things that often has also come up in addition to topped-out measures is also whether, once you have an outcome in place, do you continue to need the processes to be endorsed and measured, or is that something potentially more within the system to be keeping an eye on?

And in particular, I know structural measures will come up today, as well. And, again, I'll just point out the NQF board, at times, when we've looked at some of the structural measures in particular around participation and registries, have really indicated that, once you have the outcome measures from those registries, the

participation measures really may, again, have outlived their usefulness. So just to put that context in there for you, as well. Thanks.

MS. WINKLER: Anything else? All right. Well, thank you very much. Back to Andrew, just last few words before we get started.

MR. LYZENGA: Yes, just a last few words on just some ground rules. I don't think we need to harp on this too much. We can see the sort of ground rules we've laid out here. But I think one thing we'd really like to stress and reiterate is to base your discussion, to the extent you can, on the evaluation criteria, really ground your comments and decisions in those criteria, and to stick with the criteria under discussion at that particular moment.

Again, we're going to kind of walk through them stepwise, importance and scientific acceptability, feasibility, usability, and so on. And we'd really like to

keep the discussion focused around the particular criterion that we're discussing at that moment instead of moving on to, you know, or sort of wandering around to different aspects of the measure. During the discussion, we'll try to sort of walk through it in a systematic way. I think that helps to sort of keep the discussion focused and help us vote on what we've been talking about most recently. And with that, I think --

MS. WINKLER: Andrew, can I add one thing?

MR. LYZENGA: Yes.

MS. WINKLER: Just each of these measures is owned by some organization, and the measure steward is part of this evaluation process. As you notice, we have two reserved spots. If the measure steward developer is here with us, they will be joining the group at the table to discuss the measures. They're there to introduce the measure very, very

briefly and then respond to any questions they are able to, you know, offer their comments. They have the option of, you know, putting their card up to make a comment, just like everybody else around the table.

Again, we do realize that measure developers are really the important foundation of measurement. Without them, we would have nothing to work with. So this really is a partnership, that we really need to work with them collegially and effectively. So we do want to have good interactive conversations and meaningful conversations, feedback. They certainly have some of the most detailed knowledge of their measures to help you understand what's going on better, and your feedback can be given directly to them in terms of issues around measures going forward.

So this is an important part of this process. Again, open, transparent dialogue among all of the people involved. So I just

wanted to point that out.

And so as we'll get ready for our first measure, we've got our measure developers here.

MR. LYZENGA: Yes. And so, again, on that note, in terms of the process today, for each of these measures, we'll ask the developers to come up and just give a brief introduction of their measure. I think we've asked them to limit their remarks to around five minutes or so. And then we'll hand it over to the lead discussant. Each of you has been assigned, I believe, a couple of measures, and we'll ask the lead discussant to sort of lead the conversation on the measure, walk through each of the criteria based on this script that we've given you.

So I think we can go ahead and get started at this point. Our first measure, I believe, is number 0129. This is the Society of Thoracic Surgeons. Do we have the

developers in the room? Yes, and we would ask the developers to introduce yourselves and to actually -- and this goes for each of the committee members -- if you can, just note your name before your remarks again so that our transcriber can sort of catch who's saying what.

DR. JACOBS: Well, good morning, everybody. My name is Jeff Jacobs, and I'm a cardiac surgeon, I'm a professor of cardiac surgery at Johns Hopkins, and I am a pediatric cardiac surgeon at All Children's Hospital in St. Petersburg, Florida.

I serve several leadership roles in the Society of Thoracic Surgeons, including chairing the Access and Publications Committee for the database, as well as the Public Reporting Committee for the database. And I'll be presenting several measures over the next two days. And with me is Sean O'Brien.

MR. O'BRIEN: Good morning,

everyone. My name is Sean O'Brien. I'm a statistician at Duke University Medical Center. We serve as the data analytics center for the STS databases, so my role has been involved in designing some of the feedback reporting methods and the NQF measure submissions for the STS database.

DR. JACOBS: So should I move on and begin the introduction of the first measure? So the first measure we're discussing this morning is risk-adjusted postoperative prolonged intubation or ventilation. And this measure assesses the percentage of patients over the age of 18 who undergo isolated CABG and require intubation for more than 24 hours postoperatively.

Isolated CABG is the most common cardiothoracic operation that is done, and prolonged postoperative ventilation is a source of substantial morbidity after isolated CABG, and that provides the rationale for the

creation of this measure.

The measure is also part of our multi-domain composite for isolated coronary artery bypass grafting surgery, and we'll be discussing some of our composites later during these two days. The measure is used both for quality improvement initiatives within STS and at individual programs and also is publicly reported as part of the multi-domain composite.

During the phone call where we discussed this measure, one of the questions that was asked of us was, is there a possibility that there's an unintended consequence of patients getting extubated early in order to comply with the measure and then being re-intubated because they were extubated earlier than they should have been. So we actually went back to the database and pulled some data to examine that particular question, and I think that there's no evidence that shows that that unintended consequence exists. And

I can base that on the fact that we looked at three consecutive years of data ranging from 2010 to 2013, and patients who were extubated within 24 hours had a less than one-percent chance of being re-intubated across the board, while patients who were extubated after 24 hours, those that actually had prolonged intubation, had a rate of re-intubation of about 30 percent.

So I think that shows that those that are getting extubated early really very rarely get re-intubated. And although that's a theoretical unintended consequence, there's no data that show it actually happens. And that was the major question that was asked of us when we discussed the measure on the phone conference.

So with that, I think I'm happy to answer any questions as the dialogue evolves. Thank you.

MR. LYZENGA: Thank you. And I

believe we have Dr. Yates as the primary discussant on this one.

MEMBER YATES: This is, in fact, an outcomes measure. It's a maintenance measure first endorsed in 2007 and re-endorsed in 2011 as an outcomes. It's reasonably categorized as such in that delayed extubation, in and of itself, is an event for patient and family that has meaning in terms of how they see their health. And as an extension of that, it's a reasonable surrogate for the overall outcome in that there's evidence that delay to extubation greater than 24 hours is associated with less good outcomes overall for the patient undergoing a CABG.

As such, I would say that the evidence does not need to be further reviewed since it's not a process measure, but the evidence that exists to connect it to other outcomes is very strong.

MR. LYZENGA: Great. Thanks.

Any other comments from the Committee? Just a reminder, we're discussing evidence at the moment, which is, for an outcome measure, whether there is a rationale connecting the outcome where at least one process, healthcare process, intervention or structure that can influence the outcome in question.

MEMBER PITZEN: Collette Pitzen. I just have a process type question. I'm just curious or wondering if this is perhaps not an intermediate outcome versus a more final health outcome measure, and I wonder if anyone else has some thoughts on that because the evidence would need to be strong if it was an intermediate outcome.

CHAIR FLEISHER: Yes, I actually have a question about the re-intubation. You actually told us the rate is low, but you didn't tell us the outcome of the patients who got re-intubated within, who were extubated within 24 hours, which is actually the evidence to say

that's not relevant.

DR. JACOBS: I'm sorry. I'm not sure I understand your question.

CHAIR FLEISHER: So, in other words, you said that there was a low rate of re-intubation.

DR. JACOBS: Right.

CHAIR FLEISHER: But the real outcome is complications. So the question becomes either less than 24 hours and re-intubated, is that associated with a worse outcome than if you weren't re-intubated, because you could, because it's an intermediate outcome, ask should this be constructed differently that says less than 24 hours and stays extubated, not re-intubated. For example, the 24-hour prolonged ventilation measures actually are cumulative, as opposed to discrete.

So I'm not sure you answered the statistical relevance. It's a statistical

fact, it's not a relation to outcome per se, given it's intermediate role.

DR. JACOBS: All right. I think there's a fair amount of published data that showed that prolonged ventilation is associated with postoperative pneumonias, decreased survival, increased mediastinitis, and a variety of other complications.

CHAIR FLEISHER: Maybe I'm not making myself clear. What I'm just asking is the patients, by only asking less than 24 hours and not saying not re-intubated, like making those two discrete, it's a simple discrete yes/no, less than 24 hours initially. Is the patient who gets re-intubated, do they have a worse outcome such that you should construct the measure differently?

DR. JACOBS: I think the actual variable, Sean had it up on his computer so I can read it to you, but it's -- there it was. Yes, it's basically, it's a variable that

tracks -- the hours of postoperative ventilation time include OR exit until extubation, plus any additional hours following re-intubation. Does that address your question? I'm just reading it directly --

CHAIR FLEISHER: Yes, that addresses the question.

DR. JACOBS: Thanks.

MS. WINKLER: To respond to Collette's question again, I think this is something the Committee can consider because, as she points out, a true outcome measure, the evidence requirement is relatively limited to whether there are any processes of care or structures or something that can affect the outcome.

If, indeed, you feel it's an intermediate outcome, then the evidence expected is going to be much more like a process measure relating the intermediate outcome to more, you know, absolute outcomes, if you will.

So, again, a very good question to ask. When these measures are submitted, it is the developer who determines what the measure type is and, therefore, how they respond to the various questions. And, clearly, in this case, they designate it as a pure outcome measure and have provided the information on evidence accordingly.

But, again, it's up to the Committee to decide perhaps it should have been different. So I think that's a question Collette is raising.

DR. BUSRTIN: And just one comment from the -- I just pulled up the evidence task force report that actually delineated intermediate outcome. So at least the definition that NQF used was an intermediate outcome is a change in physiologic state that leads to a longer-term health outcome. And an outcome, obviously, is a little bit more clear, which is an outcome is the health status of a

patient or a change in health status resulting from healthcare desirable or adverse. So I think, definitionally, it does, at least my read, fit more as an outcome because it's really not an intermediate physiologic state but, oftentimes, would be considered an adverse event in and of itself.

DR. JACOBS: We would agree with that.

MEMBER YATES: Having thought about this measure, obviously, for a little bit, I think that, for that definition, you're meeting something that is obvious physiologic outcome. Breathing on your own is something that is relevant as an outcome, and I think it's very relevant to the patient and their families. As an end-user of that, the patient and families would see that as something that's an outcome. I have no question about that being, in and of itself, reasonable.

And my one question about the

one-percent re-intubation rate: is that consistent with historical re-intubation rates? Because my review, just picking up the literature, it seemed like re-intubation rates after CABG were slightly higher overall than one percent. Am I to assume that those patients are all the greater than 24-hour patients making up that lion's share of that literature, or has this team process gotten so much better that the rate is much lower now?

DR. JACOBS: I think the data that we have shows us that it's unlikely that one is going to be re-intubated if one has an isolated CABG and is extubated within 24 hours.

MEMBER YATES: Okay.

DR. JACOBS: That's a subgroup that's unlikely to require re-intubation. The ones that tend to require re-intubation are patients that are on the ventilator for a longer period of time, they're difficult to get off the ventilator, and they potentially could then

have a problem and need to be re-intubated.

MEMBER YATES: So that would be driving the overall larger number that I might have seen?

DR. JACOBS: Correct, absolutely.

MEMBER YATES: And one last comment about this being an outcomes measure, this is an ideal outcomes measure in that it's measuring a process that's many more, many more participants than just the surgeon or just the anesthesiologist. And as such, it's an excellent outcomes measure for the quality that's being provided.

DR. JACOBS: Thank you.

CHAIR FLEISHER: Cliff? And let's try -- we have a lot to get through today, so we'll focus on questions and then vote.

MEMBER KO: Just a quick question. So this is a re-evaluation of the measure. How has the performance in the database improved with prolonged intubation? Has it changed

anything, or is this where it's going to be?

MS. WINKLER: We're going to discuss that under performance gap, so why don't we wait until we do that? No, because we're going to, if you look at your script, you're going to vote on each one. You're going to do evidence and then performance gap and then priority.

So right now we want to focus in on any comments around evidence. And then we'll go to voting.

MEMBER CIMA: I wanted to clarify, it sounds like we're doing a time measure again, you know, 24 hours after operation. But when the definition was read, it seems like it's a total of 24 hours of intubation, not necessarily time to when they leave the OR, which is what the measure is asking. So I just want to make sure that the data is actually in the data set, or is someone else going to have to go back and do it? Because the way you just

defined it was 24 hours of total intubation after surgery, but that's not -- so if it's 18 - 20 hours or 26 hours after surgery, the patient gets re-intubated and stays on the ventilator 28 hours, that would be a positive in your data set the way you defined it, but it has nothing to do with the 24 hours here. So there's --

DR. JACOBS: I think it's the same thing. The measure here says percentage of patients aged 18 or older undergoing isolated CABG who require intubation for more than 24 hours postoperatively, and that's --

MEMBER CIMA: Oh, I thought it was within 24 hours.

DR. JACOBS: No, I'm reading it straight from the measure.

MEMBER CIMA: Okay.

CHAIR FLEISHER: So just as a point of clarification as we go through it, that was an important question, but that will fall under

reliability and specifications. So as we understand how this process goes, we'll be able to fit it into the right bucket.

MEMBER SIPERSTEIN: I was a secondary reviewer on this. Just to clarify that, 24 hours is cumulative. So that gets you out of the issue of a re-intubation, so that gets thrown in. And the way I see it, and I think it was stated, is that the outcome is really respiratory failure as measured by need for ventilatory support. So the outcome is not intubation, it's ventilatory failure. So if you think about it that way, it is a true outcome measure.

DR. JACOBS: Agree.

CHAIR FLEISHER: I think we're now up to voting. So -- oh, Rick?

MEMBER DUTTON: I had a dumb question. What's the statute of limitations on the 24 hours? So 24 hours intubated and the rest of their life or in that hospital admission

or --

DR. JACOBS: Hospital discharge.

CHAIR FLEISHER: All right. So just, again, a point of process. You all got your little clickers there. We've got numbers associated with each option here. In this case, it's just a binary choice: one for yes, the rationale does support the relationship of the health outcome to at least one healthcare structure, process, intervention, or service; or no, it does not.

We'll start up the voting just momentarily. You'll have 60 seconds to enter your vote, and it will display the results up there on the screen. Oh, yes, and point your clicker, I think, towards Amaru here.

MR. SANCHEZ: There will be a little green light on your remote. Once you see the light, that means we received it here. And it doesn't matter if you click once or twice, just as long as you click the number.

You don't have to press send. I know there's like a send button there. Do not press send. Just press the corresponding number.

CHAIR FLEISHER: And it will only vote once. Even if you press the button multiple times, you'll only register once.

MR. SANCHEZ: All right. Voting will now begin for criteria 1a, evidence for health outcome. One is yes, two is no. Voting will begin now.

(Voting.)

MR. SANCHEZ: Sorry. I'm having some technical difficulties here. I think we're going to have to do a re-vote here.

MS. WINKLER: We do need to ask Dr. Grover to just state that he's recusing from voting on this measure so that it's in our record.

MEMBER GROVER: Yes, I abstain or recuse myself.

MR. SANCHEZ: All right. Voting

will now begin for criteria 1a, evidence. One is yes, two is no. Voting start now.

(Voting.)

MR. SANCHEZ: Third time's a charm. Voting will now begin for criteria 1a, evidence for health outcome. One is yes, two is no. Voting will begin now.

(Voting.)

MR. LYZENGA: We're just waiting on one vote from the phone.

MEMBER JARRETT: Yes, hi. I sent it in in the chat.

MR. SANCHEZ: All right. So we'll move on to sub-criterion 1b, performance gap. And if anybody has any comments or questions or thoughts. Dr. Yates, do you have any comments?

MEMBER YATES: On the script it says opportunity for improvement. We're passing that?

CHAIR FLEISHER: That's performance gap. If you could give us your

thoughts on the performance gap.

MEMBER YATES: I heard performance and I don't -- my only question about performance gap was already addressed in terms of -- oh, that's reliability. Excuse me. The performance gap data is that there's a discrepancy of anywhere from 4 to 16 percent and the performance gap is reasonably high. And as such, I think this remains an important issue and one that is of value.

CHAIR FLEISHER: Any questions or comments? The secondary reviewer was? Robert, do you have a question? Your name tag is up.

MEMBER CIMA: No.

CHAIR FLEISHER: Okay. Are we ready to vote?

CHAIR GUNNAR: So I think the question came up earlier is what, over time, since this began in '07 and '11, what can you tell the Committee regarding what's been the

experience with this particular measure in relationship to the performance?

DR. JACOBS: Yes, I don't have that data with us.

MR. O'BRIEN: This is Sean O'Brien speaking, and I'm still in the middle of pulling up the data. But when the measure was developed using data from 2002 to 2006, the percent of patients experiencing prolonged ventilation was at 9.7 percent in the original publication, and the STS feedback report for isolated CABG patients -- if I'm reading something incorrectly I'm going to have to correct myself later, but the most recent report that I'm pulling up in front of me, in 2011 10.5 percent of patients extubated earlier and in 2013 8.8 percent. So I would say, basically, not -- I won't comment on the amount of change.

DR. JACOBS: 2011 was 10.5 percent and 2013 was 8.83 percent. So that's pretty

close, but, I mean, the data shows it dropped about a percent.

CHAIR FLEISHER: Yes, please.

MEMBER YATES: Question. Correct me if I'm wrong, but the original process measure was or, excuse me, outcomes measure was, in fact, the total time ventilated. And it has been changed to, from the time of the OR at some point over the last few iterations. I saw that in the description because it became, obviously, an issue that there was time ventilated versus time after surgery. So that may be something that --

DR. JACOBS: I think it's always been from the time you leave the OR as the starting point.

MEMBER YATES: Okay.

DR. JACOBS: Which is cumulative postoperative ventilation time, so I think the synonymous words are cumulative postoperative ventilation time equals a start time of when the

patient leaves the OR.

MR. O'BRIEN: The risk model was developed originally using data from 2002 to 2006 with different data elements. In the meantime, there's been some refinements, but it's always been a 24-hour time frame. And I know that, one way or the other, it was maybe clarification that they've been added to clarify that it was from the time leaving the OR. But I think that was obviously the original intent.

CHAIR FLEISHER: Okay. Which, again, would be specifications. But those are very important specifications.

MEMBER YATES: But it would have to do with the gap analysis because it may not have changed because it may have been a moving target.

CHAIR FLEISHER: Thank you.

DR. JACOBS: Agree.

CHAIR FLEISHER: Okay. Any other

comments, or are we ready to vote? Let's vote. You ready?

MR. SANCHEZ: Voting will now begin for criterion 1b, performance gap. One is high, two is moderate, three is low, four is insufficient. Voting will now begin now.

(Voting.)

MEMBER GROVER: I abstain again.

MR. LYZENGA: Can we have somebody turn off their mike? We can only have a few on at a time.

MR. SANCHEZ: We had 6 for high, 14 for moderate, 2 for low, and zero for insufficient.

CHAIR FLEISHER: So now we'll move on to high priority. Dr. Yates, any comments on this sub-criterion?

MEMBER YATES: I believe this to be a high-priority measure. It has an important impact on the patient and family and also has high cost in terms of prolonged intubation,

which is associated with prolonged ICU stays and prolonged hospitalization, high prevalence in that coronary artery bypass surgery is done frequently. And I would say that the impact on the health status is high, as well, in terms of severity.

CHAIR FLEISHER: Comments? I guess we're ready to vote.

MR. SANCHEZ: Voting will now begin for criterion 1c, high priority. One for high, two for moderate, three for low, four for insufficient. Voting begins now.

(Voting.)

MEMBER GROVER: This is Grover. I abstain again.

MR. SANCHEZ: We have 21 for high, one for moderate, zero for low, zero for insufficient.

MR. LYZENGA: Okay. With that, we can go ahead and move on to scientific acceptability, starting with reliability.

MEMBER YATES: The reliability is high. It's a 0/1 outcome. It's readily measured, and the data as given, in terms of past performance of the measure, would justify that statement.

MS. WINKLER: How was the measure tested for reliability? Was it tested at the level of the data element or the level of the measure score?

MR. O'BRIEN: In a sense, both, in the sense that the STS has a rigorous validation process that Dr. Jacobs may be able to describe in more detail. But in terms of the level, you know, in terms of reliability, one of the issues is statistical reliability in the sense of random sampling variation compared to true signal variation, and that was assessed, basically, for the purpose of this measure submission it was assessed by looking at the performance of the same participant measured in two different time periods, so one-year

snapshots of data back to back in two consecutive years and reporting the agreement between those two.

The agreement was higher when the data restricted to participants that had a larger number of cases. But it's in the submission material that I'll pull up, but I believe the correlation was in the 70s in terms of the Pearson correlation coefficient.

We've done additional analyses that are not included in this document to address more a formal estimate of reliability that basically gets at, kind of explain the variation in a measure that's driven by true signal variation compared to random statistical variation.

And it's possible that we can probably provide additional data that looks at the data that way. But I think that the agreement over time is basically what was included for review by this committee.

DR. JACOBS: And I would just add to that, in addition to the concepts of statistical reliability, which I think Sean addressed very nicely, there's the overall concept of completeness and accuracy of the data in the database. And this will apply to all the measures that we bring forward, and that's, I guess, another form of reliability.

And STS has a very aggressive data audit program in place that may be the most comprehensive data audit program that exists for a professional medical society. Ten percent of participants are audited every year, and a number of measures take place during that audit to assure the completeness and accuracy of the data and that the results of those audits year after year have shown that the completeness and accuracy of the data in the STS database is quite good.

MS. WINKLER: This is also the time to talk about any questions you had or further

questions about the specifications, that falls under reliability.

CHAIR GUNNAR: So just remind me and I think others, when you're a participating center and 90 percent of the, my understanding is from our call before, 90 percent of cardiac surgical programs are participants, and the ones that aren't are in the federal space or Kaiser. You have 100-percent mandated submission for all cases, correct?

DR. JACOBS: Correct. And part of the audit process is a comparison of the cases submitted to the actual operative log of the hospital to confirm that 100 percent of the cases are submitted.

CHAIR GUNNAR: So do those who are audited know that they're going to be audited, or do you just show up one day, or how does that work?

DR. JACOBS: It's a random selection of ten percent of participants every

year. So they don't know that far ahead of time, but they do get more than showing up at the door the morning of, like perhaps JCAHO visit might be. But it's not like a lot can change because they're being audited on data that's been previously submitted.

The ten percent of sites that are audited every year are selected annually. And there's some rules in place that if you were audited in the previous year, you're not going to be audited again, so it's distributed.

CHAIR GUNNAR: Do you report the results of that audit?

DR. JACOBS: We do. We have both internal documentation that gets circulated amongst STS leadership up to the level of STS board of directors and also information from the audit is shared publicly.

CHAIR GUNNAR: Last question. Obviously, as you go through this, there are a number of measures. Just for the Committee, I

think it would be helpful if you found any issues with regard to reliability in any reporting regarding any particular measure. It would be helpful. So the question here would be, in your audit, did you have any issues with the reliability of this particular measure?

DR. JACOBS: No.

CHAIR FLEISHER: Dr. Dutton, do you have a comment?

MEMBER DUTTON: My understanding is you report this at the level of the facility and the level of the cardiothoracic surgeon involved. Do you collect or analyze data based on the anesthesia team involved or the intensivist, the respiratory therapist, or any of the other participants on the team who contribute to this?

DR. JACOBS: So this measure is reported at two levels, at the level of the hospital and at the level of the cardiac

surgical program. And when it's publicly reported, it's publicly reported both at the level of the hospital and at the level of the cardiac surgery program.

The relationship between hospital and cardiac surgery program is not completely one-to-one. Most hospitals have one cardiac surgery program, but some hospitals have more than one program and some programs go to more than one hospital. And, therefore, it's reported using both of those methodologies.

We don't report this specifically stratified by anesthesiologist or ICU team or anesthesia team or bedside nurse. But we feel that the performance with this measure is reflective of the overall team process of caring for these patients. And, clearly, compliance with this measure is dependent on nursing, anesthesia, intensive care and surgery. So this is a measure that we think reflects the performance of the entire team,

and we stratify it either by hospital team or by cardiac surgical program.

CHAIR FLEISHER: Amy?

MEMBER MOYER: The specification lists that it=s for both the group practice and the facility level, I'm looking through the reliability, and it references participants, which could be either; am I correct? Is the reliability tested at all split out between, like, looking just at group practices, looking just at facilities, or is it at both?

DR. JACOBS: So the word participant when we use it for the STS database is most commonly a cardiac surgical practice. Rarely, it's an individual cardiac surgeon who's in solo practice, but most commonly it's a cardiac surgical practice. And we report the reliability stratified by the cardiac surgical practice. In most cases, that's also the hospital because in most cases there's a one-to-one relationship. But in some cases,

it's not a specific hospital because the surgical group goes to more than one hospital or the hospital has more than one surgical group.

CHAIR FLEISHER: Okay. It sounds like we're ready to vote.

MR. SANCHEZ: Voting will now begin for sub-criterion 2a, reliability. One is for high, two is for moderate, three is for low, and four is for insufficient. Voting begins now.

(Voting.)

MEMBER GROVER: I abstain again.

CHAIR FLEISHER: Fred, can we just agree that you abstain from all votes related to this measure? We'll just do it once --

MEMBER GROVER: That would be great.

MR. SANCHEZ: Thirteen for high, nine for moderate, zero for low, zero for insufficient.

MR. LYZENGA: Okay. So now we can

move on to sub-criterion 2b, validity.

MEMBER YATES: In terms of validity, it appears that the evidence does align with the specifications. The data submitted from previous experience with the measure appears to have been tested well for validity. The question of audit has already been brought up, and I think that adds to the validity, as well. And as such, I found the measure to have high validity in reviewing it.

CHAIR FLEISHER: Questions or comments? None. Shall we vote?

MR. SANCHEZ: Voting will now begin for sub-criterion 2b, validity. One is high, two is moderate, three is low, four is insufficient. Voting begins now.

(Voting.)

MR. SANCHEZ: Twenty for high, two for moderate, zero for low, zero for insufficient.

MR. LYZENGA: All right. Thanks,

everyone. So we'll move on to feasibility now at this point, I think.

MEMBER YATES: It would be feasibility because it's not a composite measure in terms of empirical analysis. The feasibility is very high in that it's an established registry from the STS. The participation is high across the country. The burden, in terms of cost, seems to be met readily by either practice or hospital. And the chart review is also apparently being met in terms of the data being routinely collected, and it appears to be a very feasible measure.

CHAIR FLEISHER: Comments? Let's vote.

MR. SANCHEZ: Voting will now begin for criteria 3, feasibility. One is high, two is moderate, three is low, four is insufficient. Voting begins now.

(Voting.)

MR. LYZENGA: We're still waiting

on one vote. Could you try to enter your vote one more time, everybody? There we go. Got it.

MR. SANCHEZ: Seventeen for high, five for moderate, zero for low, zero for insufficient.

MR. LYZENGA: And with that, we can move on to usability, criterion number four.

MEMBER YATES: Again, the STS registry has demonstrated a broad applicability and has been used by most programs involved with coronary artery bypass surgery that are outside of, say, the federal programs. The public reporting has accessibility.

The improvement over time has already been addressed, and there is slight improvement over time. And, again, there may be a question as to whether there was a shift in the 24 hours. But there is some improvement.

And the unintended consequence question that I had, which was the possibility of rate of re-intubation, as opposed to adding the hours back in, has already been addressed and I think adequately. So I would recommend it being seen as a usable measure.

MS. WINKLER: Just a question. I understand the measure is publicly reported by STS on a voluntary basis. How is the participation in public reporting been going in terms of the number of current participants and change over the last couple of years?

DR. JACOBS: This is Jeff Jacobs again. This is a very important question that will apply to, essentially, all of the measures that we discuss today. When we rolled out the voluntary public reporting from the Society of Thoracic Surgeons, participation in the initial year was 20 percent. It's now at about 50 percent, so it's gradually increased year after year.

We'd obviously like it to be 90 or 100 percent, and we have a number of ongoing initiatives that have been effective in getting us to increase public reporting from 20 to 50 percent. And I would anticipate that that number is going to continue to increase.

We could look at that as the glass is half empty or the glass is half full. I think that's probably the highest rate of voluntary public reporting of any professional medical society in the country right now, so that's the glass is half full. The glass is half empty is it's 50 percent, not 100 percent, but we're working on it.

CHAIR FLEISHER: Rick?

MEMBER DUTTON: Which 50 percent? I applaud you, by the way, for doing this, and you are absolutely right: you are farthest ahead of anybody. But is this Lake Woebegone?

DR. JACOBS: Well, I guess the best way to answer that is with some numbers. And

most measures that we publicly report are reported both numerically and stratified into a star system, which makes it more easily understandable. And the star system stratifies into one star, two star or three stars, one star being below average, two star being average, three star being above average.

And if we look at the distribution of the STS composite that includes this measure over all STS database participants, it's about 75 percent two stars, 12 2 percent one star, 12 2 percent three stars. If we look at the distribution in publicly reporting, it's probably about 8 percent that turn out to be publicly reporting at one star, rather than 15 percent.

So it's not that publicly reporting is all the three stars, some of the two stars, and none of the one stars. But it is a somewhat skewed distribution.

MEMBER YATES: This is Yates for

the transcriptionist. The question I have is that, moving forward, there's going to be a requirement for PQRS that includes, or a requirement for PQRS that allows for the use of registry data. And aside from participation and being qualified as appropriate by NQF, is the registry approved as a PQRS registry in terms of that reporting process?

DR. JACOBS: Yes.

MEMBER YATES: In which case, I would argue that the incentivization for more public reporting will become higher.

DR. JACOBS: Correct. I would agree with that. That's one of a number of potential mechanisms that public reporting will increase. We're actively collaborating with multiple states who have state-wide mandatory public reporting, working to have them transition from using administrative data to STS data for their state-wide public reporting.

In the state of Pennsylvania, for example, it has 100-percent reporting with administrative data, and there's efforts under way to transition to using STS public reporting in Pennsylvania. That will also increase the numbers. So we're getting there through a number of avenues, including the one you just described.

MEMBER SAIGAL: About improvement, I was wondering are there specific plans to share interventions or processes to reduce the rate?

DR. JACOBS: Right. So there's a committee within STS that's called the Task Force on Quality Improvement, and that's chaired by Rich Prager from University of Michigan. And one of the primary functions of that task force is to develop methodologies where data from the database can then be used to improve quality across the spectrum of the STS database. The ideal way for that to happen

would be to identify three-star programs, identify one-star programs, and let the one-star programs learn from the three-star programs. And this is an active area of work in the STS database through probably one of the most active task forces in the STS led by Rich Prager that strives to do exactly what you're talking about.

MEMBER SAIGAL: So then, to date, basically any change has been because you said a bar, and people were just aware of the bar in a general manner. Okay, thanks.

MEMBER KO: This is a follow-up of Dr. Yates' question of the PQRS. If this is a group or a facility, how is this -- are the specs changed when it's submitted for PQRS for the individual?

DR. JACOBS: So right now PQRS is just based on participation and not based on these outcome measures, and that's based on the individual. And I'm not really sure how it's

going to evolve when the PQRI or PQRS becomes based more on outcome measures and less on participation. But up until now, the STS database has been used as a tool for participating in PQRI, and that's been at the individual surgeon level.

MEMBER KO: Dr. Yates, is this a PQRS measure? Is that what you were saying?

MEMBER YATES: That was my question, but that's all evolving. There are many things going on. For instance, in terms of reporting, three months ago, if you looked at hospitalcompare.gov, you wouldn't have seen NSQIP data. As of a month ago, now you do. And, likewise, physiciancompare.gov, you know, is still working on whether or not they're going to use registry data or other measures as process measures for reporting.

So I think that's a question more for CMS than for STS, and I think the issue -- I think there are criteria, and I could be

completely wrong on this, but I think groups can report their group rates and collectively report how they do. But I'm not sure how that's going to play out with implementation of the legislation as it exists.

DR. JACOBS: Yes, I would agree with that. It's a rapidly evolving science where a lot of the changes are not in the domain of STS but in the domain of CMS. But an under-arching question really is the issue of reporting stratified by hospitals and practice groups versus reporting stratified by individual surgeons, and that plays out in this domain and in other domains and it's going to come up with other measures, as well. And I guess a generic answer to that is that, although this measure and most of the measures we're discussing are reported stratified by hospitals and physician groups, STS is actively working with DCRI to develop methodologies to have the ability to report cardiac surgical

outcomes stratified by the individual cardiac surgeon.

CHAIR FLEISHER: So I assume that will come back as a separate measure, correct?

DR. JACOBS: Yes.

CHAIR FLEISHER: That would have to come back if that was the --

DR. JACOBS: Correct.

MEMBER KO: So I have a question for NQF, and maybe Reva is the best person to answer it. When we look at a measure and, if it's a facility measure, it's going to be different than if we looked at it as an individual provider measure. Do we know if these measures are going to be in one subset or the other or both? Because it might change how we --

MS. WINKLER: There were a couple of questions in there. It really is the information reflected in the submission, and the developer determines what the level of analysis is. So there are occasionally

measures that have multiple levels of analysis, and so we expect them to address those multiple levels via testing. Sometimes, as you mentioned, or commonly, it's two separate measures for the hospital, the clinician, whatever.

The issue is we don't know how these measures may be used. What you're being asked to do is use the criteria to evaluate them whether they're suitable for use for accountability purposes, which may mean use in any of those programs, including public reporting, so they may be publicly reported. So we are providing the tools for those programs, but the programs themselves ultimately make the decision of which measures actually come into play.

CHAIR FLEISHER: But if they haven't been tested at the individual level, then they cannot be used as an NQF-endorsed measure, correct?

DR. BUSRTIN: Correct. The only thing I will add is that last year there was a pretty significant discussion at the Measures Application Partnership about whether, for some primarily hospital-based clinicians, they would be comfortable assuming the hospital rate to reflect their individual performance, particularly for hospitalists. So I think this is an issue that will likely come up in the surgical disciplines, as well, just to put that in the mix. But, again, not specific to the endorsement piece, but it is another way to at least reflect, have a clinician-level measure that actually isn't at the individual level but assumes taking on the hospitalist-level performance.

CHAIR FLEISHER: Before we go on, Dr. Erekson, can you introduce yourself briefly and tell us if you have any conflict of interest?

MS. EREKSON: Hi. I'm Liz

Erekson. I am at the Geisel School of Medicine at Dartmouth. I do have two conflicts of interest. I'm a member of AUGS, and they're submitting two quality measures, both the apical suspension and the cystoscopy at the time of prolapse.

CHAIR FLEISHER: Thank you. All right. If there are no other questions or comments, I think we can go ahead and vote on criterion 4, usability and use.

MR. SANCHEZ: We will now be voting for criterion 4, usability and use. One is for high, two is for moderate, three is for low, and four is for insufficient information. The timer starts now.

(Voting.)

MS. WINKLER: Keep pushing.

MR. SANCHEZ: We have 13 for high, 9 for moderate, zero for low, and zero for insufficient information.

MR. LYZENGA: We have nine for

moderate because we got one vote coming in late on the phone. If there are no other comments or questions on the measure, we can go ahead and move to the overall vote. You'll be voting on the overall suitability for endorsement.

MR. SANCHEZ: Overall suitability for endorsement. One is yes, and two is no. The timer starts now.

(Voting.)

MR. SANCHEZ: We have 22 yes for overall suitability for endorsement.

MR. LYZENGA: So the measure passes. Thanks, everyone. I think this is a little bit of a departure from our agenda, but I think we're going to actually take a break now. Our discussion ran a little bit over, so we'll let everybody take a little bit of a break. We'll do a 15-minute break; is that right? So we'll ask everybody to come back here at 10:30, and we'll start up with the next measure. Thanks, everyone.

(Whereupon, the above-entitled matter went off the record at 10:13 a.m. and resumed at 10:30 a.m.)

CO-CHAIR FLEISHER: It is 10:30 and we are going to try to get back on time. Right, Reva? So, where are we going next?

We were told not to be discouraged. It always takes an hour to get through the first measure. That they should have booked it correctly.

MR. ANDREW: We will try.

CO-CHAIR FLEISHER: But we know about accurate posting times. But we=ll try to move a little bit more quickly through the next few measures.

So now we=re moving on to measure number 0458. This is another STS measure, and I=ll go ahead and turn it over to Dr. Jacobs.

DR. JACOBS: Hi. Good morning again. This is Jeff Jacobs once again from the Society of Thoracic Surgeons. It was a

pleasure to have the opportunity to present the first measure. And we=ll now move on and move to measure 0458.

This measure is titled pulmonary function test before major anatomic lung resection (pneumonectomy, lobectomy, or formal segmentectomy). And a brief description is that this measure reports the percentage of thoracic surgical patients age 18 or order who undergo at least one pulmonary function test within 12 months prior to a major lung resection, which again is defined as pneumonectomy, lobectomy or formal segmentectomy.

This measure is admittedly a process measure rather than an outcome measure. But it=s felt by the thoracic surgeons within STS to be an extremely important process measure.

When we discussed this on the phone, the term was used that it=s an asymmetrical

process measure which is advantageous in that number one, it=s associated with low cost to do the test. And number two, the potential adverse outcome from the test is minimal.

So the only other piece of information I guess I could share for background before we open this up for the discussion, is the concept of why PFTs are important. As a thoracic surgeon, when one is deciding whether or not to resect part of the lung, whether it=s the entire lung in a pneumonectomy, a lobe or a segment, which is part of a lobe, the two major issues one must assess are resectability and operability.

Resectability basically means can one technically do the operation and remove the tumor. And that has to do with tumor burden, tumor size and tumor location.

Operability has more to deal with whether or not the patient is going to survive the lung resection, and be left with enough

functional lung to be able to live a meaningful life off the ventilator. And pulmonary function tests are a key test that are utilized to assess operability, whether or not a patient will be functional after a lung resection, be able to breathe, be able to breathe off the ventilator. Be able to walk up a flight of stairs after the lung resection.

So that=s the rationale for why this is an important process measure. And with that background, I think we can move forward with the formal discussion.

MEMBER SIPERSTEIN: Great. Thank you. I want to obviously minimize repetition. A lot of the issues about the STS database data collection, et cetera, have already been addressed.

Yes, this is clearly a process measure. As was explained, pulmonary function testing prior to major lung resection, has pretty much been accepted as a standard of care

for the reasons as described.

Also, it=s a useful measure as was discussed by the developers in their submission, was to compare treatment outcomes. And the whole point then is obviously to drive quality improvement.

And so I don=t want to belabor it, there were a number of publications that related to the value of doing pulmonary function testing and how they are used to both drive suitability for resection and assessing perioperative risk.

Jumping ahead to the algorithm in terms of how it should be rated. Obviously it went down the process pathway as opposed to the outcomes pathway. Yes, I felt that there was a -- they got a yes for systematic review.

They=re -- given the fact that this is regarded pretty much as standard of care, and supported in the literature, both the quality/quantity and you know, consistency

metrics I felt warranted a yes. And I feel that this then would drop into the high category.

CO-CHAIR FLEISHER: Comments?

MEMBER JARRETT: Yes, hi, this is Mark. I tried raising my hand, but I guess it=s not working. A quick question, and I don=t disagree with the you know, having been on the workgroup listening call and all that. My only question is why 12 months, and is there any difference between someone being looked at 12 months versus 6 months, versus 3 months. Has anybody looked at that.

In other words, is it giving too much leeway before the surgery, or is that adequate that if you=ve just had the 12 -- within the 12 months, you=ll get the same results as if it=s 6 or 3 months?

DR. JACOBS: I think that=s a fair enough question. And I don=t know that anybody=s actually done a formal study comparing outcomes of patients who had PFTs

within 6 months of their operation versus within 12 months of their operation.

The timing of preoperative PFTs is clearly a continuous variable and we=ve dichotomized it by making a cut off at 12 months. And I don=t know of any evidence that suggests that there=s a different cut off that would be any better. But I think 12 months is certainly as face validity amongst the thoracic surgeons involved with the development of this measure.

MEMBER JARRETT: Okay. Thank you.

CO-CHAIR GUNNAR: Again, it=s always what do you do with the information, right? The mere presence of a PFT satisfies the measurement. But it doesn=t necessarily satisfy the thoughtful and intelligent use of that measurement.

I guess what=s the connectivity between you know, two flights of stairs, the old way of measuring it, and a PFT, and whether or

not you actually did the PFT correctly with good participation and a DLCO, and then used that information in a way that it actually resulted in a decision.

So how do we make that connection as a committee with regard to this particular measure?

MEMBER SIPERSTEIN: Maybe as a correlate to that, maybe that=s jumping ahead to a question I had in a later section, you know in terms of is there data that patients who do not get pulmonary function tests, you know, are their outcomes any worse? Is that a different patient population, for example, who are healthier patients who can walk a five flight of stairs, or who are having lesser procedures out of that CPT bucket.

DR. JACOBS: All right, well I think when I trained in thoracic surgery, and I was taught by Dr. Thurber who is a long time, old time thoracic surgeon, he also taught me

this two flights of stairs rule that you=re talking about.

He=d say that if they can come in the office, walk up two flights of stairs, they=re going to tolerate the pneumonectomy. The problem with that is obviously that there=s a lot of reasons people can=t walk up two flights of stairs, only some of which is their lung function.

Some of which might be that they have bad knees. Or that they have arthritis, or any of a variety of other problems. And this at least allows some scientific quantification of why they can=t walk up those flight of stairs.

I think that there certainly is a possibility that any test that one orders could be performed wrong. That=s just a fact of the way we do business. And then there=s the fact that any test that one orders could be ignored or utilized inappropriately.

But most thoracic surgeons, when armed with the data from pulmonary function tests, are going to know what to do with it and know how to utilize that data. That=s fairly a basic concept in thoracic surgery.

And I think that despite those potential risks, that the test could be performed wrong, or that it could be performed and ignored, it=s still a very valuable measure to know whether or not pulmonary function tests were done before a formal anatomic lung resection.

MEMBER SAIGAL: That seems to make sense to me. And it has good face validity. And I guess the quality evidence in this topic was supposed to be good.

So I suppose some of this stuff was reviewed in terms of its comparative effectiveness to other sort of more heuristic measures about just walking the stairs, and the timing of it, and when it should be done was also

covered in the evidence. I would assume it was good quality evidence.

DR. JACOBS: Yes, I mean I guess if the question is was the evidence presented good quality evidence, the answer would be yes. And the supporting documentation provided is ample peer reviewed literature that=s gone through peer review and documented the importance of doing pulmonary function testing prior to anatomic lung resections.

MEMBER MOYER: I had a question relating to the quality of the evidence too. There=s a guideline recommendation that=s cited with a grade 1B, which is moderate quality evidence, and a grade 1C which is low or very low quality evidence.

And I was wondering for the studies that are listed, if you could potentially walk through them and talk about are they like a meta-analysis, an RCT, are they a series of articles listed, but it=s not clear what

they=re reporting on. And I=m looking at section 1A3 of the application.

DR. JACOBS: Well I think there=s been no meta-analysis of a series of perspective randomized trials of whether or not patients have had pulmonary function tests or not had pulmonary function tests and then compared their outcomes. So I don=t think that a prospective randomized trial would ever be carried out to obtain that level of evidence.

I thank rather than go through all the articles and talk about the data within individual articles, what I -- I would make an analogy and the analogy would be that there=s never going to be a prospective randomized trial that documents whether or not it=s a good idea to wear a parachute when jumping out of an airplane.

However, there=s probably reasonable evidence that it=s a good idea to do that based on extrapolating from other

principles. And I think the same concept exists here. Patients are not going to enroll in a prospective randomized trial of using PFTs versus not using PFTs before their lung resection, just like they=re not going to enroll whether or not to use a parachute before jumping out of an airplane.

So that might mean the highest level of evidence might never be achieved. But I think the highest level of evidence possible to support this measure has been achieved in the literature.

Was that helpful?

MEMBER SIPERSTEIN: I just want to comment that going through a number of those papers, I mean not reading the whole thing, but then kind of going through the abstract and the gist of them. They really focused on how would you interpret pulmonary function tests, and how does that guide resectability and outcomes?

The implication obviously, is that

the test is valuable in that respect. So the implication is better to do them then not to do them.

CO-CHAIR FLEISHER: So I would like to actually follow up and discuss sort of the contrapositive of what you stated in that. Patients with poor functional status, there is value in getting this test.

To understand it, to disclose a mature ACC/AHA preop cardiovascular testing guideline, when we ever always discuss echos, it=s always in patients on the low side. There is a small gap in looking at the analysis. And it=s quite small.

So is there a subset in which it=s clear there=s evidence, but to say everyone needs to get it going back to the old, that if you really have excellent exercise capacity, is there an unintended consequence of doing low value care?

DR. JACOBS: Yes. So I would

answer that by saying that if this was either an expensive test, or a risky test, it would be worth pursuing that a little bit further. But this is a very inexpensive test that has essentially no risk to the patient.

So to try to identify a subset of patients that don=t need this, really I think is not a worthwhile exercise. Because the test is so inexpensive and so low risk.

CO-CHAIR FLEISHER: Anybody else? Collette?

MEMBER PITZEN: Collette Pitzen. Maybe this isn=t the right time to ask this question. But in our committee guidebook, we=re asked to think about if something is important to measure versus important to do in clinical practice. So I just want to make sure that we have that discussion.

Dr. Siperstein said that this is standard of care. We have pretty high performance rates. So I just want us to keep

that in mind. Thank you.

CO-CHAIR GUNNAR: Yes, I would like to echo that. Because is it you know, what=s the -- the use of a stethoscope versus getting a cardiac echo, right? Use of your clinical judgement and relationship to any particular patient versus I must have a PFT to check a box.

When I know that this is an otherwise you know vibrant individual=s functional class one, who can -- who=s jogging, but having to have an isolated lung lesion that was going to get resected. But I must put him through PFT, which is an important concept of where does -- where do we -- where does forced technology versus clinical judgement, where do they intersect here?

MEMBER SIPERSTEIN: But we do that within Title CO2s and measuring pulse oximetry in every patient that=s intubated, as opposed to saying well I=m an experienced anesthesiologist, which I=m not, and I know

that the tube=s in the right place and everything=s okay.

So because the -- kind of the, as was stated, kind of the asymmetry, the low bar to do the test, and the low cost to do the test, versus the high consequence of an adverse outcome.

CO-CHAIR FLEISHER: So this gets to the question of if the performance measure -- I mean it doesn=t -- we haven=t gotten to gap yet, but if the performance measure disappeared, would the gap increase? Or is this now standard of care, which gets back to Helen=s original and Reva=s original comments?

How big a gap do we have? We can do that, but that reflects the evidence.

DR. JACOBS: So what I can provide is that we looked at data from July, 2010 until June, 2013. And out of 28,000 patients, 28,043, 26,609 had pulmonary function tests done. And 1,434 did not. So that still shows

that there=s a chunk of patients getting formal anatomic lung resection without PFTs.

CO-CHAIR FLEISHER: Rick?

MEMBER DUTTON: I=ll note that it does say in the submission that the -- in the STS database at least, this is -- PFTs are an independent predictor of badness, bad outcome. So relative to whether it=s a better test than just looking at the patient=s age or listening with a stethoscope, or what have you.

CO-CHAIR FLEISHER: So are we ready to vote? Any other comments? All right, I think we can go ahead and vote.

MR. SANCHEZ: Voting will now begin for subcriterion 1A, evidence. 1 is high, 2 is moderate, 3 is low, 4 is insufficient evidence. Timer starts now.

MEMBER GROVER: And I=m abstaining on all of the votes on this particular element.

CO-CHAIR FLEISHER: Thank you. Can we have you click your vote one more time.

Yes, one more. And remember to point the clicker towards the marrow here.

CO-CHAIR GUNNAR: Just so Dr. Erekson knows, the green light has to show on the -- when you push the button, that=s all. To make sure you=re.

MR. SANCHEZ: All right. 4 for high, 16 for moderate, 2 for low, zero for insufficient.

MEMBER SIPERSTEIN: So opportunity for improvement. As we mentioned, and the data was presented in two time frames. It was an earlier time frame from >08 to >11, and a later time frame that actually overlapped a year, from 2010 to 2013. And the performance on this measure was in the 91 to 92 percent in the earlier time frame, and moved up to 94 percent in the later time frame.

So my interpretation of this is that the compliance is relatively high. But clearly there is room for ongoing improvement.

Some limited data was presented on disparities, both sex, race, ethnicity data, and due to the I think in part to the relatively high compliance, there was at most only approximately one percent disparity noted in any of those groups.

CO-CHAIR FLEISHER: Comments?

MEMBER MARKHAM: I had the other measure in terms of the registry with the general thoracic surgeons, and I brought this up in the conversation. What, any of the -- I mean there was a very small amount of surgeons who in that registry, was this data based upon that registry?

DR. JACOBS: I=m not 100 percent sure what you=re asking. But I think you=re asking was this data -- was the data that we=ve presented based on data in the STS thoracic surgical database, and unless otherwise identified in the measure submission form, that=s the source of the data.

MEMBER MARKHAM: Right. And I was thinking, I mean exactly -- okay, if you took the total number of 28,000, which is the --

DR. JACOBS: Oh, the numbers I gave come from the STS thoracic database for sure.

MEMBER MARKHAM: Right. Now is that -- is that the total amount of these procedures performed?

DR. JACOBS: That number of 28,043 is the total number of lobectomies, pneumonectomies, or segmentectomies in the STS thoracic surgery database from July, 2010 until June, 2013.

MEMBER MARKHAM: In the bigger picture how many of these procedures, do you believe -- I mean, I=m trying to show that the performance gap is probably greater than what you=re proposing.

DR. JACOBS: Yes, I understand where you=re getting at, and I think you=re raising a very important point. So these data

come from participants in the STS thoracic surgical database. And unlike the STS adult cardiac surgery database that has a 90 percent plus penetrance, the penetrance of surgeons who do thoracic surgery in the United States is lower in the STS thoracic database, in part because thoracic surgery is also done by general surgeons who are less inclined to participate in this database.

And I think you=re absolutely right that the outcomes reported in this thoracic database that show 1,400 patients, I=m sorry, 1,400 patients not getting PFTs and 26,000 patients getting them, the gap may be even higher in non-database participants. And I think that=s the point you=re making. And I would agree with that completely.

MEMBER MARKHAM: Right. Right. So it may not be reflected.

DR. JACOBS: Right, exactly. So there=s a performance gap that we=re showing by

database participants, that performance gap theoretically is higher in non-databased participants. I would say it probably is higher, although obviously the database can=t provide that answer.

MEMBER MARKHAM: Right.

CO-CHAIR FLEISHER: But the measure would only be relevant to the database. John?

MEMBER HANDY: Well, I=m trying to be quiet as a thoracic surgeon and not shoot my mouth off too much.

There=s actually an article that was published looking at pulmonary resection in the STS database, and comparing it to larger administrative databases and the STS database is not reflective of the national experience. It=s a minority, and the results are much better. So I think that your supposition of the performance gap is much greater than what is being presented here today is correct.

DR. JACOBS: Exactly, there=s been a number of studies by STS, including the one that you=re describing, that shows that outcomes of thoracic surgery in the STS database are better than overall national aggregate outcomes as assessed from administrative claims data. Or in other words, that the participants in the STS database are the ones who tend to have the best outcomes for whatever reason.

CO-CHAIR FLEISHER: So you just identified a potential gap in a measure that would be created because it would be outside the database. Recognize that as a standing committee.

DR. JACOBS: Now, the one thing I would say is that this particular measure is not database dependent. You know we=re going to talk in the future about participation in the database. This measure just says pulmonary function tests before major anatomic lung

resection.

The data we=ve used to support the measure come in a large part from the STS database. However there=s a variety of ways one can track whether or not PFTs were done before a lung resection.

So this measure=s not dependent in any way on participation in the STS database to comply with the measure.

MEMBER HANDY: That=s fine with me.

CO-CHAIR FLEISHER: Yes.

MEMBER MOSS: Hi, Larry Moss. So current discussion notwithstanding, I=m inferring that your goal here in this process measure is to reduce the incidence of postoperative respiratory failure and/or complications.

DR. JACOBS: Agreed.

MEMBER MOSS: Can you give us, or do you have any information that would suggest that closing that few percentage points gap in

the performance of pulmonary function tests would meaningfully reduce respiratory failure and improve the desired health outcome.

DR. JACOBS: Well I think the evidence we have is extrapolated evidence from the abundance of articles and the literature, that document that obtaining this information prior to doing an anatomic lung resection is essentially standard of care because it is felt that it can reduce the incidence of postoperative respiratory failure.

But I think to get that -- to answer that question with a degree of specifics -- with the degree of precision that one would like that would really take a prospective randomized trial. And that=s not going to happen.

So I think in the absence of that, the best data that exists is the data that=s currently published in the articles that we=ve referenced.

CO-CHAIR FLEISHER: Great. It

would be important to continue moving on, or we will be here through Friday, so. All right, let=s go ahead and vote.

MR. SANCHEZ: Voting will now begin for subcriterion 1B, performance gap. 1 is for high, 2 is for moderate, 3 is for low, 4 is for insufficient. Timer starts now.

Would you please point it towards me when you=re casting your vote please.

We have 4 for high, 12 for moderate, 3 for low, zero for insufficient. 7 for low, zero for insufficient.

MR. ANDREW: All right. So we can move on to subcriterion 1C, high priority.

MEMBER SIPERSTEIN: So I think a number of these items have already been discussed. And that thoracic surgery is frequently performed. And we=ve already had discussions on the high potential consequence of lack of study -- of testing pre-surgery.

CO-CHAIR FLEISHER: Seeing no

other comments, we can go ahead and vote.

MR. SANCHEZ: Voting will now begin for subcriterion 1C, high priority. 1 is for high, 2 is for moderate, 3 is for low, 4 is for insufficient. Timer starts now.

MS. WINKLER: Keep pushing your buttons.

MR. SANCHEZ: We have 6 for high, 12 for moderate, 5 for low and zero for insufficient.

MR. ANDREW: Thanks everyone. So let=s go ahead and move on to scientific acceptability, starting with 2A, reliability.

MEMBER SIPERSTEIN: So the numerator statement that=s been pretty well stated, is the number of thoracic surgery patients greater than 18 who were undergoing at least one pulmonary function test within a year of surgery. The denominator number of patients undergoing major anatomic lung resection, and that list of CPT codes is clearly

stated in the documentation.

The exclusions are the inability to perform pulmonary function tests either due to tracheostomy or other medical comorbidities, such that the patient cannot understand or cooperate with the test. Or in patients who cannot have the testing due to urgent or emergent surgery, for example emergency surgery for lung abscess massive hemoptysis, et cetera.

The data source, we=ve already discussed, is the STS. Registry issues or concerns with the definitions or coding, I feel that there=s a fairly limited risk of subjective interpretation from inability to perform or emergent operation because those are -- those criteria are further specified in the documentation.

So I think it would be somewhat difficult to quote, you know wiggle out of, or reclassify somebody just because the test

wasn=t done.

I=m going to move on to the reliability --

MS. WINKLER: Testing, yes.

MEMBER SIPERSTEIN: Testing. So the measure as is a lot of the data elements are straightforward, and easily culled from the medical record. The score itself obviously is the ratio as described above.

There was extensive information on the test sample method of testing, et cetera. They looked at more recently, the 28,000 patients as discussed over the past three years, this included 218 separate sites. As has already been discussed in the prior measures, the auditing process that goes on to validate the input of the data, and this was showed to have a high -- the auditing was showed to have a high agreement with what was initially placed in the database.

So based on the you know, integrity

of the data, my interpretation is that it would be fairly highly reliable.

MS. WINKLER: Comment. Was the measure tested for reliability at the level of the measure score?

MEMBER SIPERSTEIN: Yes, it was looked at both the individual data elements as well as testing of the ratio itself.

MS. WINKLER: Okay.

MEMBER SIPERSTEIN: And there were several pages of data that are pretty much included with many of the STS measures that go through a similar panel of validity and reliability testing.

CO-CHAIR GUNNAR: So I have a question. With regard to the audit process. So you=ve got the cardiac programs, ten percent audited per year randomly. Are thoracic programs ten percent as well, or is that --

DR. JACOBS: Right, so there=s three STS databases that are going to come up

over the course of the next two days. The adult cardiac, the general thoracic and the congenital. All three of them undergo essentially identical audit processes with ten percent of the sites being audited every year.

MR. ANDREW: If no other comments, we can go ahead and vote on reliability.

MR. SANCHEZ: Voting will now begin for subcriterion 2A, reliability. 1 is for high, 2 is for moderate, 3 is for low, 4 is for insufficient. Timer starts now.

MR. ANDREW: Still waiting on one more. If you could revote one more -- oh, here we go, got it.

MR. SANCHEZ: We have 13 for high, and 10 for moderate, zero for low, zero for insufficient.

MR. ANDREW: All right, so let=s move on to subcriterion 2B, validity.

MEMBER SIPERSTEIN: So some of these elements we=ve already gone through.

We=ve discussed whether the specifications align with the evidence relating to the relatively low cost and morbidity of the test itself, and the high potential consequences of not doing it.

We=ve touched upon whether it was tested for validity at the data element level, measure level score or both. There was data presented on the predicted validity that kind of looked at the stability of that measure over time. So there wasn=t a lot of noise in the data.

Which brings up again, the question I asked, I don=t think I got a clear answer to is whether there were poorer outcomes demonstrated in the database for patients who are not tested, or whether that was a different population within the database that was not tested.

The test sample involved dividing the group using 95 percent confidence intervals

into a big chunk in the middle, and then the smaller groups at each end of the tail, who were the low and high performers. And they demonstrated differences between those groups.

Given kind of the clarity of the definition, I don=t think there=s any real threat to the validity of the elements. There was no risk adjustment for this. The exclusions have been clearly stated, as I did previously.

The meaningful differences was again outlined and data was presented in the submission using the high, mid and low confidence grouping. And any missing data in the database is scored to adversely affect the outcome, i.e., it would be if there=s missing data, it would be scored as not having done a pulmonary function test. Or if the case type is not specified, it would be categorized as an elective case.

And then -- so in terms of all of

those issues, documenting the validity of the measurement from you know, data collection analysis point of view, I think it met the criteria to be either you know moderate or more likely in the high category.

CO-CHAIR FLEISHER: Other comments? Okay.

MR. SANCHEZ: Voting will now begin for subcriterion 2B, validity. 1 is for high, 2 is for moderate, 3 is for low, 4 is for insufficient. Timer starts now.

We have 11 for high, 10 for moderate, 2 for low and zero for insufficient.

MR. ANDREW: Thanks. All right, let=s go ahead and move to feasability.

MEMBER SIPERSTEIN: So again, we=ve -- for those groups that are participating in the STS database, they already have all the engines in place to collect the data. Obviously very high adoption rate. It was also mentioned that many of these

procedures may not be done by thoracic surgeons who may not be participating.

But in and of itself, that separate group, the data elements would not be very difficult to collect outside of the STS system.

CO-CHAIR FLEISHER: Comments?

MEMBER CIMA: Just one on that one point. For those not participating in a community based practice, or something like that, not an integrated practice with an integrated EMR, is it necessarily going to be easy to get this type of data? So who=s going to report this?

If the hospital is reporting it, and the general surgeon in the community sent a patent to his pulmonologist friend who has a PFT testing in his outpatient clinic, not part of the hospital, does the PFT he=s faxing over to the surgeon=s office, how are we going to collect that data for them?

DR. JACOBS: This is Jeff Jacobs.

I think that most surgeons, when they dictate their operative report of a lung resection, in the dictated operative report, they would describe what the pulmonary functions are.

So that dictated operative report is going to be in the medical records at the hospital. And they=ll say this is a 45 year old gentleman with stage one lung cancer, had preoperative pulmonary function test that showed an FEV of 2 and a half liters.

So it=s going to be in the first one or two sentences of the operative report. And it will be a pretty easy fact to get right out of the hospital medical records.

MEMBER CIMA: So as for now, we=re mandating how people have to do their operative reports. Do we know that --

DR. JACOBS: No, I don=t think we=re mandating it, but I would imagine the overwhelming majority of thoracic surgeons dictate that as part of the justification in

their operative report.

CO-CHAIR FLEISHER: Amy, did you have a question?

MEMBER PITZEN: Given that -- given that the measure is submitted and tested as being from a clinical registry, is that the only source we would consider for feasibility and usability?

MS. WINKLER: Yes, I mean that=s really the thing you have to look at, because that=s what we know about the measure. I think it=s reasonable as a sideline to say that the specifications are straightforward enough that perhaps it could be used.

But really what we are evaluating, because that=s the only information we have to evaluate, is its use within the database.

MR. ANDREW: All right, let=s go ahead and vote on feasibility.

MR. SANCHEZ: Voting will now begin for criterion 3, feasibility. 1 is for high,

2 is for moderate, 3 is for low, 4 is for insufficient. The timer starts now.

We have 9 for high, 11 for moderate, 3 for low, zero for insufficient.

MR. ANDREW: Okay, let=s move on to usability and use.

MEMBER SIPERSTEIN: So the documentation includes a statement that it=s used for quality improvement both externally and internally within organizations. It is "planned to have public reporting in the future." My understanding is it=s not -- this particular measure is not currently publicly reported.

Improvement over time, I=ve reported that the data submitted shows that it=s gone from a 91 to 92 percent range up to 94 percent. And we=ve all -- unintended consequences I think would be quite minimal in this.

And I guess some of the more you know

global areas that we haven=t quite you know touched upon, have to do you know, with whether there=s certain aspects of this you know, process measure, that are you know, potentially rolled up in more outcome types of measure just in terms of the usability.

CO-CHAIR GUNNAR: Comments?

CO-CHAIR FLEISHER: I have a question. You have a risk adjusted mortality score and you have a risk adjusted prolonged length of stay score. How long do we need this process measure to -- I mean the question of process versus outcome?

DR. JACOBS: I think that=s in some ways a higher level discussion than just the discussion of this measure. For a coronary artery bypass grafting, we have risk adjusted mortality and a variety of risk adjusted morbidities, but we also have a process measure of memory utilization.

And I think it=s a very similar

discussion -- and I think it=s a very similar discussion with that process measure in relation to outcome measures and this process measure in relation to outcome measures. If one is going to have any process measures related to lung resection, I think this is probably the strongest process measure one could have.

If one is going to take the position that we should eliminate all process measures if we have outcome measures in that field, then I could say well yes, then probably this should be eliminated. But at a higher level, if there=s going to be process measures and outcome measures both, then certainly this would be the process measure for lung resections.

CO-CHAIR FLEISHER: So as Reva pointed out, this is more of a committee level decision, --

DR. JACOBS: Exactly.

CO-CHAIR FLEISHER: Particularly the lack of data regarding non-performance in outcome is in your data set, is an important question as we go forward from my perspective.

MEMBER SIPERSTEIN: I guess the reason I keep asking that question is you know, it=s data that must exist in the data set, but has not been -- not been reported.

CO-CHAIR FLEISHER: So that -- I think I=m echoing your comments. If we knew that this gap was associated with worse outcome, specifically in your data set. That=s the question.

DR. JACOBS: Okay, I don=t have that data.

CO-CHAIR FLEISHER: Yes, so that=s a committee decision.

MEMBER PITZEN: Just to comment. According to the criteria for public reporting, this measure=s been endorsed since 2008. So we=re approaching or a little bit past that six

year expectation.

MS. WINKLER: Is this measure publicly reported?

DR. JACOBS: So the STS thoracic database is about to begin a -- begin publicly reporting next year. Right now, STS has implemented public reporting in a stepwise fashion where we=ve been rolling out one measure every year.

We started with isolated CABG, then the next year we added in isolated AVR, aortic valve replacement, the next year we added in isolated aortic valve replacement and CABG. This year we=re adding in congenital cardiac surgery public reporting. And in 2015 we=re going to begin reporting a variety of measures related to lung resection from the thoracic database.

And that=s just been a strategy we=ve taken to make sure that we get it right. So in order to get it right, we=ve focused on

rolling out one new measure in our portfolio of publicly reported measures ever year.

MS. WINKLER: Thanks for the explanation.

CO-CHAIR FLEISHER: Any other comments?

MR. SANCHEZ: Voting will now begin for criterion 4, usability and use. 1 is for high, 2 is for moderate, 3 is for low, 4 is for insufficient information. The timer starts now.

MR. ANDREW: We are waiting on one more. If everybody could recast your vote. One more time if you don=t mind. Remember to point at the marrow view.

MR. SANCHEZ: We have 1 for high, 12 for moderate, 9 for low, zero for insufficient.

MR. ANDREW: Any additional comments or questions before we move to an overall vote?

CO-CHAIR FLEISHER: So I actually

would like to make a recommendation. With my perspective on CSAC. I think the information that Alan and I have been asking for will be critical for how I think about this measure in CSAC.

Whether or not there is in the patients who actually don=t get their PFTs. Whether that is associated with worse outcome. So I will have a hard time making that decision.

And the other is I would actually ask the developer to have that.

DR. JACOBS: I think the real question though is in the patients who do get their PFTs, would their outcome have been worse if they did not?

Because when you look at all the patients in the STS database that are going for lung resection and some get PFTs and some do not, and one tries to compare those outcomes, the reality is that probably the ones that are not getting them are the healthiest ones where

the surgeon says well maybe we don=t need them, because this is a healthy patient who doesn=t need them.

And therefore it might not be the fairest comparison. A fair comparison would be to randomize all patients to one or the other. And that really can=t be done because of this -- because of the parachute analogy.

So comparing patients who have gotten them to those who have not gotten them within the STS database is a fairly biased comparison.

MEMBER SIPERSTEIN: But you=d be able to do the propensity analysis to see if there are certain CPT codes, or certain other patient characteristics that led to non-testing.

DR. JACOBS: Right, you=d have to do a formal risk adjusted comparison between the two groups to try to get at the real question of if we didn=t do them and the patients who got

them, would their outcome have been worse?

And I guess the closest we could get at that would be try to do a risk adjusted comparison between those that got them and those that did not. Assuming that that might be possible with the variables that are captured in the database.

MEMBER SAIGAL: Your statistician could do an instrumental variable analysis trying to look at the unabsorbed characteristics in this database. But the bigger question, and maybe there=s no funding for that, but the big question I think in my experience with the GU measures, when the committee approved process measures that didn=t have a clear outcome link, the CSAC overturned the findings of the committee.

So I think this is an important thing for us to discuss in terms of whether we have a mandate to only look at outcome measures now. Or very strong process outcome links

measures.

MS. WINKLER: I would say both.

MEMBER DUTTON: Another way to take a shot at it, if you have groups that have shown improvement and performance on this measure over five years, have they also shown an improvement in their outcomes?

MR. O=BRIEN: This is Sean O=Brien, and I think that the ability to use indirect information like that, that sounds in a way similar to an instrumental variables analysis when one version of instrumental variables analysis would be to compare outcomes among sites that use -- that record the PFTs close to 100 percent of the time, compared to sites that use it infrequently.

But as a way of getting at kind of the true underlying counterfactual causal effect, do you want to know what the patient=s outcome had been had they received the measure.

There=s going to be a limited

ability for that type of analysis because the spread across sites, you know we do have the sites that are up at the 100 percent. There=s not that many sites that are down at the zero percent. And so there -- the kind of any causal association will be attenuated by the mix of patients that are actually -- maybe at the sites on the low end may be down in the 80s and their -- in addition, a lot of other -- when you=re comparing groups of sites that perform in different ways, there=s going to be other characteristics of the sites that are going to be different between those sites.

And then the consideration. So the other way of doing analysis is just more of a standard propensity analysis. Looking at outcomes of individual patients who have had the PFTs and those that didn=t.

And I mean I think this is a little bit subjective in the assessment of how do you value the strength of evidence generated from

observational data that when you have a treatment that=s given to 95 percent of patients and you develop a propensity model where you=re really predicted together, you now, pretty much all of the patients, or a large fraction of the patients are predicted to get the treatment.

And if you see patients that have you know, 99 percent plus predicted probability of receiving the treatment, and they don=t, well I think you have to ask well what was different about that patient? Are they different in ways that were captured in your analysis? Or could they be different that were not captured in the analysis.

And the uncertainty about that question is why there=s always questions at the end of the day with an observational study that only can be addressed by a randomized trial.

CO-CHAIR FLEISHER: Barry?

MEMBER MARKHAM: The only other

issue I have is the timing of the PFT. I mean a lot can happen in a year. And it=s 12 months, and if you=re going for a major resection, I would like to have it closer to the actual procedure than one year time limit. I think it would be more pertinent.

CO-CHAIR FLEISHER: And I just want to echo one comment that was said in the beginning. There=s a difference between whether or not something should be measured and whether something should be standard of care.

So to -- from my perspective, the argument, even the propensity matched argument, if the group who doesn=t get it, are really the group that do fine, to me that actually echoes and the standard of care is currently acceptable. That=s one person=s perspective.

So realize we=re asking about a performance metric that we are endorsing. Not whether or not this should be a standard of care

accepting certain circumstances.

MEMBER SAIGAL: You=re getting at the evidence. Basically we=re back to evidence on how important this measure is. Which was the very first thing we talked about.

So we said the evidence was good is how the committee voted. But the evidence for what? That=s the question.

So apparently the evidence is not good given what you=re saying. Given what your standards are from the CSAC.

CO-CHAIR FLEISHER: I=m representing my individual thought processes. How CSAC will vote, is --

MEMBER SAIGAL: Got it. But I think just to be frank with the committee, I mean it=s a different -- I think there=s a different sense of what=s important in terms of the standards.

So I think we should all understand what that is so that we can make use of this

venue appropriately.

MS. WINKLER: Let me just respond to Chris= comments. Indeed from the Board of Directors and supported by the CSAC, there is definitely as Helen mentioned earlier, a hierarchical preference for measures.

So I would agree with you Chris, that the preference for outcome measures is absolutely there. Also, process and perhaps structure measures that have really solid evidence association.

But I think that -- I think the question that you=re asking is in the face of having outcome measures, what is the value of process measures? And I think that=s the right question to ask as you look at the entire portfolio of measures.

Because volume of measures is not necessarily a good thing. So the question is do we have the right measures. So you=re asking the right question. And this is why we

bring you all here to grapple with you know, sometimes thorny issues.

MEMBER SIPERSTEIN: Let me ask Lee and Reva. Just in terms of looking at this as an isolated measure. Is there you know philosophically a higher or a different criteria for endorsement? i.e. if a measure has been out there for six years, is there a you know, requirement or desire to prove its, you know, utility or effectiveness with data, as opposed to a measure that=s being proposed for the first time that in some ways isn=t out there to be road tested?

MS. WINKLER: Actually, if you read through the criteria carefully, you will find that there are comments about expectations for measures undergoing maintenance review. In other words measures that have been endorsed.

And so things like testing, it really is the expectation that measures coming in for maintenance review will be more

thoroughly tested than they were during their initial evaluation. That they will be tested at the level of the measure scores so that we can really understand how they=re being used.

The assessment on usability, you know, how is it being used? If it=s not being used, why not? Certainly, so -- you could say that your expectation on all the criteria are likely to be a little bit stronger. Though it=s actually culled out very specifically in the details of the criteria for certain criteria, so.

You=re right. Because I think one of the questions is what=s the usefulness of the measure? What=s it been, what=s its continued use? What else has happened in the universe? The context that it exists in.

When process measures were all we had, then that=s what we had. But now that we=ve got many more outcome measures, the question, you know everything should really

prove itself of having value in the portfolio.

MEMBER SIPERSTEIN: But I guess to oversimplify, if you were to not endorse a message, would it be sending the wrong message to a certain group who may not be understanding the subtlety of some of this discussion?

CO-CHAIR FLEISHER: So it=s an interesting thing. Helen did say we have reserve status for some if it meets all the criteria. And the question that is a question -- CSAC has not -- it=s still wrestling with this question of when should process measures -- if that process will not drive the outcome and we have the outcome, but we can=t prove that it drives the outcome, when should we stop endorsing it or put it on reserve status?

So we don=t have an answer. And in fact this committee=s thoughts, one of the reasons to put someone from CSAC on a committee like this is to hear your thoughts. So the question becomes do you think if we didn=t

endorse, or do endorse this measure, that driving improvement through the measure, will lead to improved outcome?

If the answer is it would, that=s helpful for CSAC to say I think we should continue endorsing this measure. If it=s not, you shouldn=t endorse the measure. If you=re unsure, but you think it meets all the other criteria, that=s what we=re wrestling a lot with right now.

DR. JACOBS: I would just add that I believe that a lack of NQF endorsement for this message would -- and let me say that again, the lack of NQF endorsement of this measure would really send the wrong message.

Because this is something that is taught as standard of care. And the unintended consequences of not endorsing it I think could be potentially harmful.

MEMBER JARRETT: This is Mark. You know and I appreciate that. But I you know

having living with the world of how many things we all measure, I think that if something=s standard of care, making people measure it all the time is not necessarily the right answer culturally of where we want to go.

I think it comes with how you message it, that it=s not just something that we=re going to measure on a regular basis. We retire measures all the time like aspirin in the emergency room. But that doesn=t mean that we -- that aspirin in the emergency room for an acute MI doesn=t count.

So I think we have to be careful that we just don=t keep it for the sake of well that=s the standard of care and that=s the only way we keep the standard of care going.

MEMBER YATES: This is Dr. Yates. The -- I think part of this is that we kind of skipped over the first algorithm for process measure which is looking at the scientific evidence, the scientific review. And in fact

I=ve pulled up where the -- they=re not cited, but the actual level 1B, level 1C recommendations come from the Journal of Chest in 2013.

So that=s a -- that=s post -- that is status post the initiation of this measure. And from that, you gather at best they have moderate to low evidence to show anything. And there=s only six citations given.

But my argument is that surgery is something that is learned, or the experience in surgery is accretional and our experience in taking care of patients is accretional. And at this point in time, you=re never going to get level 1A data in a prospective randomized study asking thoracic surgeons not to get PFTs on their patients versus getting them.

You=d have to -- even trying to subselect the ideal population that doesn=t get the PFTs, is kind of playing guts ball with patient safety when they at least at 1B and 1C

level, the experts in the field feel that it=s an important thing to do based on moderate and low evidence and gave it a level 1 criteria.

And so going back to that evidence in terms of this process measure, I think it at least has moderate validity, and it=s never going to be great validity in the fact that I doubt that we=re going to see a level 1 trial that=s going to do anything that proves the questions that are being asked.

And I would -- and so quibbling over whether it=s standard of care, some things that are standard of care in surgery are learned from experiences, one being a surgeon at Vanderbilt said you know son you don=t have to learn about all your mistakes by doing them, you can read about a few of them.

And I think this is one of those that I don=t think we=re going to revisit in terms of getting level 1A evidence to show that it has value. But as such, it still remains something

that people shouldn=t be skipping.

It=s an end organ being taken out. And knowing what it does before you take it out, probably has some value. And anyway, that=s why I would say. I can see why they made it 1B and 1C despite the lack of a large amount of evidence.

Sorry to get on my soapbox.

MEMBER MOSS: So I agree with what Dr. Yates said about the evidence. But -- so the STS database is one of the most comprehensive and meaningful and useful databases in all of surgery.

And you folks with your collective expertise have decided that if we=re going to measure results of lung resection, we=re not going to do it with an outcome measure. And that there are those maybe aren=t ready for prime time, and this is best addressed with this process measure.

Could you help us understand how you

got to that, and why you chose to go down this road, and not an outcome measure?

DR. JACOBS: Well, we are in the process; we have risk-adjusted mortality and morbidity measures for lung resection and for esophageal resection. And we are in the process of developing composite measures to go with those.

MS. WINKLER: Larry, to answer your question, we do have and they are NQF-endorsed. You will notice under thoracic surgery in your portfolio there are at least two risk-adjusted outcome measures from STS for lung resection.

CHAIR FLEISHER: Other comments?

MEMBER KO: So, is our job on this Committee -- and I'm sorry, going back to this again and again -- to just look at the individual measure by itself? Or should we look at it, if there is a process, do we look at it in the context of out outcome? Or is that what CSAC does?

Because, you know, not all outcomes are the same type of -- yes, the hierarchy, I understand that. But like, for example, VTE, we would not have a VTE outcome measure and we would preferentially use process likely.

And if there is a process to outcome link, and it is just one big process to the outcome, that's different. When there's 100 different processes, that is going to link something to something like mortality.

It is a little difficult for us in this, or for me, to look at this just by itself, unless we know that the CSAC is going to do a good job of that.

MS. WINKLER: Actually, this is sort of the new responsibility for a standing committee. That is why I discussed the portfolio review this morning and have given you the list of measures. So, it is meant to be a reference for you to see what else is in the portfolio that may provide the

environmental context of each individual measure.

So, you are evaluating each measure against the criteria, but it is not in the absence of understanding the greater context. And so, it is absolutely in your purview to ask the question of, when we have outcome measures, do we need the process measure?

We are looking to your expertise on the various topic level, the clinical areas, to really help us understand that. And I will agree with you, it raises the bar for the challenge for the Committee. But, yes, not individually in isolation; we do want to see the greater context.

MEMBER KO: So, that is clear, but it is hard to vote on this without knowing --

MS. WINKLER: Oh, yes.

MEMBER KO: It is like you vote on a diver in the Olympics and nobody gets a good score in the beginning because you're always

waiting for later. And that is kind of the quandary we are in now.

MS. WINKLER: Right. Yes.

CHAIR FLEISHER: So, I would say we are still in the development stage of where we want to be, but that is why I asked the question. So, if you vote to endorse or approval -- and I am making this up as I speak, so Reva can comment -- then, if you have concerns or a lack of concern that you want CSAC to recognize and to debate in the context of looking at the overarching concept, then that should be brought up here. It shouldn't be deferred. We should inform CSAC of our thought processes, correct?

MS. WINKLER: Yes, but I would put it even a little bit more differently. Your recommendations are to the NQF membership large. So, it is not even just CSAC; it's everybody.

And so, your rationale and how you

have applied the criteria is going to be important for end-users and other audience members to really understand what NQF endorsement means. And so, that really is your responsibility here, to explain and to raise some of these issues. Yes, some of them are really challenging, no doubt about it.

But, as a standing committee, this is going to be your role to help us grapple with those. And it is not an easy one; it is not a slam-dunk. But it is absolutely on the table for you.

CHAIR FLEISHER: John?

MEMBER HANDY: I just wanted to comment that there are other outcome measures that are in the portfolio that aren't just mortality. So, the 14-day length of stay after lobectomy, because of all the bad things that can happen to you after lobectomy are infrequent enough, this is a composite measure that says that things aren't going well. So,

this is not just mortality; this is also morbidity capture.

CHAIR FLEISHER: So, any other comments before we vote?

(No response.)

Recognize that these comments will be captured in part of the report that will go to CSAC and the public, and there is public comment at the end of today. So, we may hear from the public also. STS may want to be prepared to address some of these questions, both during the public comment period as well as CSAC and the Board.

Did you have another comment?

DR. JACOBS: No.

CHAIR FLEISHER: No?

MR. SANCHEZ: Voting will now begin for overall suitability for endorsement. One is for yes; 2 is for no. The timer starts now.

(Vote.)

MR. LYZENGA: We still need one

more vote in the room, if you can click your vote again.

(Vote continues.)

Let's try one more time.

(Vote continues.)

There we go. Thank you.

MR. SANCHEZ: Eight for yes; 15 for no.

MR. LYZENGA: And I should note -- sorry -- that I believe 15, I think that falls right in our gray zone or just at the edge. We have got a new sort of status for when a measure falls between 40 and 60 percent of the Committee voting to recommend it. That is called, what we say is that consensus is not yet reached. And we will put that forward through the rest of the process in public comment, CSAC review, member voting, et cetera.

And I think we will reconvene after the public comment period and take another vote on it at that time after we receive public

comments on it. But it will go out for public comment with the status of consensus not yet reached.

CHAIR FLEISHER: And it is important to recognize that we will be reviewing the final report drafted by the staff. So, this conversation is critical, getting to, Cliff, your question. So, it will be important to make sure that your thoughts are actually in that report, so that both the public and the CSAC are comfortable that they understand the thought processes that led to that, though.

MR. LYZENGA: Dr. Gunner actually corrected me on my math. We are right above the 60 percent level. So, this will actually not be recommended for endorsement.

DR. JACOBS: Can I just put one more thing in the record? Yes, I would just like to have it documented that I think that the lack of endorsement of pulmonary function testing

prior to anatomic lung resection can have some unfortunate unintended consequences. I think that one should be really careful when one decides that NQF is going to make a statement that that is an issue that is not worth endorsing anymore.

MR. LYZENGA: Thanks, Dr. Jacobs.

And we are actually going to do a little bit of shuffling around on our agenda here. We are going to move Measure No. 0453 up to the front of this block of measures. Our developer representative from CMS is going to have to drop off at noon. So, we are going to allow their representatives from AUA and AUGS, who very kindly agreed to delay the review of their measure until we are done with 0453. So, we are going to move to that one at this time.

And I think we have representatives on the phone from CMS. Do we have you on the line?

MR. BRATZLER: Yes. This is Dale

Bratzler. I'm chairing another NQF call at noon.

(Laughter.)

MR. LYZENGA: Thanks, Dr. Bratzler.

Well, go ahead and give your introduction of the measure.

MR. BRATZLER: All right. Very briefly, this is a major urinary catheter removed on either postoperative day one or postoperative day two, surgery being day zero, in patients who have had surgery. It is part of the Surgical Care Improvement Project measure set, a measure that has been in place now for several years.

We implemented the measure after a variety of studies showed that urinary catheters were often left in patients postoperatively for prolonged periods of time. In a survey that we have done in the past, more than 50 percent of the patients had a urinary

catheter left in place for more than two days.

And really reviewed the literature, and a number of studies demonstrated that, frankly, the majority of time there wasn't a need for continued urinary catheterization. So, the measure has been in place and used by CMS. It is publicly reported as a part of the Hospital Inpatient Quality Reporting Program.

And I will be happy to answer any questions about the measure.

CHAIR FLEISHER: I will just disclose, like Fred, I am a member of the SCIP Technical Expert Panel and helped create some of these measures. So, I am going to recuse myself.

MR. LYZENGA: Dr. Erekson, I think you are the lead discussant on this one?

MEMBER EREKSON: Thank you.

My first question, just as an overall question because I am not as familiar with the SCIP, participation in the SCIP, what

is the burden for hospitals and participating in SCIP? Do they have to actually pay to do this? Because the SCIP data is a formalized chart review. It is not just billing data.

MR. BRATZLER: Right. So, this is a chart-based, chart-extraction-based performance metric. It is a part of the Hospital Inpatient Quality Reporting Program. There are a number of metrics that CMS -- it is a voluntary program, but hospitals that don't participate in the program lose part of their annual payment update. So, virtually very close to 100 percent of eligible hospitals in the United States participate in the Hospital Inpatient Quality Reporting Program because there are financial ties to Medicare payment.

MEMBER EREKSON: Thanks. That's helpful.

So, I would just echo what the developer, if we are moving on to the evidence here, what the developer was talking about in

terms of the quality of evidence.

My other question on the quality of evidence is, from the references they cited, the bacteria that develops is within two to ten days, but this measure is very dichotomous at the two-day period.

And this goes into some of the semantics we were discussing earlier about pulmonary function tests at twelve months, three months, or six months. If they could give us a little bit of insight into the two days?

MR. BRATZLER: Yes. So, it is well-known that the No. 1 risk factor for catheter-associated urinary tract infection is duration of catheterization. And when we actually first studied the use of urinary catheters in postoperative Medicare patients, we actually looked at the association with urinary tract infections in our study.

And what we found was two days was

the inflection point. Once the catheter was in for more than two days, the risk of urinary tract infection increased fairly dramatically. So, they are often not manifest for the first couple of days, but two days was the inflection point where the risk of infection goes up.

MR. LYZENGA: Any other comments or questions related to evidence?

MS. WINKLER: Just in terms of rating this evidence, in terms of what is presented in the evidence attachment, the quality of evidence, do we have a systematic review; are we looking at a guideline; do we have details on the quality, quantity, and consistency of the evidence, and do we have strong evidence that this process of care impacts or relates to patient outcomes? Those are the exigent issues we need to address and be sure everybody understands in evaluating the evidence for this measure.

MR. BRATZLER: I don't know if that

question was addressed to me or more a comment.

MS. WINKLER: Dale, if you will just hold off a minute, I was addressing it to the Committee.

MEMBER EREKSON: In my review of this measure, it does not seem that I have found a systematic review, although you do have very, very consistent guidelines. You also have the CDC guidelines on the catheter-acquired urinary tract infections. But I did not find a systematic review that encompasses all postoperative surgical patients. There are randomized trials or I believe randomized trials in orthopedic surgeries, in particular.

MS. WINKLER: And from the guidelines, do we have grading of the evidence?

MEMBER EREKSON: Let me get that for you.

MS. WINKLER: Andrew has pulled up the evidence attachment for this, and you can see the CDC guideline under 1a4.2, talking

about a Category IB recommendation that states that, "For operative patients who have an indication for an indwelling catheter, remove the catheter as soon as possible, preferably within 24 hours, unless there are appropriate indications for continued use."

So, this is the evidence that they are presenting to support this measure.

MR. LYZENGA: Any other comments or questions about evidence before we vote?

(No response.)

Seeing none, let's go ahead and vote on 1a.

CHAIR FLEISHER: Fred, are you abstaining or are you voting on this? You are part of the SCIP TAP still?

MR. LYZENGA: You can vote on this one, Dr. Grover.

MEMBER GROVER: It wasn't listed; that's all, but I am happy not to vote. That's no problem.

(Laughter.)

MR. SANCHEZ: Voting will now begin for Subcriterion 1a, evidence. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient evidence.

The timer starts now.

(Vote.)

MR. LYZENGA: Could I ask everybody to submit your vote again? We are still waiting on a couple.

(Vote continues.)

All right. We're good.

MR. SANCHEZ: We have 11 for high; 10 for moderate; 1 for low; zero for insufficient evidence.

MR. LYZENGA: So now, we can go ahead and move on to 1b. This is performance gap or opportunity for improvement.

MEMBER EREKSON: So, when you look at this measure, when it was initially proposed, there was over 50 percent of

postoperative surgical patients were not getting their catheters removed within two days of surgery.

As the developers have implemented this measure, they should a consistent performance, consistently greater performance in this measure. Across 2012 to 2013, we went from 96 percent compliance with this measure to 97.7 percent compliance with this measure. So, we are getting close to being at a topped-out status, but I think that there is still a gap there of patients that don't necessarily need those catheters in place.

CHAIR GUNNAR: So, I have a fundamental question. Just remind me. So, if they abstract the chart and no Foley catheter 48 hours post-surgery, patient still inpatient, meet the criteria.

Are there criteria that allow me to actually document in the medical record why I might maintain that would be justifiable

reasons why the Foley catheter should retain or should be still in place?

MR. BRATZLER: Yes, there are.

CHAIR GUNNAR: Okay.

MEMBER JARRETT: This is Mark.

I think it is based on NISN reporting and they do have exceptions for certain indications.

MEMBER SAIGAL: So, then, the 97 percent rate that we are seeing now does not include those exclusion people? Those are all people that should have had the measure being met?

MEMBER JARRETT: I believe so, yes. I believe that should be the denominator.

MEMBER SAIGAL: That seems pretty high to me in terms of it being topped-out.

MR. LYZENGA: Any other questions or comments before we vote?

MS. WINKLER: Yes, just I will put that in context. Dale, what is the current

rate of exclusion in this measure? How many patients get excluded?

MR. BRATZLER: So, I am actually looking. I have a file, and I am really struggling to find it at the moment. But I have a file where we have actually done a breakdown of the number of patients excluded.

So, remember, there are certain operations that are excluded. Urogenital operations are excluded from the measure. And then, we have tracked the actual performance on the metric of those that actually have the catheter removed versus those that have documentation of a reason to leave the catheter in.

The catheter removed is consistently improved. I believe about 17 to 18 percent of the cases end up in the numerator because there is a documented reason to leave the catheter in.

But I am looking for the file. I

just don't happen to have it in front of me, but give me a moment and I will break in when I find it.

MEMBER KO: I have a question. With the NQF, when measures have topped-out previously, at what percent compliance? Is there a ballpark? Is 97 in there?

MS. WINKLER: Yes, 97 is in there. I mean, usually, it runs around above 95. This is always a question that the committees struggle with, is how many, because how many extras can you improve on the margin? I mean, there is no set answer. There is no absolute threshold. It is a judgment call on your part.

MEMBER KO: Well, I will tell you, at UCLA we had a measure that was 97.8 percent, and that was read, and we severated on that. So, when we top them out, that is definitely a good thing.

MEMBER JARRETT: Yes, this is Mark.

You know, I would agree with you

because I keep going to all my hospitals and saying, if you are up 97-98 percent, I don't want you putting a thousand resources into getting that one patient per month because you are never going, you may never get there.

However, I think we have to look at it that this measure is almost part of a bundle in terms of trying to lower catheter-associated urinary tract infections throughout the whole hospital. And therefore, there may be some value in maintaining it at least for another year, as most hospitals across the country struggle to get the total number of catheter-associated urinary tract infections down. And that slippage on this one might have impact later on.

So, I would not see it something continued for the next five years, but I think since this bundle has been so, you know, is really on everybody's forefront right now, it may be something worth keeping.

CHAIR GUNNAR: Barry?

MEMBER MARKMAN: Yes, I think it is inherent in most of the CMS measures, because I reviewed one, that they are going to be in the 97-to-98 percentile anyway because it is like mandatory reporting.

But I think there is a tremendous amount of value to the CMS measures going forward because they lead to better outcomes. I think the collection of the data, even though the performance gap is low, is still important.

MR. BRATZLER: This is Dale.

I did find the distribution list. So, in a single quarter, out of about 238,000 cases, 17.6 percent of the cases were excluded because of a reason. So, that could be a urogenital operation or documentation of a reason to leave the catheter in place. So, it is about 17.5 percent.

CHAIR GUNNAR: So, the exclusion list is 17.5 percent, which supports the fact

that, for those who are in the measurable category, 97 percent is actually the weight from 50, where it started, now 97 percent?

MR. BRATZLER: Right.

CHAIR GUNNAR: Dr. Dutton?

MEMBER DUTTON: I wanted to pick up on Barry's point. If you are reporting the data to Medicare in order to get paid for it, why would you ever report data when you were out of compliance? So, I am more interested in knowing how many, of all of the eligible cases out there, are reported. Do you know that, Dale?

MR. BRATZLER: So, out of all of the eligible cases that are actually being reported, I don't know that. I mean, we think that reporting is fairly completely because, remember, there is validation of the reporting. So, hospitals are randomly selected for validation, abstraction of medical records.

The cases are picked to be in the

denominator based on ICD-9 principle procedure codes. So, if the patient has one of the operations that is in the denominator, then the case is eligible for selection for the performance measure.

So, we have looked at the validation results before, and we find the reporting is usually pretty good. Now, you know, I think that with any performance measure -- I don't care what the performance measure is -- there can be unintended consequences or gaming of the measure. And I think the one thing that we have been concerned about is that a clinician can document a reason to leave the catheter in. We don't try to judge the clinician on what that reason is, but we have been tightening up those criteria a bit.

But the percentage of that that have been excluded did go up from, 2009, it was at about 14 percent, and it has gone up to about 17.5 percent. So, I suspect some of that is

because of more documented reasons to leave the catheter in, whether those are clinically-appropriate or not.

CHAIR GUNNAR: So, just for process, since I think we are ready to vote, if I have got this correct, if we have a -- oh, I'm sorry, sir. Go right ahead.

We have another comment.

MEMBER SAWIN: Well, in Washington State catheter-associated UTI is a reportable data mark. So, what is the state of that nationally? And if that is really the end-point we are shooting for, why do we need a process measure rather than the outcome measure?

MR. BRATZLER: So, that is a good question. And so, in the CDC -- I'm sorry -- in the CMS Value-Based Purchasing Program and in the Hospital and Patient Quality -- well, in the Hospital and Patient Quality Reporting System, catheter-associated urinary tract infections

are reported only for patients in the intensive care unit. They are not for patients in general med-surg beds. So, it is using the NHSN system, but at least at this point it only includes ICU patients.

CHAIR GUNNAR: Barry?

MR. BRATZLER: And I apologize. I actually chair the other NQF Committee, and we are going to start in four minutes. So, I really need to call in.

I think there are some OFMQ staff that may be able to answer other questions on the call.

CHAIR GUNNAR: So, we are going to move on, but just to clarify, the performance gap, if we vote as a Committee that this is low at this point, okay, then we will make a decision about whether or not it should go in reserve status or not? Did I get that correct?

MS. WINKLER: Yes.

CHAIR GUNNAR: Okay.

MR. BRATZLER: Thanks. I will be back on your call later this afternoon.

MR. SANCHEZ: All right. Voting will now begin for Subcriterion 1b, performance gap. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient.

The timer starts now.

(Vote.)

MR. SANCHEZ: Can everybody just submit their vote again, just in case?

(Vote continues.)

We have 1 for high; 4 for moderate; 17 for low; zero for insufficient.

MS. WINKLER: Okay. This now prompts the question, because you have decided that the gap is insufficient to meet NQF's criteria. This is where the option of reserve status comes in.

So, before we even answer that question, just to tell you what would normally happen, the failure of this measure, this

subcriteria fails the measure. It does not go forward. Okay?

However, if you decide that this measure is potentially a candidate for reserve status, we would then use the evaluation. Because in order to qualify for reserve status, it has to hit all the criteria solidly. So, it has to meet all other criteria really good in order to meet the reserve status. As Helen mentioned, this should be an exception, not something you do frequently.

So, the question to you at this point is, because we sort of broached that question, do you think this is a potential candidate for a reserve status and you want to continue the evaluation to be able to get to that point or not?

We can do it by a show of hands. How many want to continue the evaluation for a potential reserve status?

(Show of hands.)

This is why we vote with the clickers. I can't do a half-hand.

(Laughter.)

One, two, three, four, five, six, seven, eight, nine, ten, eleven, twelve, I think. Yes, it looks like it is at least a basic majority. So, we will go ahead and continue the rest of the evaluation.

Priority?

MEMBER EREKSON: So, when you are looking at this measure for priority, it is looking at these catheter-associated UTIs. It is a process measure looking at preventing the outcome of catheter-associated UTIs and the potential consequences of those.

CHAIR GUNNAR: Any discussion?

(No response.)

Hearing none, vote.

MR. SANCHEZ: Voting will now begin for Subcriterion 1c, high priority. One is for high; 2 is for moderate; 3 is for low; 4 is for

insufficient.

The timer starts now.

(Vote.)

We have got 12 for high; 6 for moderate; 4 for low; zero for insufficient.

CHAIR GUNNAR: So, does that meet the criteria for moving on, Reva?

MS. WINKLER: Yes.

CHAIR GUNNAR: Okay, we will keep going.

MS. WINKLER: No, keep going. Reliability.

CHAIR GUNNAR: Reliability.

MEMBER EREKSON: So, CMS actually conducted quite extensive reliability testing, going through identifying all the hospitals that participate and, then, using a validation sample of about 1,000 hospitals within that.

And in terms of the exclusion criteria, which seems to be the topic that we are focused on the most, which is the 17 percent

of people not having a documented reason for keeping the catheter in place, their kappas were actually fairly excellent. They were at the 91 percent, which for kappas is fairly impressive.

MS. WINKLER: Are there any questions or discussions about the specifications? That is part of reliability also.

CHAIR GUNNAR: Dr. Temple?

MEMBER TEMPLE: So, I actually have a lot of issues with respect to this measure and this specific issue because I don't think the specifications are any good. I think that if you look at the spreadsheets -- and I didn't spend time looking at the GYN issues; I looked specifically at colorectal -- they are excluding patients who have hemorrhoidectomies, who have fulguration of warts, who have fistulotomies, fistulectomies. These are patients who should never have a Foley

catheter, and they are put into the exclusion criteria.

And if you go through the colorectal procedures, at least half of them should not be in the exclusion category. And so, with that, if you have got patients walking around with Foley catheters two days after these small perinanal procedures, it is not measuring what we want to be measuring.

And moreover, I also think that the exclusions, while they include these ICD-9 codes, they also include just a physician or LIP documenting that the Foley was kept in. So, it is really more of a documentation type of measure than it is actually truly measuring what we are really wanting, is getting catheters out of patients two days after the surgery for the appropriate cases.

So, I take issue with the specs of the measure.

MS. JOHNSON: This is Wanda from

OFMQ.

We are updating that table, that exclusion table, which is 5-16. We did recognize that we do have procedures in there that should not cause exclusions.

As far as it being a documentation measure, we do allow any physician, APN, PA documentation of a reason for allowing the Foley to remain in. We did try to keep it physician, APN, PA documentation.

MEMBER PITZEN: This is Collette Pitzen.

I have a comment related to the numerator specifications. In terms of the evidence that was presented, I am wondering if within postop day one or postop day two perhaps might be too long of a timeframe.

And a question: are patients included in the numerator postop day zero? And could there be a future consideration if this measure was to continue to be used to shorten

that timeframe, based on the idea of less than 24 hours?

MS. JOHNSON: So, that is something that we can bring up with the Technical Expert Panel to discuss. We do have quarterly Panel meetings to discuss the specifications. And so, this is something that we would move forward to the TEP.

MR. LYZENGA: All right. Seeing no other questions or comments, let's go ahead and vote on Subcriterion 2a, reliability.

MR. SANCHEZ: Voting will now begin for Subcriterion 2a, reliability. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient.

The timer starts now.

(Vote.)

CHAIR GUNNAR: Folks, vote again.

(Vote continues.)

Make sure the green light goes on. There you go. You've your 22.

MR. SANCHEZ: One for high; 13 for moderate; 8 for low; zero for insufficient.

MR. LYZENGA: So, I think that may now fall into our consensus not reaching --

MS. WINKLER: That is really 64 percent.

MR. LYZENGA: All right. That will pass the subcriterion then.

So, let's move on to validity.

MEMBER EREKSON: As this measure, so things to consider for the validity I think is the exclusion criteria. And the one thing that convinced me that this measure was still maybe doing some good is when we talked to our colleagues at CMS and asked them, "Are we just getting better at documenting why we are not taking the catheter out? Are we actually taking more catheters out?" And I believe that was the response that we got on the Workgroup call that said, "We are actually taking more catheters out."

And so, even though we have all this exclusion criteria, and absolutely, in the colorectal data I would defer to colorectal colleagues, but when I look at the GYN data, there is definitely procedures that don't need to be excluded. Laproscopic oophorectomy does not need to be an exclusion procedure. At least we are taking a lot more catheters out.

And then, if you look at the validity testing that CMS did perform, the critical datapoints were extremely good when they performed these 12 case audits at these selected hospitals. The only datapoint that didn't have a slightly low discordance is the patient participating in a clinical trial, and is that the reason why the catheter was not removed?

MEMBER MARKMAN: I am not sure if this comes under validity, but how many reinsertions of catheters? I mean, the question is, are they being pulled and, then,

how many are being reinserted because they had to be pulled within the 48 hours? I am just curious if they have data on that.

CHAIR GUNNAR: Well, if I understand it correctly, there must be documentation if the catheter in -- it is either the catheter is in place postop day two, yes or no, right? And if it is in, is there documentation for an exclusion that would allow that catheter to not be marked as failing to meet the metric?

MEMBER MARKMAN: Right, if you pull it between the 48 hours, and then, 24 hours later you are going to put it back in, will that -- yes, I'm just asking the developer.

CHAIR GUNNAR: That is the correct question asked during the validity portion of the --

MEMBER MARKMAN: Which is where we are at now.

(Laughter.)

CHAIR GUNNAR: Okay.

MS. JOHNSON: This is Wanda.

If they do remove the catheter within the first two days, and then, have to reinsert it, they will still answer yes. And so, it does show that the catheter was removed, and they will pass the measure. We do not collect data on whether they reinserted it, though. We just don't have that data collection.

CHAIR GUNNAR: Okay. Are we ready to vote?

Oh, Dr. Yates?

MEMBER YATES: To that point, it also probably does not collect the incidence of number of times somebody has straight-cathed. In orthopedic literature, if somebody decides to pull the catheter at the end of surgery that was there because it was a spinal or for volume, the incidence is straight-cathing does go up. There is debate over that, whether there is a

value or no value of that within the first 24-48 hours.

But not knowing the straight-cath incidence is also some, though, because there are cost issues in terms of nursing time and everything else doing that. So, that is sort of along the validity lines, but it has something to do with hospitals that are comfortable straight-cathing on a regular basis, and they have a higher level of taking the catheters out less than two days.

One other thing, though, is that 17 percent exclusion criteria for the surgeon or the physician or the nurse practitioner documenting the reason, I would be curious to know from CMS, do they know whether or not that that documentation is prospective to the catheter being left in or is it retrospective to the catheter being left in, and whether they have data as to whether the retrospective documentation has gone up. And likewise, do

they know for a fact that they are capturing all of the criteria that are used or all of the reporting, and have they looked to make sure that there is validity that all the documentation is captured in terms of keeping the catheter? Because some places may be getting very good at retroactively justifying the retention of the catheter.

MS. JOHNSON: The documentation must be present on postop day one or postop day two. So, they should not be going back in and adding it as a late entry.

MEMBER YATES: Has that been audited, though?

MS. JOHNSON: No, we have not collected that data to find out whether they are going back and adding late entry.

CHAIR GUNNAR: Okay. Dr. Markman, do you have another?

Sir? Yes.

MEMBER MOSS: I asked this question

on the Workgroup call. Is CMS suggesting any effort to track the reasons for leaving the catheter in over time? So, there could be a learning process to ultimately determine what is valid and what is not, and maybe develop a more sophisticated measure.

MS. JOHNSON: The problem with that is that we are going away from chart-abstracted measures. And this would have to be something that is collected in the EHR as a reason for allowing. It could be collected then if they have a discrete field. But that is very difficult for the 50 vendors across the United States to come to a compromise and determine what rationale they should include in their systems.

But the goal is to get to EHR-only measures. And so, we would not be able to collect reasons unless they put them in discrete fields in the EHR system.

CHAIR GUNNAR: Okay. Any other

discussion?

(No response.)

Thank you very much for that explanation.

I think we are ready to vote on the validity.

MR. SANCHEZ: Voting will now begin for Subcriterion 2b, validity. One is for; 2 is for moderate; 3 is for low; 4 is for insufficient.

The timer starts now.

(Vote.)

CHAIR GUNNAR: We're missing a vote. One more time.

(Vote continues.)

MR. SANCHEZ: Yes, 5 for high; 8 for moderate; 9 for low; zero for insufficient.

MS. WINKLER: So, this does not pass validity. That takes it out of the possible realm of any reserve status. And so, it is now failed on two criteria.

So, we can finish. We are done.

Public comment?

MR. LYZENGA: Operator, can you open the phones?

THE OPERATOR: Yes. If you would like to make a comment, please press *, then the number 1.

(Pause.)

There are no comments at this time.

MS. WINKLER: Anybody in the room? Anybody back there?

(No response.)

Okay. I just want to look at our agenda, folks. So, we are clearly not progressing along at a particularly speedy rate. That is not terribly unusual. Hopefully, these first few measures have you given an opportunity to discuss a lot of generic issues, but we all need to collectively work to focus and keep things moving. And while I know that all these conversations are fascinating,

we really do have to focus on the work that we need to do today.

So, what I am going to suggest is we are scheduled for lunch from 12:30 to 1:00. Lunch is here. The question is, do we want to do a working lunch?

CHAIR GUNNAR: How about 15 minutes?

MS. WINKLER: Yes, right, I was going to say maybe 15 minutes. We reconvene at 12:30. And then, we will focus-in on these three remaining GU measures. And then, we will move into the afternoon agenda. Does that work for everybody? Okay.

Reconvene at 12:30. Lunch is available back here.

(Whereupon, the foregoing matter went off the record at 12:16 p.m. and went back on the record at 12:33 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

12:33 p.m.

CHAIR FLEISHER: Okay, we are going to restart in about 30 seconds.

The first piece of business is that, frequently, when these committees get together, and it is very clear we were talking about the phone calls and how effective they were. Think about that during the day, and it may be worth talking to any of us up here about how to make those phone calls more effective, the Workgroup calls.

But I realize part of that is this group has never been together. And now, we have some idea of how we are approaching these measures.

The second thing is, frequently, we arrange for a dinner, which is on your own that a certain amount can go against our expense report, realizing that we are on sort of the government dime. So, therefore, there are

federal limits.

What is the amount?

MR. SANCHEZ: It is $36.

CHAIR FLEISHER: Okay. Right. So, we have a reservation made, and we would like to know how many would like to join the group. And we would just throw 30 credit cards onto the table, and they actually know that, right, that we will be individually paying?

So, what is the information that we have for tonight?

MR. SANCHEZ: So, the reservation is at this restaurant called Mio, which is right nearby on Vermont Avenue. It is probably a block away, just one block up, and then, a left. And it is right now for 7:00 p.m.

I could tell you what the cuisine is, if you want, but it is really good.

MR. LYZENGA: Should we get a quick just hand count? Who would like to join us for dinner?

(Show of hands.)

All right. Thank you.

MS. WINKLER: Okay. As we get started on this next group of measures, these are all GU measures that have come to us in a slightly-convoluted pathway. And so, we will be approaching them slightly differently.

Chris Saigal was the Co-Chair of the effort when we tried out something that was a pilot project, and we learned a lot. We are not doing it this way anymore. But where developers brought to us their measures, and the initial review was of the importance criteria, the evidence, the gap, and the priority.

The idea being, if they passed, then it would be worthwhile to spend the resources to go test the measure and finish the development. For those that didn't pass, perhaps they need to go rethink from the beginning.

And it was an attempt to find a process that would help support measure development and not use resources unnecessarily by testing everything that ultimately is not going to pass.

So, like I say, we're on another version and another approach, and we are not doing it this way anymore. But we have to finish up what was done.

So, these three measures were initially evaluated during that stage one process. They have met the importance criteria. So, we are not going to repeat that. All right?

So, we are going to jump directly to scientific acceptability, the reliability of the measure specs. So, for the lead discussants and the measure developers who will join us, you will need to give an introduction about what the measure is, so everybody knows what we are talking about. But we will not be

going through the 1a, 1b, 1c discussion or voting. Okay? That has already been done. They got to you by passing those.

If there are any questions about any of those, I am hoping maybe Chris will be able to fill in any blanks as necessary.

So, do we have our measure developers here or on the phone?

MR. MORGAN: I'm here. My name is Dan Morgan. And I was one of the measure developers, and I think there is one other measure developer on the phone as well.

MS. WINKLER: Okay.

MS. PULLIAM: And I'm Samantha Pulliam, and I am another of the measure developers for AUGS.

MS. WINKLER: Okay. So, great. We have got you on the phone. Good to know.

All right. So, the first measure we are going to start out was 2038, performing vaginal apical suspension at the time of

hysterectomy to address pelvic organ prolapse.

And who is the lead discussant?

MEMBER TEMPLE: I am.

MS. WINKLER: Oh, great. Larissa, are you okay, have finished swallowing?

MEMBER TEMPLE: Sure. I'm doing good. Thanks.

So, I will just give a very brief overview of the measure.

MS. WINKLER: Hold on a second. Let's let our developers make a couple of comments first.

MEMBER TEMPLE: Sure.

MS. WINKLER: I'm sorry.

MR. MORGAN: Okay. This is Dan Morgan again.

I think this first measure that we are discussing is the use of colpopexy at the time of hysterectomy for prolapse. This is an area that has a high impact and that the rate of re-operation is significantly higher among

those women who do not undergo colpopexy. That has been shown by multiple studies, as we went over the importance part of the application.

We went ahead with a testing for these patients, and we recruited patients from four different places around the country, and then, were able to retrospectively look at the experience and how frequently these were done, and then, to get some data to be able to speak to the validity of the measures. And that is what I understand we are going to talk about at this point.

Is there anything else that you would like me to speak to?

MS. WINKLER: Larissa?

MEMBER TEMPLE: You want me to start? Okay.

So, the essential premise is that with pexy we can decrease, after hysterectomy for prolapse, we decrease recurrence and we decrease the need for repeat surgery and the

morbidity of having recurrent prolapse in women. It is an increasing procedure and increasing morbidity, especially as our population demographics change.

So, speaking specifically towards the reliability of the instrument, if we look first at the specs, the specs of the instrument are quite reasonable. They provided the CPT codes for the numerator, which includes procedures, hysterectomies which include the various pexy procedures. And they accept any type of pexy. So, they haven't chosen one pexy procedure over another. It is just if you pexy at all, that is how I think of it.

And the denominator is hysterectomies for women who have prolapse. The prolapse codes are fine with ICD-9 codes and they seem pretty exhaustive. They exclude women who are having their THs for cancer and, also, for obliterative procedures.

And so, it is all electronic. It is

all data from billing. So, it is fairly straightforward in terms of the numerator and denominator.

Where I have issues -- and this is where the developer is going to be very useful -- is how they did their reliability and validity testing. So, they originally set out to identify the prevalence of colpopexy by getting four hospitals with 300 surgeons, and they identified 4,200 women who had had TH for prolapse between 2007 and 2011.

And they reported that the prolapse pexy rate was 74 percent. But, then, what they also discovered is, then, they wanted to go do a smaller chart review. Then, they discovered one of the four hospitals that they evaluated had used codes in a weird way. So, they excluded two-thirds of the patients. So, they went from 4,000 down to 1400 patients because one hospital was not using the right CPT codes.

The developer on our Workgroup call

sort of assured us that this was due to an institutional error, but not something that would be systematic. So, it was just one institution. It wasn't sort of something you would expect to see if you surveyed more groups.

They reported that, to test the reliability, they looked at three hospitals. So, then, one surgeon left. So, they had three surgeons to evaluate the reliability. Each surgeon evaluated 33 records to look at the frequency of the EMR being correct versus the operative report. And they found a kappa that was quite high. It is .92, so very high. So, the reliability of the measure is good.

They are using patient-level data. And so, I would say it is moderate data, but I think it meets the reliability criteria.

MS. WINKLER: Any comments from anybody else, comments or questions?

Fred?

MEMBER GROVER: Yes, just one

question. On the administrative data, then, that was the CPT codes just to identify who the patient population was, and then, the rest of the data was abstracted either through an EMR or clinical records? Is that it? Is that what I understand? Or am I wrong?

MEMBER TEMPLE: My understanding is it was all electronic; it was all billing records. And they went back to identify how accurate the billing records were by going back and reading the operative notes. But it is all --

MEMBER GROVER: So, it is verified?

MEMBER TEMPLE: -- verified, yes.

CHAIR FLEISHER: And just your comfort with the fact that 25 percent were incorrect and whether that is or is not generalizable and systematic?

MEMBER TEMPLE: I am troubled by it in the sense that coding hysterectomies with colpopexy shouldn't be that difficult, and it

was a huge change in the numerator or the denominator. The developers tell me that it was one hospital. It would have been nice to have seen them go out and check four more hospitals or three more hospitals. But that is up for discussion I think.

CHAIR FLEISHER: So, was that brought up on the call?

MEMBER TEMPLE: It was.

CHAIR FLEISHER: And their response was? And their response on the call, if they could make a comment?

MS. PULLIAM: Sure. So, we had the four hospitals, and there was one hospital with billing codes that were so dramatically different from others that it really triggered a rethinking of the way that hospital had coded, and has triggered a recoding and a readdressing of this within the hospital structure itself because it was so incorrect.

And so, our feeling was that that

was such an exception that we didn't think that it was something that would be likely to happen in other institutions.

To address the issue of going out and looking at other hospitals, I think that would have been an ideal. However, given the time constraints of this presentation, that wasn't feasible at this moment.

MEMBER GROVER: I guess this brings up -- I mean, that was one out of four hospitals. So, going forward, how are you going to be sure that your data is accurate, and what kind of an audit process are you going to have established, and so forth?

MS. PULLIAM: Well, I mean, I think our work -- and, Dan, perhaps you can speak to this more -- but our work has basically underscored how an audit process might be accomplished once this measure is in effect in terms of chart review and comparison of those bits to the actual billing codes.

MEMBER TEMPLE: When I looked at the whole measure, the group of measures, what I noted is that this is probably, of the measures there is, when I read other -- they talk about developing a pelvic floor registry database. This is the only measure where, theoretically, you wouldn't need a database because it is all captured electronically. But my understanding -- correct me if I am wrong -- is probably eventually the data will be pulled from that type of registry. But, for now, it is talking about electronic capture of billing codes.

I guess the real question to the group is, do we believe that it was just that one hospital that made the error in their coding and billing or do we think that it potentially we need to get more data? Because that is really the issue about the reliability, right?

MR. MORGAN: And as one of the developers -- this is Dan Morgan again -- we

have polled our members extensively, trying to look to see if in any way that this type of coding that this one hospital did was supportable. And it really did not use the codes in the correct way. Systematically, this hospital collected the procedure with the one code, and they are able to figure that out by having the operative note reviewed. But it just was not supportable and it is not an alternative way to go about capturing that. And that was something that was gleaned from asking many institutions across our own professional society.

MEMBER CIMA: This is a real issue when we look at this. The only reason I am aware of this is because we have been dealing with this when we look at our multiple other hospitals.

And especially in GYN surgery, it has become a real issue. There was a recent study by Manatt's that looked, they

prospectively looked at coding of Caesarian deliveries at 11 different hospitals. So, C-sections you would think are a very straightforward thing to code. And they found that only 66 percent of the cases actually had it coded correctly, and only 13.7 percent of the ICD-9 codes were assigned correctly. So, this is a big issue if you are going to be doing this just off these coding issues.

I know we have an obstetrician colleague here, but, I mean, when we start pulling into stuff like this, when we are going to be going completely off the coding, that becomes a huge issue.

CHAIR FLEISHER: Okay. Actually, unless you wanted to comment, I would ask, Larissa, do you have a proposal that, if this makes it through, of what should be provided before the next step?

MEMBER TEMPLE: You know, I really appreciate the cost and time that goes into

pulling all these records, but I think it really behooves them to probably do another sample of hospitals perhaps. I don't need 2400, 2,000 cases, but probably a random sample in three hospitals that shows the right coding would be reasonable, if they sort of went back to their dataset of like 100 cases per hospital, the three hospitals.

CHAIR FLEISHER: Can I get a response from the developer?

MR. MORGAN: I mean, the cost and the time associated with that, because it is completely unfunded times, and each of these hospitals, in order to request the data, it is a significant burden. I think we need to look at things in the future, but I am very concerned about our ability to recruit more hospitals and be able to have the resources to put forward for that.

CHAIR FLEISHER: So, the answer is no? I am just being very explicit.

MR. MORGAN: I would suggest no.

CHAIR FLEISHER: Okay. Thank you.

Barry, did you want to --

MEMBER MARKMAN: Was the coding done by the hospital or was it done by the individual surgeon? And the followup to that question is, you said there were some chart audits of the coding or there were not, of the ones that were reliable, not the ones that you had to exclude?

MEMBER TEMPLE: So, it looked like the charts, once they got rid of that hospital, it looked like when they did an audit of operative note versus the -- it was accurate and the agreement was very high. So, that looked good.

But the real problem is, as Bob talked about, you know, getting the codes right from the beginning.

CHAIR GUNNAR: So, a question for the developer: is it there, although not

explicitly stated, wouldn't it be an expectation of this measure that two things would happen if it was actually endorsed. One would be that the actual occurrence of the additional -- you know, ensuring that what you want is going to happen.

The other would be is that, in fact, you will get better coding; the country will code better with regard to this. So, the usefulness of historical data actually is not very -- it doesn't really have much relevance.

CHAIR FLEISHER: Collette?

CHAIR GUNNAR: It has a starting point, but this didn't exist before.

MEMBER PITZEN: I just have a question of the OB/GYN colleagues in the room. Are these CPT codes fairly straightforward or are they subject to bundling? Because a methodology that is going after CPT procedure codes can't be very reliable. So, I am just really curious if it straightforward or subject

to bundling, or what the issues were.

MEMBER LEVY: So, for this measure, the CPT codes are reliable; the ICD-9, not so much. And one of the issues is, of course, that there might be two or three reasons for doing a hysterectomy, some of which may be coded, some of which may not. And I think that is an issue.

The second thing is it has not been necessary to code accurately in OB/GYN for payment purposes. And so, hospitals have not really spent the resources to train folks to code these things well.

And I think to your point, when it becomes a measure, then they will train people to accurately code.

CHAIR FLEISHER: My only comment would be that this is a criteria that we are setting which may be relevant in lots of domains. So, when we look at reliability, if you are saying it is okay that 25 percent fail, and that the measure developer feels that that

was good enough and is not willing to look further -- I'm hearing some bias here -- then, the question is, is that the standard that is going to apply for any measure that is CPT-code-based, because they will get better over time?

MEMBER MARKMAN: The other issue is, have you prepared for the incoming change in the ICD-10 and the coding that will be -- I think they pushed it back another year, but it is going to change.

MR. MORGAN: As a developer, I can say, yes, we provided all the codes that would be a transition of the codes from ICD-9 to ICD-10.

One other thing I can say about the coding at the fourth institution that might be helpful is that, when we realized and we learned what codes they were using to capture colpopexy, it was remarkably accurate. It was as accurate as the other institutions. So, we

had a sensitivity of 82.4 percent. And their sensitivity, using the codes that they used systematically, but not correctly, was higher than that, actually.

CHAIR FLEISHER: Larissa, do you want to comment or end the discussion? No? No?

Any other comments?

(No response.)

Are we prepared to vote? Or are we going on to the next phase here? Yes.

MR. LYZENGA: Let's go ahead and vote on reliability.

MR. SANCHEZ: Voting will now begin for Subcriterion 2a, reliability. One is high; 2 is moderate; 3 is low; 4 is insufficient.

The timer starts now.

(Vote.)

MR. LYZENGA: We are still missing one. If you could enter your vote one more

time?

(Vote continues.)

MR. SANCHEZ: We have got 1 for high, 8 for moderate, 10 for low, and 4 for insufficient.

MR. LYZENGA: So, we need to sort of do the calculation here again, but I think that may either fall in our gray zone or fail.

Okay. So, we are going to move forward and evaluate the rest of the criteria.

MEMBER TEMPLE: So, moving on to validity, the validity testing was done with the same methods as the reliability. So, they use the same dataset to assess the sensitivity and specificity caused when they get a predictive value.

They report decent numbers. They break it down by procedure. But, again, the issue is that they only evaluate three out of the four hospitals.

And they found differences by

volume. So, they found different rates of colpopexy based on surgical volume, which is speaking to sort of face validity. But, again, I think that the issue speaks to the fundamental question of the exclusion of that one hospital.

CHAIR GUNNAR: Comments?

MEMBER LEVY: I just had a comment on some of the ICD-9/10 codes that they are using for validity. I think that, for example, rectocele and urethrocele would not be conditions that would require an apical suspension. And so, I just have an issue with some of the inclusion codes that they have got for this performance measure.

CHAIR GUNNAR: Barry?

MEMBER MARKMAN: How many procedures were there in the count, 300, or from two hospitals? Is that sufficient numbers?

MEMBER TEMPLE: The validity testing was done on 99 patients.

MEMBER MARKMAN: Ninety-nine

patients.

MEMBER TEMPLE: Yes, yes. Yes.

MS. WINKLER: Just in response to how many, I mean, again, NQF does not establish number threshold, minimums. I mean, the testing is to make the case that the measure does what it is expected to do. It gives you reliable and valid results. So, again, you're the audience.

MEMBER SAIGAL: Well, on the other hand, I would say that it is expensive to develop those measures, but a mandate on the part of the societies. So, I mean, what is the cost to develop them?

MR. MORGAN: The cost to build them I can speak to a little bit. I mean, we have had to spend both compensating people for getting the data and requesting it is somewhere between $20,000 and $25,000 for those four hospitals.

CHAIR GUNNAR: I don't think

that -- we know it is very expensive, and part of our job is, and NQF, in looking at the new process through their Kaizen process, was to facilitate, but we are setting the standards.

MEMBER CIMA: The point is that I know it is expensive to develop it, but we are bringing together a proposal that basically looked at 100 or 200 patients to a national committee that is going to be setting a standard for a national problem or a national issue. Yes, it is probably a real problem, but I am not sure we can actually say in scientific way, give scientific feasibility to the measure, whether or not on 99 patients or 100 patients, it is just I am not even sure how it got past the first hurdle. Well, no, you know, the first criteria 1, criteria 2, based on that sample size.

MS. WINKLER: That sample didn't play into it. The importance criteria is evidence, gap, priority.

MEMBER SAIGAL: I think the

evidence is around published literature on many more patients. This is specifically about the performance of the measure as described in a sample of patients to see how it performs. That is my understanding of it.

MEMBER CIMA: Okay.

CHAIR FLEISHER: Other comments?

MR. MORGAN: We did 99 patients where we did two abstracters doing reliability testing. We had 638 patients that we did explicit operative note review comparing that to the CPT codes. So, it really was much more than 100 patients. And that was a representative sample of the 4,238 that we looked at, which was a consecutive sample from four years of any surgery, any vaginal or any hysterectomy done for prolapse. So, we tried to be very systematic in the sample that we drew on. This was not a small sample from just a few hospitals.

CHAIR FLEISHER: Comments?

Barry?

MEMBER MARKMAN: What hospitals were they? Were they university settings or --

MR. MORGAN: We tried to get hospitals from different groups. So, we had University of Michigan partners, which is the Harvard system, and we had Geissinger Health System in Pennsylvania, and we had Southern California Kaiser.

MEMBER SAIGAL: And one of those institutions was not coding correctly?

(Laughter.)

MR. MORGAN: That's correct.

MEMBER SAIGAL: That seems really -- that is an issue then. I mean, if those institutions are having problems, then you have got to wonder if the mom-and-pop operations are doing it, I think, in the country. And 100 patients and you find a problem, that is an important signal I think.

CHAIR FLEISHER: Okay, I think we -- Barbara?

MEMBER LEVY: Well, I would just point out that Kaiser has just begun using CPT and ICD. I mean, they have no reason to code those correctly, and they could be using a totally internal system. I don't know the Kaiser system, but if that were the outlier, that would not surprise me. And so, maybe the developer can tell us if that was the outlier. Because if it was, that is not a surprise at all. They have no reason to use that coding system.

MR. MORGAN: You are correct, it was Kaiser, and that is why it counted for such a large proportion of our sample that we had to exclude.

As we have talked to Kaiser and tried to figure this out, that has been exactly their point to us.

CHAIR FLEISHER: Comment?

MEMBER LEVY: I mean, I would just

say that, overall, I would think using CPT coding for validity makes good sense and should be reproducible.

MEMBER TEMPLE: I guess it is just disappointing that if Kaiser is using a different system, that they chose to use that to do the testing. And I appreciate that it had high numbers. I appreciate it probably had a physician champion willing to do the work. But it just doesn't look great when it comes here.

And, you know, if it was validity testing, reliability testing on 99 people and you had perfect data, it is one thing. But when you have to throw out 25 percent of the data, and then you have it, it is a problem. So, it is unfortunate and it is a lot of work. It is a small society, and they really wanted to make a good measure. And it is disappointing, but I think that, to push this through, you either need a registry or you need to get some more data so we can test it.

CHAIR FLEISHER: Well, it sounds like we should vote first. That is your opinion, but you have evaluated the measure.

Amy, did you want to --

MEMBER MOYER: In looking through the validity testing, I am not necessarily seeing a place where you are making any kind of a reliability or a minimum recommendation at the individual surgeon level. Is there a place where you saw that this was a valid measure at that level?

MR. MORGAN: I'm sorry, I had a hard time hearing, if it was directed to the developer.

MEMBER MOYER: Sorry. I will get closer to the microphone.

In looking through the testing, I am not seeing anything at the individual surgeon level where you are talking about reliability or a minimum number of pieces, kind of how a meaningful measurement. And I was wondering

if I am just missing it or if there isn't any data on that.

MR. MORGAN: We have provided in 2b5.2 some of the ranges, the number at which a surgeon at the fifth percentile would do colpopexy, 21.4 percent of the time. And then, there is somebody was performing colpopexies more frequently, at the 95th percentile, was doing them 96.4 percent of the time. So, we were trying to get at that, that there was that gap across surgeons.

CHAIR FLEISHER: Are we ready to vote?

MR. SANCHEZ: Voting will now begin for Subcriteria 2b, validity. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient.

The timer starts now.

(Vote.)

Zero for high; 5 for moderate; 14 for low, and 4 for insufficient.

CHAIR FLEISHER: So, that measure will fail on the validity criterion.

Are there any comments from the Committee to the developer?

MS. WINKLER: Question to the Committee: if they were to bring you back some additional data, would you be open to revisiting the measure? Is it really just a lack of numbers that is problematic for you as opposed to the actual construct of the measure itself?

CHAIR FLEISHER: Can you comment, having looked at this?

MEMBER TEMPLE: I think I have said everything I need to say. I think they need more -- I would be very comfortable looking at this measure again with more data, accepting that the codes may not be as good as registry, but I think we could still take it. I think that, if we got that, we would even do it with billing data.

CHAIR FLEISHER: Collette?

MEMBER PITZEN: This is Collette.

Yes, I would just appreciate if the developer addressed the additional feedback about exclusions and that the exclusions be appropriate for the population.

CHAIR FLEISHER: All right.

MEMBER GROVER: A couple of things. I think it is really important for professional societies to be coming forward with procedures and with metrics to measure quality. And I think the evidence for this is excellent.

I think the area where you really -- this is complicated, and you can't really, I mean, the Kaisers across the country, I mean, a very high percentage of our patients, for example, in Colorado are Kaiser-insured. So, you can't really exclude them. I mean, to me, that would be another issue.

What you are trying to collect is really a fairly simple thing. I would

encourage you to think about establishing a clinical registry in your database. And even if the surgeon filled it out and you audited it, I mean, you want to know if the right procedure was captured and whether they did that technique. Two questions really.

So, that would be my advice, for whatever that is worth.

CHAIR FLEISHER: John, you looked like --

MEMBER HANDY: Yes, I think it is really, to me, this seems like an important clinical problem. This is a black box to me. So, I was reading it sort of for the very first time, and it is amazingly lack -- it penetrates to it seems like a foundational procedure with this particular type of problem.

So, I think that there is nothing fatally flawed with this particular proposal. It is just not supported well enough. So, I think that, in answer to Reva's question, it is

we would consider it again with just better support.

CHAIR FLEISHER: So, the positive thing for the developer is we are a standing committee for years at least, if not three, this group. So, that means many of us may be here for a while. So, for the developer's sake, should you bring this back, it would not be another group who would be seeing new problems. And I assume on the phone call you got feedback on the other issues that were of concern, but, hopefully, you got the feedback today.

Okay, next measure.

Thank you.

MR. LYZENGA: So, the next measure is 2052. This is an AUA measure.

Do we have a representative of the AUA? Oh, in the room here.

MR. DMOCHOWSKI: Yes, I am Roger Dmochowski.

So, would you like me to begin?

Hello?

CHAIR GUNNAR: Yes, please do.

MR. DMOCHOWSKI: Okay. Yes, so in brief discussion of this measure, this is a percentage of stress incontinence surgeries in women for which cystoscopy was used during the surgical procedure to reduce complications.

The definition is, the numerator, surgeries for which cystoscopy was used during a surgical procedure. The denominator is all stress incontinence surgeries done in female patients adult ages, age 18 or older. And the exclusion for the purpose of this measure was concomitant surgery for prolapse.

A brief history: in 2009, the AUA entered into a partnership with the American College of Obstetrics and Gynecology, under the PCDI's independent measure development process, to look at a variety of measures related to stress incontinence interventions.

In June 2010, at the American

Urologic Association Headquarters in Baltimore, a multistakeholder panel was convened. The stakeholders included representatives from gynecology, urology, geriatrics, family medicine, urogynecology, and nursing. And this group was brought forward to derive, based upon the AUA's evidence-based clinical practice guidelines, a measure set that could be used for women undergoing surgical interventions for stress incontinence.

In 2011, after deliberation, the panel finalized and voted approval for five measures. In June 2012, those five measures were submitted to the initial phase of NQF's GI/GU Pilot Project. Only the cystoscopy measure was approved to continue to the next phase in November of 2012.

In 2013 and 2014, utilizing a third party, testing was conducted of that measure set in four single specialty large practice

groups, and that is the data that you have.

And then, in March 2014 through present, this measure was submitted to the NQF's Surgery Call for Measures.

We feel that this measure is extremely important as a safety protective measure for women undergoing interventions for stress incontinence because of increasing reports and the significant literature support for the lack of recognition without cystoscopy of lower urinary tract injuries related to stress incontinence surgery which makes management of those complications postop very difficult in terms of reconstructive procedures. So, it is much easier to manage complications when they are recognized at the time of the procedure rather than at some point in the distant future after that procedure.

That is a brief summary.

CHAIR GUNNAR: Who is our discussant?

MEMBER SAIGAL: That is me.

So, in terms of reliability, as Dr. Dmochowski mentioned, they tested those measures in four sites, and they have a good sense of what it means to do that in terms of finding their patients. About 150,000 patients were part of these practices, urological practices, and about a third of the women in most practices have some form of stress incontinence. They got about 159 surgeries to evaluate.

And they sent out abstracters to look at the EMRs of these practices. There was 100 percent agreement between two raters about whether the measure was met. So, the kappa was one.

And they were able to abstract the data from two out of the three that could report, and the other groups felt that they could report them with some modification of the EMR.

There were no exclusions identified by the abstracters. It is pretty rare to have an exclusion for something like this.

CHAIR GUNNAR: So, your feeling is that the reliability is high?

MR. DMOCHOWSKI: I would say so.

CHAIR GUNNAR: Any other discussion?

(No response.)

Vote.

MR. SANCHEZ: Voting will now begin for Subcriteria 2.a, reliability. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient.

The timer starts now.

(Vote.)

CHAIR GUNNAR: We need one more.

(Vote continues.)

I think we've got it. Okay.

MR. SANCHEZ: Sixteen for high; 6 for moderate; zero for low, and zero for

insufficient.

CHAIR GUNNAR: Validity.

MEMBER SAIGAL: Okay. So, validity. So, we talked about the exclusions. There's really a few exclusions that really make sense.

One thing that is relevant here is that the CSAC, when it talked about approving this, wanted to remove the exclusion for other concomitant surgeries at the time of this measure, but this measure is very similar to the use of cystoscopy during procedures for prolapse correction and the measure developers elected to keep these as separate measures. So, the exclusions for this measure do exclude prolapse surgery, which was decided after public comment.

And the rationale there is that there is different levels of performance, the different specialties that perform these surgeries, and they should be measured

separately.

I think that is reasonable, and I think it can be looked at again after there is some time in the field with these measures to see if they are that different.

And there is no missing data, and so, I think that the validity is high.

CHAIR GUNNAR: Any other discussion?

(No response.)

Okay, I think we are ready to vote.

MR. SANCHEZ: Voting will now begin for Subcriterion 2b, validity. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient.

The timer starts now.

(Vote.)

Sixteen high; 7 moderate; zero low; zero insufficient.

MEMBER SAIGAL: Okay. So, feasibility. These are CPT codes, and,

ideally, they could be looked at during billing, but the procedure of cystoscopy is bundled with the sling surgery. So, that is not a possible thing. You don't want to start routinely billing for things that are bundled; you would be accused of fraud.

So, this would be either EMR extracted, which they have shown is possible in the EMRs that they evaluated. I think that it makes sense. It is a pretty straightforward procedure, cystoscopy, and, certainly, the surgery itself, the sling surgery is captured in the EMR pretty accurately.

There is also a possibility to use this as part of their registry which is being developed, the ACA registry, in the future.

So, I say that it is generated during care and it is an electronic source. So, I guess that makes it high.

MS. WINKLER: Question: Chris, in terms of the data source, even though it is

abstracted out of EHRs, I mean, we are talking about an EHR as a medical record. So, it could also be done in paper records. I mean, there is nothing obligatory about the EMR use. Because there is a distinction between that and a true eMeasure that is developed specifically and with certain kinds of specifications.

MEMBER SAIGAL: So, it is not specified as an eMeasure. So, that is an important distinction. It can be taken out of an EMR feasibly. That would require some work on the part of the EMR owner. Or it can be taken out of a chart paperwise, which is obviously much more labor-intensive.

CHAIR GUNNAR: Just as an aside question to the developer, do you know in the ICD-10 codes whether or not the distinction occurs between a sling with or without the cystoscopy?

MEMBER SAIGAL: I think that is a CPT code.

CHAIR GUNNAR: Well, I know it is a CPT code. What I am saying, when the ICD-10 -- in Europe there are no CPT codes, right?

MEMBER SAIGAL: Yes.

CHAIR GUNNAR: So, it is somewhat do the math downstream.

MEMBER SAIGAL: Uh-hum.

CHAIR GUNNAR: When ICD-10 gets so voluminous over --

MEMBER SAIGAL: Right.

CHAIR GUNNAR: It is just a question.

MEMBER SAIGAL: I don't know.

CHAIR GUNNAR: It is probably inappropriate for this moment.

Anybody else have a question?

Collette?

MEMBER PITZEN: Just a question for clarification. Then, there isn't a plan to use CPT codes to identify the numerator cases

because of the bundling issue?

MEMBER SAIGAL: Well, to the degree which they are captured in EMR using a CPT code during the process of care -- so, cystoscopy can be coded that way. That is how it would be captured.

CHAIR GUNNAR: So, again, your recommendation regarding your assessment of feasibility?

MEMBER SAIGAL: I think it's high. I mean, they showed that they could do it in these practices. There was just a moderate amount of work. It is unclear to me how much work it takes for each EMR to do this, but it is certainly feasible.

CHAIR GUNNAR: Any other discussion?

(No response.)

I think we're ready to vote.

I was talking about 10.

MEMBER YATES: Yes, but it will be

there, too, if it is there for ICD-9.

MR. SANCHEZ: Voting will now begin for Criteria 3, feasibility. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient.

The timer starts now.

(Vote.)

We have 7 for high; 16 for moderate; 1 for low; zero for insufficient.

CHAIR GUNNAR: It carries the day.

MEMBER SAIGAL: Okay. So, usability and use. This is proposed to be used as a PQRS measure and, then, within the ACA registry as an internal measure within AUA.

And we don't have any improvements. We didn't talk about a gap, but there is definitely a gap. So, even the highest estimates in this data are 81 percent, but in literature it is lower in terms of performance of this procedure. So, we could see improvement with measurement.

And I am not sure -- the unintended negative consequences are basically maybe a small increase for infection, but it hasn't been documented anywhere. So, I think it is a pretty small chance of an unintended negative consequence.

So, I would say that it is high in terms of potential accountability.

CHAIR GUNNAR: Any discussion?

MEMBER MOYER: I would just echo the previous comment, that this is a really low-cost, really low-harm way of identifying something that could have really potentially very catastrophic and life-altering impacts. And so, I would agree with the usability, that it is something that is very easily done with very low risk.

CHAIR GUNNAR: Any other?

Oh, Amy?

MEMBER MOYER: I apologize if this isn't the right point, but I am curious, from

the measure developers, and looking at reduction of complications at the start of the measure, is there a plan for a measure of the incidence of complications at some point, an outcome measure?

MS. POPE: I think we would be open to that. We just wanted to see this measure through initially.

CHAIR GUNNAR: So, any other discussion?

(No response.)

We are ready to vote.

MR. SANCHEZ: Voting will now begin for Criterion 4, usability and use. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient information.

The timer starts now.

(Vote.)

CHAIR GUNNAR: There we go. You've got it.

MR. SANCHEZ: Fifteen for high; 9

for moderate; zero for low, and zero for insufficient information.

CHAIR GUNNAR: Any further discussion?

(No response.)

I think we are ready to vote.

So, does the measurement meet NQF criteria for endorsement?

MR. SANCHEZ: One is for yes; 2 is for no.

The voting timer starts now.

(Vote)

Twenty-four yes; zero no.

CHAIR GUNNAR: So, the recommendation is passed for this to go on as a recommendation for endorsement.

The next measure.

MS. WINKLER: Next is Measure 2063.

CHAIR FLEISHER: Is the developer in the room or on the phone? Is the developer on the phone?

MR. MORGAN: Yes. Dan Morgan here again for AUGS.

CHAIR FLEISHER: Great. Do you want to give us an overview of the measure?

MR. MORGAN: This measure is the use of cystoscopy at the time of hysterectomy for prolapse. Unrecognized lower urinary tract injury is a significant morbidity for patients in that it will lead to re-operations, readmissions, and significant cost, as well as suffering for the patient.

So, we targeted this as a safety patient measure that would allow us to try to recognize that injury or encourage providers and surgeons to recognize the injury at the time of the initial event, and that way, be able to repair the event at that time.

There have been studies that have shown that, if you do cystoscopy and recognize injury, that there is a significant cost savings. Per case, for unrecognized injury,

the cost is estimated to be about $54,000. And when looking at cost-effective analyses, if the injury rate is greater than 2 percent, then the use of cystoscopy will be cost-effective.

With prolapse surgery, there have been several consecutive cohorts of patients involving, grouped together, 2,000 patients, but the risk of injury to the ureter or bladder is about 5 percent. So, we are several-fold over that need.

So, we targeted looking to see how frequently cystoscopy was used in this cohort. And then, we would like to see this go forward as a measure for patient safety, and we think we would decrease the likelihood that somebody would leave the operating room with an unrecognized injury.

We think that it is especially germane, as we talk about these measures separate, that the incontinent surgeries tend to result in bladder injury; whereas, the lower

urinary tract injuries that happen with prolapse surgery are much more often those related to the ureter. So, we are looking for different injuries, and that is one of the reasons why we have tried to describe and give importance to these being separate measures.

We hope to be able to eventually get a Category II code that would allow us to do this through CPT codes, but at this point we would need NQF endorsement to do that. And we have the same issue about bundling of the CPT codes for cystoscopy with these procedures for hysterectomy for prolapse.

And I thank you for the opportunity to present it.

CHAIR FLEISHER: Suzanne? Barbara?

MEMBER LEVY: Yes, I am the discussant.

So, my understanding is we have already passed the importance criteria. So,

we are not going to go through those.

In terms of reliability, they looked at 638 patients, and the kappa was huge. So, I don't think we have an issue at all with reliability.

This will require chart review because, as he said, the cysto is always bundled with these codes unless there is a separate diagnostic reason for doing cystoscopy. So, it will be quite difficult to pull this out of administrative data.

MEMBER CIMA: For clarification for my colleagues in urology or gynecology, when you do a cystoscopy if you are looking for a ureter injury, how often do you detect that on cystoscopy, especially if it is somewhat of a cursory cystoscopy? I mean, in the previous one, they talk about only 30 percent being picked up, of bladder perforations being picked up on cystoscopy. Now I am wondering about, if we are trying to do this for ureter injuries,

I have never heard cystoscopy as being a gold standard for ureter injury.

MEMBER LEVY: So, what typically we do is inject indigo carmine or some sort of dye and, then, watch for bluey flux from the ureter to detect whether or not the ureter has been kinked or obstructed by the surgery. So, that is typically what happens.

It is relatively reliable. Eighty-nine percent or so of the time you pick up the ureter injury. If it is a thermal injury at the time of hysterectomy using some sort of thermal device, electrosurgical device, that is going to be delayed injury, and you won't pick that up. But the data are around 85 to 89 percent.

MEMBER YATES: It is no longer a moot point, but there is an ICD-9 code and there is an ICD-10 code for the act of a cystoscopy.

MEMBER LEVY: Yes, but they are bundled into the primary code. So, they are

never coded.

MEMBER YATES: They're --

MEMBER LEVY: They are not used for billing purposes.

MEMBER YATES: They're not recorded.

MEMBER LEVY: They're not recorded.

MEMBER YATES: Well, what I am saying is that they may be bundled under the CPT, but are they separate in the ICD?

MEMBER LEVY: No, no, no. They're not -- they are bundled for the purposes of payment. There are separate codes in CPT and in ICD-9 for cystoscopy.

MEMBER YATES: Right.

MEMBER LEVY: When it is done alone, it is coded. When it is done in conjunction with another ICD-9 procedure code --

MEMBER YATES: Nobody --

MEMBER LEVY: -- it would not be coded separately.

MEMBER YATES: Nobody bothers or is --

MEMBER LEVY: Correct. Well, it is actually considered fraudulent if you are doing it for billing purposes.

MEMBER YATES: For the hospital?

MEMBER LEVY: Correct.

MEMBER YATES: Okay.

CHAIR GUNNAR: No, the question I think earlier was, so the superpubic sling procedures as an ICD-10 code in the future, is there one with or without cystoscopy? That was the question. And we don't need to get off-track with that.

MEMBER YATES: Okay.

CHAIR GUNNAR: I am looking, actually, to see if that is the case, but we will --

MEMBER LEVY: My understanding is

that there is not. It is inherent in the procedure. It is one of the steps of the sling procedure. So, it is not included as a separate --

CHAIR FLEISHER: So, let me just get NQF staff -- when we go to ICD-10, it needs to be -- it doesn't come back necessarily to this Committee? It just needs to go through a process internal to NQF, and that is not relevant -- I mean, it is a different process? We should focus on the reliability of the measure as specified, correct?

DR. BURSTIN: It may change materially from ICD-9 to ICD-10. We will make that determination. We may come back to you with guidance as needed. But, in general, we have already seen for all the new eMeasures, for example, they are all using ICD-10. But, as issues come up with that translation, we will certainly call on you for input.

MEMBER LEVY: But for this measure

at this point, it is chart review or looking at the electronic medical record and pulling out from the operative report that cystoscopy was performed.

CHAIR FLEISHER: And that is what we should focus on, the reliability of that --

MEMBER LEVY: Yes.

CHAIR FLEISHER: -- in this Committee?

Chris?

MEMBER SAIGAL: I would just add, I think what would happen is you would do a cysto, and if you see efflux of urine out of the UO and you were concerned, then you would give the indigo carmine. And you usually can see efflux of urine if someone is well-hydrated.

MEMBER TEMPLE: I just want to make one quick comment. When we looked at this measure, they used the same dataset as for the prolapse. And so, they went from the 4,000 to the 600 to do the chart review. Do you want to

comment on why you think this is better reliability than the previous measure?

MEMBER LEVY: Well, I think that this one will require chart review to specify it; whereas, the other one didn't. So, I think the fact that it requires chart review gives it the reliability and validity because, whether it is Kaiser or anybody else, we are not relying on the numbers; we are relying on the operative report itself.

CHAIR FLEISHER: So, that gets back to the initial comment that we are reviewing the measure as specified. We are not changing the measure at this point. We can ask questions of the developer, correct? And the developer can choose to change the measure, but they own the measure at this point.

So, thank you, Barbara.

Chris, did you have another comment? Or no?

Are we ready to vote?

MR. SANCHEZ: Voting will now begin for Subcriterion 2a, reliability. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient.

The timer starts now.

(Vote.)

We have 10 for high; 12 for moderate; zero for low; zero for insufficient.

MEMBER LEVY: So, similarly for the validity, they found cystoscopy was performed in 84.5 percent and detected a bladder or ureteral injury in 5.8 percent. So, that fits with the literature, and the large quantity of literate, that would state that we are looking at a valid measure. We are looking at something with a gap. We are looking at something that will distinguish care.

And again, we have the same issue with starting with 4,000 and coming down to 638. But, since we are looking at this with abstracted chart review or looking at the

electronic medical record to specify cystoscopy, I don't think it is an issue for this.

CHAIR FLEISHER: Great.

Comments?

(No response.)

If none, let's vote.

MR. SANCHEZ: Voting will now begin for Subcriterion 2b, validity. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient.

The voting timer starts now.

(Vote.)

Fifteen for high; 7 for moderate; zero for low; zero for insufficient.

CHAIR FLEISHER: Next.

MEMBER LEVY: So, in terms of feasibility, again, I think this is certainly feasible. It would be nicer to have it in a registry or something that is easier, but chart abstraction that will have to be done for now.

If they can go get -- my understanding was that CPT II are actually going away. I hope that is not the case, but if it is the case, that is going to be a problem. And specifying this in a registry format would be much easier to use. Nevertheless, I do think it is feasible as specified.

CHAIR FLEISHER: Comments?

(No response.)

Hearing none, let's vote.

MR. SANCHEZ: Voting will now begin for Criteria 3, feasibility. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient.

The voting timer starts now.

(Vote.)

Six for high; 15 for moderate; 2 for low; zero for insufficient.

CHAIR FLEISHER: You can go ahead.

MEMBER LEVY: And again, for usability, I think that for public reporting,

for internal quality assessment, this will be a useful measure. It will be something that will help to distinguish care.

And I don't have any further comments on that.

CHAIR FLEISHER: So, that was your usability comments?

MEMBER LEVY: Correct.

CHAIR FLEISHER: Okay. Anything further?

(No response.)

MR. LYZENGA: Okay, let's vote.

MR. SANCHEZ: Voting will now begin for Criterion 4, usability and use. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient.

The voting timer starts now.

(Vote.)

MR. LYZENGA: I think we are still waiting on a couple of votes. If you want to try to cast your vote again, everybody?

(Vote continues.)

There we go. Thank you.

MR. SANCHEZ: Twelve for high; 11 for moderate; zero for low; zero for insufficient information.

CHAIR FLEISHER: Okay. Any comments before we vote on endorsement?

(No response.)

Hearing none --

MR. SANCHEZ: Voting will now begin for overall suitability for endorsement. One is for yes; two is for no.

The voting timer starts now.

(Vote.)

Twenty-three yes; zero no.

CHAIR FLEISHER: Great. Thank you.

I think what is also nice is in the documents will be some of the comments regarding where this measure may need to go three years from now. So, I think those,

without spending too much time, those are useful directions for the developer.

Where do we go next?

MS. WINKLER: We will begin the afternoon agenda, which, hopefully, again, I would ask everybody to really be time-sensitive, so we can get through this, or we truly will miss dinner tonight.

So, the next measure is 0178, improvement in status of surgical wounds. This is a measure from CMS.

Do we have the developer on the line?

MS. DEITZ: Yes, this is Deborah Deitz from Abt Associates.

MS. WINKLER: Great.

MS. DEITZ: And my colleagues from Acumen are also here.

CMS sends their regrets. They were on, but they had to drop off a few minutes ago.

MS. WINKLER: Okay. Thanks,

Deborah.

MS. DEITZ: So, shall I give a brief intro?

MS. WINKLER: Please.

MS. DEITZ: Okay. This is an outcome measure that reports on the improvement in the status of surgical wounds in the home health setting. So, specifically, the percent of episodes of home health during which the patient has a better status of surgical wounds at discharge than they did at that they entered home health.

The measure is calculated based on data obtained from the Home Health Outcome and Assessment Information Set, the OASIS C, which is a core standard assessment dataset that home health agencies collect as part of their own comprehensive patient assessment.

And information on the healing status of surgical wounds is used to calculate this measure. That information is recorded in

the OASIS items as part of the normal clinical practice.

CMS currently publicly reports this outcome measure for Medicare and Medicaid patients on their Home Health Compare website, and they have been doing that for a number of years, I think since 2002. Consumer can, then, review and compare agency performance on this and other home health measures.

According to the 2013 data that we just analyzed, about 25 percent of all the home health patients had a surgical wound, and about 13 percent of patients showed an improvement in their surgical wound during their home health episode.

Home-based surgical wound care follows very well-known principles, tenets, including keeping the wound clean and dry, avoiding activities that cause skin torsion and tension near the wound, lifting restrictions, nutritional intake, and patient education on

signs and symptoms of wound issues, like deterioration or infection, that need to be reported.

Between the July 2010, between the 2010-2011 measurement period and the 2012-2013 measurement period, at the agency level the mean risk-adjusted performance rate for this measure increased from 86.2 percent to 87.9 percent.

Basically, because of the high prevalence of surgical wounds among home health patients and because there are agency practices that are associated with high-quality care, CMS thinks it is really important to continue publicly reporting this measure.

CHAIR GUNNAR: So, thank you very much.

Who is the discussant?

(No response.)

Andrew is not here. Who is the secondary now?

We have to put on our home healthcare hats.

MS. WINKLER: Yes. Who else was on the Workgroup and listened to the initial conversation and might be willing to step up and talk about this measure?

MR. LYZENGA: There are a few comments from the Workgroup here on the screen that we summarized in the document. I could read those off, if you like.

CHAIR GUNNAR: So, the Workgroup summary is:

Measure demonstrates room for improvement -- national average at 89 percent -- as well as demonstrating differences in racial disparities.

Although evidence to support this measure, morbidity and mortality associated with surgical site infections and complications is provided, it was difficult to understand the types of wounds to be assessed.

Further, it would be useful to understand the frequency and duration of caregiver visits and how that correlated with wound improvement, as well as how improvement correlated with office visits.

So, the developer, any response to the second Request for Information, understanding the types of wounds that were assessed and the frequency and duration of caregiver visits in relationship to wound improvement?

MS. DEITZ: Yes, I think we addressed those questions at the time of that call. And the response was mostly centered around the fact that physicians are the ones that write the orders for the care of the surgical wound, and the nursing staff of the home health agency carry out those orders and, also, use their nursing judgment to determine whether the physician needs to be informed of any changes necessary for continued

improvement of the wound.

You know, the type of wounds that are being cared for are --

CHAIR FLEISHER: Can I ask, we are questioning whether we can do justice to you as the developer. Would you be willing to call in tomorrow and we can assign somebody else overnight to review this, so we can appropriately evaluate that? Would that be acceptable to you?

MS. DEITZ: I think that makes sense.

CHAIR FLEISHER: Okay. So, we are going to table this until tomorrow. Reva will get back to you, or somebody, right?

CHAIR GUNNAR: Anybody want to volunteer for -- Dr. Markman will volunteer.

CHAIR FLEISHER: Okay.

CHAIR GUNNAR: Thank you, Barry.

CHAIR FLEISHER: Yes, we apologize that the Committee member was not here, but we

think that is best for you, if we wait until tomorrow.

MS. DEITZ: Okay, very good. And you will be letting us know at what time?

CHAIR FLEISHER: We will be emailing you and ask when you are available. We will work around your timing for your assistance in this.

MS. DEITZ: All right. We will check our emails. Thank you.

CHAIR FLEISHER: Okay. And maybe CMS can join us.

MS. DEITZ: Okay. Perfect.

CHAIR FLEISHER: Yes.

It is all yours. I am recusing myself.

MEMBER DUTTON: Yes, I am actually putting on my developer hat. So, I will recuse myself from any voting on this, but would be happy to address questions from the point of view of the developer.

And Matt Popovich and Maureen Amos are also here from the ASA.

CHAIR GUNNAR: I'm sorry, who from the developer? Rick, will you --

CHAIR FLEISHER: Actually, it can't be Rick. It cannot be Rick.

MEMBER DUTTON: Okay.

CHAIR FLEISHER: It cannot be Rick. So, Matt and Maureen.

CHAIR GUNNAR: Who would like to begin?

CHAIR GUNNAR: Well, he can't.

CHAIR FLEISHER: He can't.

CHAIR GUNNAR: Just make sure your microphone -- oh, there you go.

MR. POPOVICH: The measure is perioperative temperature management. The measure has been slightly altered from 36 degrees Centigrade to 35.5 degrees Centigrade.

It went through a significant period of review by ASA members, as well as

members over the years. It has been supported by the ASA House of Delegates.

And we are more than happy to answer any questions that you may have related to changes to the measure.

CHAIR GUNNAR: Who is the lead discussant? Yes?

MEMBER OLSEN: So, you really weren't very specific on the scientific background. I mean, you gave a lot of global things in your review, but not the specific documents about why the change in the measure.

MR. POPOVICH: Well, the measure was changed to remove the process aspect of the measure, which was to use a warming device, and instead, it focuses more on the temperature of the measure. And so, our members felt that that was more important to look at the outcome aspect of the measure rather than the process part.

MEMBER OLSEN: Certainly, there is

a lot of literature that has been published, numerous articles on temperature control.

MS. WINKLER: All right. Why don't we go ahead and talk through the criteria? So, the first one is evidence. Summarize the evidence, and then, your evaluation of the rating of it.

MEMBER OLSEN: I think that the evidence is well-documented in the medical literature, probably some 40 or 50 different papers, a lot of them randomized control trials documenting the impact of temperature control during the post-anesthesia recovery period.

CHAIR GUNNAR: So, your recommendation would be that the evidence is --

MEMBER OLSEN: The evidence is supportive.

CHAIR GUNNAR: -- high?

MEMBER OLSEN: It is high.

MS. WINKLER: I think we need to go into it in just a tad more detail. So, in terms

of the criteria, in order to evaluate it, we need to know whether we are looking at a systematic review, whether we have information on the quality, quantity, and consistency. And since many of these are based on guideline recommendations, at a minimum, we need to at least establish with the grading for that guideline recommendation and the evidence that supports it is. So, we do really want to have a brief conversation around that, please.

MEMBER OLSEN: I believe the criteria was 1c throughout or in that general category, for at least using the grade category.

MS. WINKLER: Okay, and what does one see, actually, indicate? Scroll down a bit on the document. You will see they usually put the grading. Yes, there it is.

So, you have got two levels of the recommendation. One is class, which is the level of the recommendation, but the actual

level of evidence at "c" means "evidence from case study, standard of care, or expert opinion involving very limited populations".

So, how strong is the evidence for this measure?

MEMBER OLSEN: I thought the evidence was strong for this measure. They did provide some data from the NACOR registry, at least on the compliance with the process measure.

Although only about 25 percent of anesthesiologists are tied in electronically in the post-anesthesia recovery period, probably the compliance rate in that area is around 97 percent.

CHAIR GUNNAR: Any other discussion?

MEMBER YATES: Yes, we are talking 1c data. Are there any prospective randomized trials of patients that have been allowed to get cold or warm? I'm not being facetious, but we

are really talking about 1c data, which would be low-level evidence, but thought to have reasonably-high importance to the reviewers. There is at least one paper out there that talks about warmer patients coming out with cognitive difficulties.

MEMBER TEMPLE: Collette?

MEMBER PITZEN: Hi. Collette Pitzen.

And I have a clarifying question. Has the numerator changed from the application that we are looking at? Because this was submitted as a process measure that could be either active warming techniques are used or maintenance of body temperature. As I am understanding from your introduction, it is now strictly the outcome of temperature.

MR. POPOVICH: Yes, the numerator has been changed, and there is a CPT code that is now associated with the outcome aspect of the measure, that has removed the process part of

the measure.

MS. WINKLER: So, just to clarify what Collette asked you, the information that this Committee has to evaluate is not accurate, is not up-to-date compared to the current --

MR. POPOVICH: So, the evidence in the NACOR, as well as the CMS 5 percent files, is based upon how it is currently coded and has been currently coded in the past. The changing of the temperature of just looking at 35.5 degrees Centigrade is what we have proposed to change this measure to.

The evidence and the study is conducted at 35.5 percent or 35.5 degrees Centigrade. There is evidence in studies that were conducted at the end of the evidence chapter. I believe we put that in the eighth section of the evidence. 1a8.2 does have studies that look at 35.5 degrees Centigrade.

CHAIR GUNNAR: So, Dr. Fleisher has informed me that he, in fact, is the lead author

on the randomized control --

CHAIR FLEISHER: Second author.

CHAIR GUNNAR: -- second author on the randomized controlled trial in cardiac surgical patients. Non-cardiac?

CHAIR FLEISHER: Surgical patients with cardiac outcomes.

CHAIR GUNNAR: Non-cardiac surgical patients with cardiac outcomes. So, the point being is that level C evidence is not correct. There is far better evidence than level C evidence.

CHAIR FLEISHER: Absolutely.

CHAIR GUNNAR: As presented, though.

Right now, what we are working off of is what we have in front of us.

Dr. Grover?

MEMBER GROVER: Just a quick question. In your anesthesia literature or OR literature, what is the tipping point in terms

of temperature, evidence-based? I mean, did you have a reason for picking the 35.5 or 36? That is all I'm asking.

CHAIR GUNNAR: I think we are a little hamstrung here because the people who are actually representing the developers are muted in a way. So, I don't know how you want to --

MEMBER DUTTON: I would be happy to answer science questions.

DR. BURSTIN: I would say, if you are speaking purely to the science behind this, that is fine. But you cannot speak to the measure, how it is constructed, or any of those issues as the developer.

MEMBER DUTTON: All right. As I understand the science question, it was, is there an absolute cutoff temperature where outcomes change? And the answer is, no, it is a continuum. The colder you get, the higher the incidence of infectious complications,

MIs, shivering, et cetera.

MEMBER KO: And what was the science for 36 versus 35.5, the science from 36 down to 35.5? And what do you gain from how much more compliance is here with 35.5 versus 36? What is the tradeoff?

MEMBER DUTTON: I don't want to get myself in trouble, but our consideration in the measure is the scientific evidence. And the 20 papers in 1a8.2 more strongly support 35.5 as a cutoff, if you are going to pick an arbitrary number.

CHAIR FLEISHER: Yes, sir?

MEMBER YATES: Just in terms of evidence, the people that are muted on this, can they just at least point us to a clinical guideline from the anesthesia professional groups or, say, Cochrane database review? So, a high level of a --

CHAIR FLEISHER: The American College of Cardiology Foundation, the American

Heart Association guidelines on perioperative cardiovascular care published in 2006, updated shortly thereafter in 2009, Fleisher first author, recommends temperature monitoring.

The randomized control trial of perioperative maintenance of normothermia with cardiac outcomes in 300 patients, abdominal, thoracic, or vascular surgery patients, were randomized to a hypothermic group, 35.4, which is where they got to, versus a normothermic group, which was warming, active warming versus passive warming. It showed a risk reduction, 55 percent risk reduction in cardiac events. PubMed ID No. 9087467.

Nothing further.

CHAIR GUNNAR: Dr. Moss?

MEMBER MOSS: I was happy to see that you did not exclude children in this --

MS. WINKLER: Use your microphone, please.

MEMBER MOSS: I was happy to see

that you did not exclude children from this measure. But it looks like virtually all the evidence that at least I saw, anyway, was adult-based. This is a particularly important clinical issue in infants and children, and the ultimate criteria might end up being even more restrictive than you are suggesting.

Can you make some comments about the deliberations regarding children and how you would see this measure applying?

MS. AMOS: It was not intentional. Our intent was not to exclude children, but we would like to see this measure adopted for children as well, of course, taking into consideration that we said all patients for this particular measure.

MEMBER MOSS: I would love to see a measure in this area for children, but perhaps it could be constructed based on pediatric-specific data.

MS. WINKLER: Just to point out,

the specification measure is that children are included. This is not age-defined.

MEMBER MOSS: That was my point, exactly.

CHAIR GUNNAR: So, just a point of order. It was before an endorsed process measure; it is now in its current format moving to an outcomes measure. So, is it viewed as a new measure?

MS. WINKLER: We see the evolution of measures all the time. That decision would usually be arbitrary. The fact is the measure as they are presenting for their maintenance review is what they have presented to you. It looks like it has become an intermediate outcome measure rather than a process measure. Okay. It doesn't have to meet a new number or anything. No reason not to just go with the flow here.

CHAIR GUNNAR: Kelsey?

MEMBER McCARTY: I was wondering if

in the literature you cited or any other literature it reviews potential unintended consequences of increase in surgical site infection rates or other types of infections due to increased warming.

MR. POPOVICH: Can you please repeat the question? Is it an increase in surgical site infections from the literature on the warming aspect of it?

I think with the evidence that we have provided and the studies, it does demonstrate positive benefits of maintaining normothermia in those patients.

CHAIR GUNNAR: Any other discussion?

Oh, Dr. Reede?

MEMBER REEDE: You asked the question if it is reported now at a high level, at 36 degrees Centigrade, on arrival to PACU, that 30 minutes before, 15 minutes after, and it is.

So, as an anesthesia professional, I am wondering what kind of confusion we have when we go down to 35.5. I know what a struggle we have now to keep an operating room warm and to keep a patient warm. And where is the tipping point now? I don't see the value.

CHAIR GUNNAR: I think that was a similar question to what was asked earlier, and the response from the developers was that 35.5, although arbitrary, appeared to be a breaking point and applied to all. The measure, although not with a great deal of pediatric evidence, is applied to all ages.

Did I capture everybody's points?

So, I think unless there is further discussion, we vote on the evidence.

Amy?

MEMBER MOYER: I had the same intermediate outcome thought, and it almost felt like you did yourselves a disservice by including that guideline, which is mostly about

specific interventions, which we are not looking at since it is not a process anymore.

But when you go down to like 1a7.7, you start talking about the RCTs, about warming the patient, which is now what we are actually kind of measuring, that intermediate outcome of normothermia.

So, I am a little confused on how to evaluate the evidence, like what path we are going down. So, I am not sure how to vote.

MR. POPOVICH: Just to address the difference between the process and the intermediate outcome measures, when we were completing the application for maintenance review, there wasn't a choice for an intermediate outcome in the evidence. The other two documents, the application only lists our process for an outcome.

So, we did recognize that this was an intermediate outcomes and it may not be a true outcome measure. The process is moving

forward in that direction. But that is the choice that we made, is to label it that way, even though we do recognize that it is more of an intermediate outcome measure.

CHAIR GUNNAR: Collette?

MEMBER PITZEN: This is Collette.

So, the data that was provided, is that based on the prior specified measure of active warming or normothermia? Or is the data provided with the fairly high rates of the outcome measure?

MS. WINKLER: Collette, I think we will talk about that under gap and under testing. So, maybe we could focus on evidence and get through that.

CHAIR GUNNAR: Dr. Yates, did you have any other?

MEMBER YATES: No.

CHAIR GUNNAR: Okay. One more time, further discussion?

(No response.)

We will take it to a vote on evidence.

MR. SANCHEZ: Voting will now begin for Subcriterion 1a, evidence. One is for high; 2 is for moderate; 3 is for low; 4 is for insufficient evidence.

The voting timer starts now.

(Vote.)

CHAIR GUNNAR: Somebody is missing. Please try again.

(Vote continues.)

Okay.

MR. SANCHEZ: Five for high; 11 for moderate; 5 for low; zero for insufficient evidence.

CHAIR GUNNAR: Dr. Olsen?

MEMBER OLSEN: Yes, the performance gap is primarily just not knowing what -- well, only about 25, less than 50 percent of all the surgeries are captured by NACOR. You have to be on an electronic

healthcare reporting system to report into that system. So, it certainly runs into a lot of patients that are not reported that way, especially in smaller hospitals that don't have electronic medical access.

But in those hospitals that did report, it is reporting around 97 percent compliance rate.

CHAIR GUNNAR: So, the performance was the warming process, right, before?

MEMBER OLSEN: Correct.

CHAIR GUNNAR: So, it is 97 percent were actually actively found to be -- so, there was no gap, or a topped-off gap, as a performance measure, as a process measure? I'm sorry. A process measure?

Comments from the developer? Any relationship or thought or evidence that you have in relationship between viewing this, 97 percent as a process versus what do we know about it as an outcome?

MR. POPOVICH: Well, the way that the measure was written was combine the process with the outcome. So, the 97 percent is part of that.

The question here is the amount of providers who do not report this measure or have not reported the measure. It is how this will fair out in the future with just the outcome aspect of it, of the temperature. I don't want to project into the future of what that data might be.

MEMBER McCARTY: Just to clarify, so we don't have any data in terms of the current state, what percentage of anesthesia patients have a temperature at 35 degrees or above in that 45-minute period? There is no baseline? Or do you have a value for that?

MR. POPOVICH: We don't have a value for the 35.5 in this application. And part of that is just how to record the measure in the future.

With that said, I think the reiteration of several studies that looked at 35.5 degrees Celsius as the temperature, as well as overview by the AMA PCPI Anesthesiology and Critical Care Work Group that worked on this measure for two years, as well as many of our members, again 35.5, it is an issue of coding and actually gathering that data.

MEMBER McCARTY: I guess I am just wondering, in terms of if we are supposed to assess this performance gap, I am just wondering how much of a problem is this. So, do we only reach those temperatures 50 percent of the time? Do we reach it 99 percent of the time? It is hard to assess without knowing today how well people do with that.

MR. POPOVICH: Right. You are looking at the difference between the warming device as opposed to a 35.5 degree, right. And we don't have that evidence presented in the NQF application that we filled out.

CHAIR GUNNAR: No, go ahead.

MEMBER PITZEN: So, just a process clarifying question. Are we to evaluate the gap based on the data that we have right now, which is very high performing and looking to be topped-out, or how do we proceed?

MS. WINKLER: I mean, essentially, you're right, it is a challenging question because you need the data on these measure specifications. Now the question I would ask is, have you tested these measure specifications and what were your results? That would at least be minimum data.

MR. POPOVICH: There were no codes available for 35.5 degrees Celsius for the CMS 5 percent file. And as far as I know, NACOR has not measured the 35.5 degrees in a significant pattern or a significant time period prior to this.

Again, this was recently reviewed by the AMA PCPI group from 2010 to 2013,

approved by our House of Delegates just over a year ago. So, this is still in the process of gathering data.

MEMBER SIPERSTEIN: But I guess the question I guess we are trying to dance around, but haven't asked, on the prior measure, how many met criteria based on 36 versus how many met the criteria based on a warming blanket? So, how many made it into the PACU with a temperature of 36? Because that will give us at least a back-of-the-envelope in terms of what the gap is.

MEMBER CIMA: Having sat through multiple coders meetings with our team on this, if they had documentation of an application of a warming blanket, they passed. They didn't even look at the temperature after that point.

So, if they put it on and documented it within 15 minutes of the case starting, it didn't matter what their temperature. Their temperature could have been 30 at the end of the

case. They still passed.

So, the data that are available are not going to be able to distinguish this because of the way the previous measure was designed.

So, to your point, Reva, you said it is sort of we go with the flow. But this is actually a new measure. This is absolutely a new measure, and it should be just sort of skidded through based on the fact that we don't even know if there -- I mean, I am assuming there's a performance, I know there is a performance gap. I mean, but we just don't have the data on it.

MS. WINKLER: Right. And also, the other thing that has become clear that wasn't was it doesn't sound like these new specifications have been tested. So, that is a problem. With changing your specifications under a previously-endorsed measure, you still have to keep up with all the other criteria. So, it truly is problematic at this point.

DR. BURSTIN: But, to be clear, the performance gap itself could be from the literature. There is nothing that says it has to be from your data, just to be a stickler on process.

So, in fact, if you guys have suggested there is a performance gap known of patients not being adequately warmed, that is sufficient for performance gap. I think there is a larger issue that is now being sort of unearthed about the question of, is there actual data on the new measure and how it performs as just the outcome and whether that has been tested.

CHAIR GUNNAR: That clarifies the question in the room right now, which we have no answer to.

MEMBER McCARTY: So, another question I have about the data, so in the literature that shows that there are positive effects of patient warming, is the definition

in the literature of warming effects the same as your definition of having at least one measurement of a certain threshold within that 45-minute period? Are those aligned?

MR. POPOVICH: Yes. Yes, it is.

MEMBER McCARTY: Okay. Thank you.

MEMBER CIMA: Although there is new data that is coming out that says it is the aggregate temperature, Rick will say, over the entire case as opposed to the last 45 minutes. It is probably much like with the antibiotic measures. It is not a single event. It is the time course over the operation. Being 35 at the end of a seven-hour operation is different than being 35 at the end of an one-hour operation. And so, there is a time relationship that is not included in these measures which is probably why they are relatively weak.

CHAIR GUNNAR: But, again, to go back to what we have now from a performance

point of view, we have 97 percent being reported as the number, and which we are reacting to. So, with no real evidence to support the question about what is known as far as gap as the actual temperature, which is now the intermediate outcome that we are asked to evaluate it against, should we vote as it stands?

MS. WINKLER: Yes, I think we do have to proceed because these are the rules of engagement, if you will, for all the measures to be treated equitably. So, you are asked to use the information presented in front of you to make your evaluations against the criteria.

CHAIR GUNNAR: So, let's vote.

MR. SANCHEZ: Voting will now begin for Subcriterion 1b, performance gap. One for high; 2 for moderate; 3 for low; 4 for insufficient.

And the voting timer starts now.

MS. WINKLER: Dr. Dutton, on behalf

of the measure developers, just letting them know that, given the issues that we have raised, rather than continue chewing up time, they are going to withdraw the measure for current evaluation, and, hopefully, bring it back having addressed the issues.

Does that work for everybody?

The next measure is Measure 0465, perioperative anti-platelet therapy for patients undergoing carotid endarterectomy.

MR. LYZENGA: Actually, sorry, Reva, this was apparently included by mistake. This one was withdrawn as well. We will move on to the next one.

MS. WINKLER: Oh, okay. Hey, we're catching up just fine.

(Laughter.)

MR. LYZENGA: Now it is 0527, unless we want to do a break.

MS. WINKLER: I know.

This puts us ahead of our agenda,

and I think the question we want to ask is, are our measure developers for those first measures from CMS, is Dale Bratzler back to be able to do those? Or perhaps it is time we can take a break and get a hold of him, so that we can get started on the measures.

Let's see. Go ahead and take a break. We will reconvene you if we can find CMS or some of the other developers.

(Whereupon, the foregoing matter went off the record at 2:21 p.m. and went back on the record at 2:40 p.m.)

MS. WINKLER: Okay, folks, we've got the developer on the phone, so we can get started again.

Wanda, are you there?

MS. JOHNSON: I am.

MS. WINKLER: Great! Super! Do you know when Dale will be available?

MR. BRATZLER: Yes, this is Dale, I just got here.

MS. WINKLER: Thank you, Dale. We are running you ragged today. Thanks so much for coping. We just got further ahead of our agenda than we expected. So, we are reconvening the group and we will get started momentarily. Thanks so much.

Dale, quick question. When we get to the PCPI measures, could you speak to those as well?

MR. BRATZLER: Yes.

MS. WINKLER: Okay, thanks.

MR. BRATZLER: We had them up on the website. I'm not logged into the website. I just called in in a hurry.

MS. WINKLER: Okay.

MR. BRATZLER: Let me see if I can get back into the website.

MS. WINKLER: Okay, great. Thanks.

Okay, if we are reconvened, okay, we are going to start with Measure 0527. And this

will be the first of the three measures from CMS around antibiotic prophylaxis.

So, Dale, did you want to just say something about the measure by way of introduction?

MR. BRATZLER: So, give me 0527 is the -- which title? I still don't have the website up yet.

MS. WINKLER: This is received one hour prior to surgical incision.

MR. BRATZLER: Okay, so this is one of the very first of the SCIP antibiotic performance measures that focused on delivery of antibiotics within 60 minutes prior the incision. Originally started the measure in a pilot project back in 2002 and then it went national in 2005, with the Deficit Reduction Act change.

The performance measures looks at most antibiotics initiating the dose of antibiotics within 60 minutes before the

incision. But for certain long half-life antibiotics, vancomycin and fluoroquinolone, it is 120 minutes before the incision.

Performance will become the metric word about 55 percent when we first started measuring the performance on the measure and it has increased very substantially over the years since we implemented the measure.

There are certain categories of patients that get excluded from the measure but the principle exclusion are those patients who have documentation of an infection because we assume that those patients are receiving antimicrobials for treatment, not prophylaxis.

So, I would be happy to answer any questions.

CHAIR GUNNAR: Who is our discussant? Dr. Cima.

MEMBER CIMA: So, as Dale said this is the granddaddy of measures. I am just going to go through really quickly. So, this is 0527

prophylactic antibiotics within one hour of surgical incision.

According to the guidance for evaluating clinical evidence, this is a process measure, not a direct outcomes measure. However, it has a substantial body of literature both experimental literature, as well as population based, as well as randomized trial literature supporting it, as well as in 2013 the CDC, American Infection Society, Hospital Pharmacists sent out a joint guidance saying that this was Level 1A supporting this for surgical patients who receive antibiotics within 60 minutes, with the small exception of those long-acting agents as best practices strongly supported in the literature.

So, from an evidence point of view, it would be rated as high.

CHAIR GUNNAR: Any discussion? Hearing none, let's vote.

CHAIR FLEISHER: I am choosing to

abstain, since I am on the SCIP Technical Expert Panel. It is not required, but I have chosen so.

CHAIR GUNNAR: So noted.

MEMBER GROVER: I'm on the SCIP Committee but I wasn't on the technical advisory panel that developed this. So, I will trust your judgment, as chair.

CHAIR GUNNAR: I have no issue that you -- I would think you would be able to participate in this vote.

So, let the record note that Dr. Grover will be a part of this vote and Dr. Fleisher will not.

So, can we go to the vote?

MR. SANCHEZ: Voting will now begin for subcriterion 1a, evidence. One is high, two is moderate, three is low, and four is insufficient evidence.

The timer starts now.

(Voting.)

CHAIR GUNNAR: Try it one more time.

MR. SANCHEZ: Twenty for high, one for moderate, and zero for low, and zero for insufficient evidence.

MEMBER CIMA: In regards to opportunity for improvement, when this originally was rolled out in the mid-2000s, appropriate dosing administration was around 50 percent. Clearly, given the focus on this over the last decade, has driven that upwards significantly. It is now, most of the national studies show it to be at 98 or greater percent data. The last five quarters of data provided by the developer show it at 98 percent, which is fairly constant.

And a few, very small number of hospitals are just around the 95 percent. So, if you use 95 percent as your marker of success, then everybody that is reporting this is in compliance.

I will say that there is some literature out there saying that what is reported and what is reality are different but we can't go into that. We have to go off what we are provided. And so this raises the question of whether or not this is a topped out measure.

CHAIR GUNNAR: So, your recommendation is what?

MEMBER CIMA: Well, it is hard to separate what I would like to see happen and what the thing -- if we used 98 percent, which we had used previously, saying that was a topped out measure, then I would this is a topped out measure. Does it mean it should be retired as a measure, I would say no but that is not what the question asked of me.

MEMBER SAIGAL: Can I ask a question? So, if we vote that this thing has low performance gap, then we automatically move looking at this as a reserve measure?

MS. WINKLER: We'll ask the question as we did before, whether you want to consider it as a reserve because there must be measures that don't pass the gap that you don't want to go forward, perhaps. So, it is a secondary question after doing the evaluation of this criterion.

CHAIR GUNNAR: Dr. Handy?

MEMBER HANDY: Well, on the other hand, it is so hardwired that everybody is successful. Why do we measure it anymore? This is going to be a recurring theme with all these antibiotic issues. That is the same one that I am presenting, too.

CHAIR GUNNAR: Dr. Grover.

MEMBER GROVER: Just explain to me a little bit. Because I remember I used to go nuts trying to get everybody to comply with following the evidence-based literature in our department and deliver these drugs close to the time of the incision.

Now, if we say it is topped out and put it on reserve, will they still have it in place or does that mean they can have it in place, the measure, or what are the ramifications?

MS. WINKLER: Well, I think that is one of the things that we are going to start evaluating, given that reserve status has been around a while. But remember that ultimate decisions for implementation are with the folks which implement them, which tends to be after the NQF endorsement.

I think you are seeing that folks like CMS are retiring measures in some areas. So, this is a very dynamic and evolving process. So, it is something to think about.

I guess one thing we might want to ask the measure developer is, okay, this measure is done very well. It has probably been very successful at doing this but what is the next generation of measurement in this very

important area?

Have you considered -- because these measures only address a certain subset of procedures. Have you considered expanding the denominator, creating a composite measure, replacing with the outcome? I mean what is the thinking around measurement for this topic area? Because really --

MR. BRATZLER: This is Dale. So, I can't speak for CMS. There may be a representative from CMS on the call. And I think my perception, this is Dale's interpretation of what he sees and that is the movement is towards outcome measures, focusing on surgical site infection rates. And as you know, there are certain surgical site infections that are a part of the Hospital and Patient Quality Reporting Program that CMS has in place through the National Healthcare Safety Network. So, I think that is the general direction that CMS is headed, at least based on

what you see in the published rules.

That said, I would say for this particular measure, this one probably is hardwired. The delivery of the antibiotics has been hardwired into many different specialties, the anesthesiologist, the circulating nurses or others that makes sure this happens fairly routinely.

I'm not as convinced when we get to some of the other performance measures that are coming up for discussion such as discontinuation, that we might not see fairly substantial slippage if we aren't looking. This one I don't know because this one is very a systems-based measure.

MEMBER JARRETT: This is Mark. I tend to agree. I think things like this that have been hardwired in, if we are going to have to continue measuring these because they really weren't hardwired, we are never really ever get to outcomes because we are just going to be

looking at processes for the next generation.

So, I think we have to assume that people have hardwired it. I think it is up to individual institutions to go back and audit the process every once in a while to make sure that sustainability is there. But when we are up to 98, 99 percent to keep measuring it, it becomes a blur with a thousand other measures that we have to do. And I don't think it adds real value at this point.

CHAIR GUNNAR: Yes, so it is interesting. So, part of this, and maybe my perception is that during this time span the advent of universal protocol, time our checklist, all the things we do are sort of hardwired now. That took years to actually make standard process. But in fact, now, is the case antibiotic prophylaxis is in that. So, it may well be that this tracks with sort of the overall safety culture of our surgical programs generally. So, that is a reflection.

Dr. Yates.

MEMBER YATES: I would be concerned about assuming the hardwiredness of this because I think there is a certain amount of what they call in physics a Heisenberg effect and that the act of observing causes something to occur. And being under observation does raise the ante in terms of the anxiety over making this happen, especially when you get into alternative drugs cephalexin such as vancomycin or ciprofloxacin, something of that sort.

The second thing that I would observe is that there has been a carrot attached to the end of this stick in terms of it being associated with value-based payments. And as we move forward with value-based payments to the hospital, the process measures are going to become smaller compared to what they are now, as we move to outcomes measures. And you are going to see, moving to hospital-acquired

conditions, i.e., complications of surgery, in particular infections, becoming the outcomes measure. But as the process measures become smaller and smaller parts of the value-based payments, you want to see whether or not you have a falling off, as people maybe don't value it or worry about it as much.

And finally, it would be great to have CMS analyze this. Right now, it is not this committee's business to be concerned about whether this is a threshold for them or whether this is a moving benchmark. But since it is a moving benchmark that moved right up against, the ceiling, it may be important to establish, at some point, before putting this out into pasture what the actual flux is in terms of actuarial risk at the edge of 98, 99, that not hospital -- I mean there will be hospitals with 100 but there is going to be statistical anomalies that hospitals fall below that benchmark and maybe keeping this alive for a

little while longer, CMS can establish something like a threshold of 98 percent to protect hospitals against that anomalous fluctuation that occurs at any edge of statistics.

So, those would be my points.

CHAIR GUNNAR: Dr. Grover, did you have your sign up?

Any other discussion?

MEMBER JARRETT: This is Mark again. And I think one of the points -- and Dr. Yates, I agree with a lot of what you said. But I think one of the points made earlier is perhaps we ought to think in terms of composites or bundles so that you take a bunch of these process measures, along with getting to the outcome that you want as the next step. And maybe that should be what is done, rather than looking at every individual measure along the line.

Because again, I think, that yes, if

people know you are looking at it, they tend to do a little bit better but we can't look at everything. And at some point we have to draw the line. And I know it is hard. Do you draw it this year? Do you draw it next year? But I just see more measures coming down, especially as value-based purchasing is changing. And my concern is that we are just going to have people looking for that last two percent, which may be a small number of cases and putting a lot of resources to that and not really bring resources into more major issues. So, that is my only concern about moving this, for example, either into a composite or into a reserve.

CHAIR GUNNAR: Dr. Asher?

MEMBER ASHER: It was just to amplify that comment. My observation is that the resource issue is becoming a very significant one at even the major medical centers. And the persistence of something

like this keeps these more substantive measures from being implemented.

And I was reflecting on this JAMA article that came out a few months' ago, just cataloguing the number of measures that are out there. It is just astounding, just between the CMS and Joint Commission. It seems to me if the intent really is to do things that are going to be more significant in terms of moving the needle, we need to make room for those measures.

CHAIR FLEISHER: So one of the comments from the perspective of sitting on CSAC, is we are not -- how they are used is different, to some extent, than what our job is. No, you don't think?

DR. BUSRTIN: I think he is speaking directly to the criteria, which is, is it topped out or not.

CHAIR FLEISHER: Right. That is the question.

DR. BUSRTIN: But part of the logic

of that, the genesis of that criterion, in fact, is to remove measures that are no longer are adding value. So, I think that is the question. So, I think speaking to that is fair.

CHAIR FLEISHER: Yes, we are saying the same thing.

MEMBER CIMA: I would just on one side, if you look at the three antibiotic measures, the one that has the strongest level of evidence to support it is this one, as far as actually doing what you want it to do, which is to decrease surgical site infections. There is none of these other measures have the level of evidence that support them for reducing surgical site infections.

Number two, there is some data that if you use appropriate prophylaxis, amongst antibiotic choices, there is support for that that does show improvement in surgical site infections.

Number 0529, which is part of the bundle, extending antibiotic use, that is, for antibiotic stewardship reasons and a number of other reasons is not good for patients. But if you are asking what was this measure designed to do is to try and help reduce surgical site infections, which is the largest number of hospital-acquired infections in surgical patients and the largest cost of morbidity in surgical patients. This measure is the only one of all the measures that we do that actually has strong Level 1 evidence to support what we are asking it to do.

And although it has been hardwired, I can tell you from touring around multiple, multiple hospitals, it is not as hardwired as we would like to think, and as recent data on the Surgical Safety Checklist has showed us, that people use it but don't really use it.

And so, making the reliance on it is saying that it is actually a tool. I think, we

are going to see slippage and it may be a problem.

CHAIR GUNNAR: Any other discussion?

MEMBER SAIGAL: What do you mean by use it but not really use it?

MEMBER CIMA: There is a lot of new papers that are saying people just sort of check the boxes and aren't actively engaged in using it. And that just because you have a checklist and everyone participates, whether it actually translates into the actual safety or outcomes that were initially proposed, such as the paper that came out of Canada recently, it is a question of are they actually buying into it. Are the processes changed?

MEMBER SAIGAL: You said they are actually not doing it but they are reporting that they are doing it. That is what you are saying?

MEMBER CIMA: However you want to

take it, they are not actively engaged in what it is really intended to do.

CHAIR FLEISHER: That is the Ontario paper and that is a surgical checklist with Lucian Leape's editorial. That is a different question about antibiotic timing. That is whether or not the preoperative checklist works.

MEMBER CIMA: I'm just saying just because we think it is hardwired, doesn't mean it is. And I tell you there are a lot of hospitals out there where it is not hardwired.

CHAIR GUNNAR: Yes, I just want to get us back focused towards this vote. Dr. Levy?

MEMBER LEVY: So, I think this has topped out as a measure and I think what we really care about are surgical site infections and I think we really need to be moving in that direct. And how a hospital or a system would choose to use this or not internally to ensure

that their real outcome measure, which is reduction or elimination of surgical site infections would be up to them.

But I think at some point, we have to say this is topped out. And I don't disagree that it should be in reserve status so that we can continue to follow that but we really need to be moving to outcomes.

CHAIR GUNNAR: I am getting pressure from my left to call for the vote. So, unless any other discussion, I think we can vote.

MR. SANCHEZ: Voting will now begin for subcriterion 1b, performance gap. One for high, two for moderate, three for low, four for insufficient. The voting timer starts now.

(Voting.)

CHAIR GUNNAR: We need one more. Try it again. Make sure the light turns green.

MR. SANCHEZ: We have got three for high; three for moderate; 16 for low; zero for

insufficient.

CHAIR GUNNAR: So this fails on an endorsement. But now the question to the committee, do we vote should it go forward as a reserve status. And so a show of hands for yes.

(Show of hands.)

MS. WINKLER: No, what it means is you will continue to evaluate the measure and determine your final recommendation. Otherwise, it is over.

CHAIR GUNNAR: We haven't gotten to the next part yet, which is, it has to hit all the other points, criteria, to then actually be recommended for reserve status. But would the committee wish to go forward with reserve status evaluation? Yes. That's it. Okay, so we go forward.

For the record, I think that was unanimous.

MEMBER CIMA: To continue, then,

whether or not this is priority in how it would be established, the construct of it -- no, that's not. That is one for composite.

It does meet a healthcare goal, reducing surgical site infections. It has been shown to have high impact on outcomes over the years and in multiple studies. And this has been adopted widely without much difficulty across the country. So, in that sense, it is something that would be considered something a high priority measure.

CHAIR GUNNAR: Any other discussion? Hearing none, shall we vote?

MR. SANCHEZ: Voting will now begin for subcriterion 1c, high priority. One is for high, two of for moderate, three if for low, four is for insufficient. The voting timer starts now.

(Voting.)

CHAIR GUNNAR: I think we are good.

MR. SANCHEZ: We have got 16 for

high; three for moderate; three for low; zero for insufficient.

CHAIR GUNNAR: We carry it forward.

MEMBER PITZEN: I have a process question. I just need to understand reserve status better. Does that mean that data will still be collected and reported or is it just a measure that is designed to be a really great measure and people can use it if they want to?

MS. WINKLER: It will depend on the programs that use it to determine how they react to the reserve status. But again, it is meant to add sort of a bit of a warning label saying this might get you a lot of good information.

MEMBER CIMA: As it relates to reliability, this has been one of the original measures that have been evaluated. It does have a very clear numerator and denominator. It is mainly designed around a specific group of procedures. Not all procedures require preoperative antibiotics. That would be a

different discussion. But the main groups that have identified a large cohort of patients, exclusion, criteria, are well-established, easy to identify. So, the measure itself report has a high reliability in reporting what it is designed to report, which is clearly antibiotic administration in a certain cohort of patients with appropriate exclusion criteria. So it actually is easily extractable in a way that is highly reliable across institutions.

CHAIR GUNNAR: Questions?

MS. WINKLER: Just a comment on the criteria. Reliability does include anything having to do with specifications. But in terms of the testing, it looks like this measure was evaluated at the data element level against the gold standard of the chart. So that means that will apply to both reliability and validity of testing at the data element level. And so, therefore, the highest rating for that would be

moderate when testing for both reliability and validity.

CHAIR GUNNAR: Any further discussion? Take it for a vote.

MR. SANCHEZ: Voting will now begin for subcriterion 2a, reliability. One is for high, two is for moderate, three is for low, four is for insufficient. The voting timer starts now.

(Voting.)

MR. SANCHEZ: We got nine for high; 14 for moderate; three for low -- zero for low; zero for insufficient.

MEMBER SAIGAL: So a point of clarification here. Does that mean -- does it move forward?

MS. WINKLER: Yes.

CHAIR GUNNAR: Dr. Cima, validity. It follows a similar pattern.

MEMBER CIMA: Yes, basically it is very similar with the -- it is based on what Reva

said also about the source of the data but it is not a complex problem to identify, so that it is valid on that point. It is also valid as far as administration and how you would implement this and operationalize it.

MS. WINKLER: Yes, validity would also encompass things like do the measure specifications reflect the evidence. Is there an alignment there?

Also in the assessment of threats to validity, such as how exclusions are handled, how missing data is handled, any risk adjustment, if necessary, not so much for this measure but in general.

So, there are an assessment of threats to validity would be the other aspects of validity, aside from the actual testing.

CHAIR GUNNAR: Any other discussion? Hearing none, we will take it for a vote.

MR. SANCHEZ: Voting will now begin

for subcriterion 2b, validity. One is for high, two is for moderate, three is for low, four is for insufficient.

The voting timer starts now.

CHAIR GUNNAR: I think we're good.

MR. SANCHEZ: Twelve for high; 11 for moderate; zero for low; zero for insufficient.

MEMBER CIMA: So in regards to feasibility of collecting this data, it is elementized data. It can reside in electronic medical record, which can be pulled, or on a paper record. In the cases of a paper record, it can be resource-intense to pull it but it is, basically, two discrete fields and then we have to pull all the exclusion criteria on the patient type. But it is relatively straight-forward data and is accessible relatively easily in a standard medical record.

So, it is high to moderate.

CHAIR GUNNAR: Any further

discussion? Hearing none, vote.

MR. SANCHEZ: Voting will now begin for criterion 3, feasibility. One is for high, two is for moderate, three is for low, four is for insufficient.

The voting timer starts now.

(Voting.)

CHAIR GUNNAR: You got 22. We are getting faster.

MR. SANCHEZ: Eighteen for high; five for moderate; zero for low; and zero for insufficient.

MEMBER CIMA: This is on the usability and transparency. It is used on multiple websites as part of Hospital Compare. It is multiple state requirements, depending on the state require public reporting of this. It is accessible to the public in multiple venues. It has been shown, over time, to show improvement in how people have been performing. So, it meets all of those criteria from an

original inception, question now moving forward but originally, it meets all those criteria.

CHAIR GUNNAR: Any other discussion? Hearing none, we will take it for a vote.

MR. SANCHEZ: Voting will now begin for criterion 4, usability and use. One is for high, two is for moderate, three is for low, four is for insufficient information. The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have 20 for high; one for moderate; two for low; zero for insufficient information.

CHAIR GUNNAR: So this question is really about should this NQF criteria for endorsement really meet -- is there a status. Right?

MS. WINKLER: Correct. I mean reserve status is still endorsed but it does

have that extra thing attached to it as reserve status. So, that is the question to you is it should be recommended and reserve status endorsement.

CHAIR GUNNAR: So any additional discussion regarding this before we vote on reserve status?

Hearing none, let's go ahead and vote.

MR. SANCHEZ: Voting will now begin for endorsement -- or potential for reserve status. One is for yes; two is for no. The voting timer starts now.

(Voting.)

CHAIR GUNNAR: We are just waiting on two more, if you can enter your votes again.

(Voting.)

MR. SANCHEZ: We got 21 for yes; one for no.

CHAIR GUNNAR: So, the recommendation of the committee is to maintain

this particular measure in reserve status.

MS. WINKLER: The next measure is 0528, essentially the same in this group of measures around the selection of antibiotics for surgical patients.

CHAIR FLEISHER: Can I make a suggestion? Which is, actually, look at 0269, ASA? Are they coming back or are they not? You are here. Because it is essentially the same measure, it is just physician-level. And this might allow a discussion that is concordant about it. And there really are pairs of measures that are -- essentially the hospital measure with the PCPI, the physician measure. Make sense? That way, you may be able to go through this much quicker.

MEMBER DUTTON: And I will recuse myself from 0269.

CHAIR FLEISHER: As I will also.

MEMBER DUTTON: Matt and Maureen are here.

CHAIR GUNNAR: So, to get us squared away, I think we are going to move to 0268, Perioperative Care: selection of Prophylactic Antibiotic: First OR Second Generation Cephalosporin. That one?

CHAIR FLEISHER: No, 02696.

CHAIR GUNNAR: Okay, 0269, Timing of Prophylactic Antibiotics - Administering Physician (ASA). And Dr. Fleisher and Dr. Dutton will recuse themselves.

But just for -- we have to consider them separately. So, I guess the question is, CMS was in line. So, you want us to do these one after the other? It just is a similar discussion.

CHAIR FLEISHER: My proposal to the committee is, having listened to this, they are paired, essentially. So, hopefully, you can quickly go through this and discuss. I mean if you are going to things to the hospital-level measure, should you act differently? And I am

opening up to the committee, should you act differently for the physician-level measure.

CHAIR GUNNAR: So, is it your recommendation that developers for both of these measures simultaneously present?

CHAIR FLEISHER: So, I am just saying --

MS. WINKLER: He is changing the order.

CHAIR FLEISHER: I mean, does anybody have any comments? It is just a proposal that you would want to look at these --

MEMBER SAIGAL: I agree.

CHAIR FLEISHER: Okay, thank you.

CHAIR GUNNAR: Anyone disagree? Hearing none, we will now move to 0269 and the ASA developers are presenting. Would you like to make some opening comments, please?

MR. POPOVICH: Sure. Thank you. This measure was first developed in 2006 and

endorsed in 2008. The measure looks at administration of antibiotics, administration prior to surgical incision. We do cite a significant number of studies and guidelines for administration. The studies do go back to 1957 and there are over a thousand studies concerning prophylactic antibiotic administration.

The potential downside of a patient not receiving an antibiotic is infection and there is a significant amount of data showing a strong association with this process measure with patient outcomes. Thank you.

CHAIR GUNNAR: The discussant is Dr. Moss.

MEMBER MOSS: So, I will try to move through this pretty quickly. The evidence here is the same clinical practice guideline that we just discussed. Nothing really much to add to that, other than the point that the evidence clearly establishes a link between

appropriate timing of antibiotics to reduce surgical site infections but it doesn't really speak to the fact that surgical site infections are multi-factorial and this is just one of many factors which are not addressed. But I would still say the evidence level is high.

CHAIR GUNNAR: Any discussion? We will carry on and vote regarding evidence.

MR. SANCHEZ: Voting will now begin for subcriterion 1a, evidence. One is for high, two is for moderate, three is for low, four is for insufficient evidence.

The voting timer starts now.

(Voting.)

CHAIR GUNNAR: We're there.

MR. SANCHEZ: We have 19 high; two moderate; one low, zero for insufficient evidence.

CHAIR GUNNAR: So, Dr. Moss?

MEMBER MOSS: So moving on to performance gap. Similar findings here over a

three-year period where this has been measured. The performance has gone from 93.7 to 94.9 percent, not a very significant change.

Just one comment under this category with respect to disparities. I wanted to ask the developers why children were excluded from this measure.

MR. POPOVICH: It is how the measure was originally written in 2006 and carried on through the past few years.

CHAIR GUNNAR: Dr. Yates?

MEMBER YATES: If I read the measure submission correctly, the biggest gap is that only 50 percent of anesthesiologists report on this.

MR. POPOVICH: Yes, reporting the measure is in the 50 percent range. And it depends -- I think that there were also discrepancies within the Medicare population reporting, as well as other payer reporting.

MEMBER YATES: And following

through on that, it is somewhat surprising to me that this measure would measure the anesthesiologist performance, when in fact the preoperative administrative of antibiotics may be the responsibility of the ordering surgeon, may be on call to the OR, maybe a hospital issue, much more than an individual anesthesiologist.

Out of this performance gap, is there any analysis of the effect of the hospital's performance, since there is a co-related measure with this? Have they been linked to see whether or not the hospital performance overwhelms the performance of the anesthesiologist, per se?

MR. POPOVICH: We haven't presented evidence comparing the two within this measure but we can always check to see if that data is available and report back to the committee.

MS. AMOS: We also pulled, as part of the data, we used the five percent Medicare

files, including NACOR data that was presented in this documentation.

CHAIR GUNNAR: Dr. Reede?

MEMBER REEDE: Thank you. Inside NACOR, then, I believe that CRNAs and anesthesiologist assistants are also reported and they are also administering antibiotics?

MS. AMOS: That is correct.

MEMBER REEDE: So, the administering physician, should it be administering clinician, if we are going to look at a specific anesthesia provider administering?

MS. AMOS: It is all-inclusive. And you know we recognize that in the title it says administering physician but it is actually provider.

CHAIR GUNNAR: Ms. McCarty.

MEMBER McCARTY: One concern I have about this measure, and I should have raised it during the previous discussion, is that in a lot

of the electronic capture of this documentation, you are actually recording documentation of when antibiotics were delivered, as opposed to when they were actually delivered.

And I have seen and I have heard about, and I have even been to discussions where people talk about because there is reimbursement rates tied to this about how it is very easy to just change the time stamps manually, in order to comply with this measure.

And one I would like to advocate that we don't put this into reserve status like we did with the hospital-wide one because I do feel that when you look at it at the physician level that you can, sometimes pull that out and provide coaching about how to actually deliver on time and not just document on time. So, I just wanted to raise that point and advocate for this measure.

CHAIR GUNNAR: Dr. Grover.

MEMBER GROVER: My question is on that 50 percent of anesthesiologists or their staff that are reporting, that they are reporting all of their cases. Can you document that or is there cherry picking?

CHAIR FLEISHER: If I can just comment, being the chair of an anesthesia department that doesn't report but has 99 percent compliance on this measure, it is purely an issue of when they say 50 percent don't report, they don't reporting anything. It is people not using those codes. That is the function.

MEMBER JARRETT: This is Mark. If you are getting documentation that is lacking, what are we really measuring?

CHAIR GUNNAR: Collette?

MEMBER PITZEN: Collette Pitzen. Maybe this isn't the right time to ask this question but I just wonder, do we need two almost identical measures that are really

captured in a different way or stratified differently?

MS. WINKLER: Collette, your question led into mine and I think it is important that we look at the specifications for these measures because, in fact, if you look at the denominator on what patients are being captured, the numbers of procedures in this measure is quite a bit larger than the procedures captured in the hospital measure.

So, I just wanted to be sure the committee was aware of that.

The other thing is I wanted to verify from the developers is you have indicated the level of analysis for this measure is not only the clinician, either group or individual, but also facility, which means hospital level.

MR. POPOVICH: The NACOR does collect that data as well.

CHAIR GUNNAR: Yes, question?

MEMBER SAIGAL: So how does it relate to your comment about fraud, if the facility and the physician are reporting, there should be some concordance there.

MS. AMOS: I'm sorry. Could you repeat the question?

MEMBER SAIGAL: Well, there was a concern raised about somehow gaming the system.

MS. AMOS: Right.

MEMBER SAIGAL: And then I was wondering if it is reported but the facility level -- data you get from the facility and the doctor to make sure that they are concordant, is that what you are saying?

MS. AMOS: So facilities could report this information but the provider could as well. So, a group, a physician within a group, an individual physician or a facility hospital could report.

I think what you are asking is if there is some cross-checking between what the

physician reports and what the hospital reports. And to my knowledge, we have not done that cross-check.

CHAIR GUNNAR: Dr. Ko?

MEMBER KO: I wanted to follow up on Reva's comment. That where the numbers are different, the specs, the denominators are different. Is that supposed to be different or is it just operationally that it is hard to keep up with all the codes and inclusion/exclusion? And if that is the case and they should be the same, would this be something to harmonize and then have different levels?

CHAIR FLEISHER: So being there at the beginning, if I can, so SCIP was created with a very defined set as number of procedures. And when the ASA and the American College of Surgeons got together for the first time the PCPI, they chose to expand it.

So it is more that SCIP has always stayed with a very small group of procedures and

has focused on that, while when ASA and ACS and actually ANA, and AUA, most of the people who are affiliated, not represent people in the room, it was chosen to be a much wider and more inclusive group.

So, I think it is just historical but I don't see SCIP changing their perspective and going to a larger group.

Dale, do you want to comment?

MR. BRATZLER: Yes, well, again, I think that has been mainly the administrative decision not to expand the denominator. As we said, the SCIP denominator was originally designed to pick common operations performed on Medicare patients. It was never meant to be comprehensive for all operations that should receive antimicrobial prophylaxis.

MEMBER KO: Well, it is a purview of this committee to potentially suggest that? I mean as a lot of these things are coming down in the QCDR, the PQRS stuff from CMS is

including a lot of non-CMS patients. It seems like they are going that way. Is this something that we could potentially suggest? Because now is different than 2006.

MS. WINKLER: Yes, Cliff, I mean essentially is the fundamental question around harmonization, which is going to be when we get through all of these measures is now going to be the question before you as well is harmonization, alignment, consolidation, call it whatever you want to. But is there some way we can make more sense out of all these measures?

CHAIR FLEISHER: And I guess that also gets to the attribution question. So in 2004, you had SCIP independently trying the voluntary process. You had PCPI and PQRS developing. And in order -- and at the time, there wasn't the stick to the hospitals quite as much.

So, we now, whether there should be

joint attribution or you really need to create separate categories but they exist to date. That is part of the historical context.

MEMBER MOSS: So, those are valid points about potentially expanding the denominator but we just voted 22 to one that this measure was topped and should be in reserve status. So, how do we reconcile those two issues of sending the message that this should be in reserve status but yet it should be expanded?

MS. WINKLER: I think that this is a difficult pathway. I think one of the interesting things about this measure is that they are specifying it also at the hospital level, not only at the clinician level.

So, you have a measure that includes a larger number of levels of analysis, as well as a larger number of procedures captured in the measure. So, when it comes to looking at harmonization or perhaps we are talking about

competing measures and whether one would be better going forward compared to others, this is why we will probably have to have sort of iterative conversations as we go through these measures to see at the end of the day, what do we really see going forward as the best group of measures to achieve our end goal?

MEMBER CIMA: Well, just to go to that point and what we just finished talking about, we talked about one of the main reasons to move off of the other one was well, it takes resources. It takes time. We should move on to more important things.

And now we have almost the exact same measure. It is going to take the same amount of resources, the same amount of thing and so I am not even -- does the argument from the last one 20 minutes ago now fail or no longer is valid because we have expanded the denominator and made it even more difficult to do what we said we didn't want to do?

I mean so, that is why linking these two, it is almost like when one goes the other has got to go, unless we say it is more important. But the reasons we gave for getting rid of the other one was that it was topped out. And this is 95, 96 percent or 94 percent. I still don't know how they can be that discordant, other than because it is a cross, a bigger denominator, probably. But the resources and the things that we said were going to be needed to do the other one are going to be the exact same resources, only more so.

So, I am just -- I am not following the logic here of even continuing the discussion.

MEMBER MOSS: I second that.

MEMBER YATES: Just for anybody listening in, it is not -- the measure that was just put into retired pasture with reserve status was CMS SCIP, not the hospital ASA measure, which we passed over to go straight to

the clinician provider measure, which is what we are talking about right now.

MS. WINKLER: No.

MEMBER YATES: Or am I wrong?

MS. WINKLER: Yes.

MEMBER YATES: We looked at CMS SCIP as the first measure.

MS. WINKLER: Correct. And all we did was reorder the measures you are looking at.

MEMBER YATES: Right. But we reserved --

MS. WINKLER: And so this measure is -- the reserved measure is the SCIP measure.

MEMBER YATES: Right.

MS. WINKLER: Now, you are looking at another measure that originally started out at clinician level. But I think since it transferred from PCPI over to ASA and they have data from the NACOR registry, they are able to say it is a hospital-level measure, too.

MEMBER YATES: Right. No, I

understand that. But someone had just said that we had just retired the hospital measure. And we are not talking about that hospital measure yet because we skipped over it.

The one we retired was the SCIP hospital measure.

MS. WINKLER: There isn't any --

MEMBER YATES: They are both the same. So they crossed over?

MS. WINKLER: I don't know --

MEMBER YATES: I thought ASA, I thought that the measure before this that we didn't discuss is a separate measure from ASA.

MS. WINKLER: It is.

CHAIR FLEISHER: I think all we are saying is that the measure can be aggregated at the hospital level for this measure.

MEMBER YATES: Right.

CHAIR FLEISHER: So, they are the same measure --

MEMBER YATES: Right.

CHAIR FLEISHER: -- just aggregation.

MEMBER YATES: But we haven't retired that part of or that aspect of that measure yet.

CHAIR FLEISHER: We haven't discussed it at all.

MEMBER YATES: Right. Right, I understand. I am just clarifying that because someone spoke otherwise.

And I would just double -- I would just second what was already said just now is that if the one is deemed as topped out, I would agree 100 percent that this is topped out as well.

CHAIR GUNNAR: Dr. Moss.

MEMBER MOSS: So, this is -- I would suggest this is probably the most visible of all surgical outcome measures. And that the country is probably looking to leadership, looking for leadership from this group about

where we are going with respect to surgical site infections. I would just suggest that after eight years of are we giving the antibiotics on time, we need to send a message it is time to move on to outcome measures.

CHAIR GUNNAR: Dr. Ko.

MEMBER KO: So maybe I can just inform a little bit about that from NSQIP. We tried to develop an outcome measure and then we harmonized it with the CDC.

And so we were -- it was easier to do the colorectal SSI outcome measures, risk adjust it with what was in the NHSN data set and we were looking at reliability of distinction, a very high level of statistical rigor.

And beyond that, it was hard to pick any other procedure to do SSI just because of the rates and -- because of the rates of the infection and the numbers that are being done in hospitals. The one that was far down the list was hysterectomy and that is why that is

the second one. Beyond that, we couldn't really do -- something that would pass this committee outcome measure in SSI for a procedure.

CHAIR GUNNAR: Ms. McCarty.

MEMBER McCARTY: So again, the fact that this is an individually measured metric, there is very few measures out there that can be done at the individual measure. And I think the fact that it is at 93 percent and not the 97, 98, that we see for the hospital one, means that there is still some room for improvement.

And in terms of hospital culture and accountability and being able to drive improvement, oftentimes that is done with feedback to the individual. And what better place to start with moving in that direction towards individual reports than with a measure that we are all so accustomed to and very comfortable with, like prophylactic antibiotics.

So, in terms of the NQF mission of trying to think forward in terms of where do we want to go with this committee and what types of measures do we want to endorse, I would say thinking about individual metrics might be a good place to focus our efforts and this is a really good one to start with.

CHAIR GUNNAR: I'm not sure what your -- so, let me ask it this way. To cut through all the other voting, I guess I will take it as a motion on the floor and it was actually seconded, was that this particular measure be assigned to reserve status. Just get rid of all the other voting. We could go through it but -- let's just have a show of hands. Because if you don't have the show of hands, then we will take each one of these and go forward.

So, do people want to place this particular measure in reserve status as we did CMS?

MEMBER JARRETT: I am voting yes. You can't see my hand.

CHAIR GUNNAR: Okay, somebody take that --

MS. WINKLER: I get 18 yes.

CHAIR GUNNAR: So just for a -- can we -- do we pass that on to you as a recommendation?

MS. WINKLER: I think we can take that as a committee action. So we do have to go through the rest of the criteria, yes.

CHAIR GUNNAR: Okay, I was trying to avoid that but apparently we can't.

MS. WINKLER: They are different measures. They test differently. They have different data sources. They really could have different results at the level of scientific acceptability. There could be different issues.

CHAIR GUNNAR: So anymore discussion on -- do we pass on performance or

do we have to vote on it?

DR. BUSRTIN: Basically, I think we will just assign what you guys just did for the first measure for these particularly categories and let you move on to reliability. How about that?

CHAIR GUNNAR: Great. Dr. Moss.

MEMBER MOSS: So, just in the interest of time, I don't think there is anything really to add to the reliability here that wouldn't apply from the other measure. This does require the anesthesiologist to personally answer the question or check a box and then that box needs to be translated into the electronic record. The developers have shown in the measure that that is possible and can be --

MS. WINKLER: The important characteristics of looking at testing results for reliability and validity, as well as the measure specifications is what are the data

sources, and how was it tested. Was it tested at the level of the data element or tested at the level of the major score? There is no reason to think those were the same as the prior measure. They each function independently. So, you need to assess reliability on those criteria, on the results of the testing, as well as any comments about the measure specifications.

CHAIR GUNNAR: Any other discussion on reliability? Dr. Markman.

MEMBER MARKMAN: My question is -- or my issue is well, we have one that is a data collection but we have one that is an individual recording. And why only 50 percent of the anesthesiologists participate? What is the crux of it in terms of why don't they do it? Why do you have such a --

CHAIR FLEISHER: I mean one percent given the burden -- until NACOR was developed, the burden to actually do the work to submit was

greater than the value of doing it for billing companies. So, it was that simple.

While the hospital extracts SCIP, it is such a small percentage, because they can even do a random -- SCIP is a random sample of all cases. This would be G-codes in a much larger state when it is that simple.

I think with the establishment of the registry, that will change.

MS. WINKLER: So for everybody's information to help with this criterion, how was this measure tested? Was it tested at the level of the measure score? Did we do reliability testing of the performance results or were there testing done at the level of the data elements?

MR. POPOVICH: Well, as was stated, it is that the provider actually check a box and it was submitted successfully.

So, it was the performance score of the five percent file, as well as the NACOR

scores that are provided in the charts, in the data.

MEMBER MOSS: So, my interpretation of the 50 percent question, and developers please correct me if I read this wrong, was that the 50 percent issue is the fact that this is a voluntary reporting system and some people choose to participate and some people don't. But when it is reported, it is done so in a reliable and accurate fashion and stands up to auditing.

MR. POPOVICH: Yes.

MS. WINKLER: We have got a lot of data on the actual performance result but the testing would test reliability. And there should be some statistical assessment of the reliability of the measure, common -- like a signal-to-noise analysis, some kind of a statistical assessment of that reliability. Where would I find that?

MR. POPOVICH: The signal-to-noise

assessment has not been provided in this document.

MS. WINKLER: So, we don't have a testing at the level of the measure score. So, do we have testing results for reliability at the level of the data element?

CHAIR FLEISHER: Rick can answer the question.

MEMBER DUTTON: I can speak to how the data is collected in the registry, the National Anesthesia Registry from which a lot of this report comes. This is harvested from the records of the participating institutions in anesthesia practices and it is reliable at the reporting level.

You can see in the data fields provided that it matches quite closely. In fact, it matches almost exactly with what is reported in CMS and the Medicare data.

MS. TIERNEY: Sorry. I just wanted to comment on the reliability question.

So, this measure used to be a PCPI measure and we worked with ASA to transfer the stewardship over.

And at some point a few years ago, maybe two years' ago, we provided testing data, reliability testing data for this measure. So, we could provide that information to ASA and you could, potentially, submit that for consideration, if that was possible.

I just wanted to make sure everybody was aware that has been done. We had submitted it before. I know it wasn't part of this submission but it is an option, is possible and NQF would allow that.

CHAIR GUNNAR: Any additional discussion about reliability.

MS. WINKLER: I am just wondering without that, that really you can't evaluate reliability. If that was submitted a few years ago, we might be able to find it before tomorrow and take a look at it.

I am hesitant about pushing too much to tomorrow but, nonetheless, I think it is going to be difficult for you to make an assessment of reliability without the data.

So, if you want to table this either to tomorrow or we do have a follow-up conference call on the 9th, we may end up tabling it to that point, too. But at least you would have, it sounds like, information to work with, which you don't have now.

CHAIR GUNNAR: And that would extend to validity as well?

MS. WINKLER: Yes, I presume.

MEMBER SAIGAL: It sounds like either we table it or kill it. Because the answer here it is insufficient then, it wasn't submitted.

So, I, personally, believe that there probably is data to look at. It just wasn't properly managed.

So, I think we should table it and

give it a chance because it is an important measure. But I don't know what you think about that.

MS. WINKLER: Is anybody opposed to tabling it until we can see if we can capture that data? I don't see any indications for that. Okay, thanks.

Okay, so to keep us back on agenda, I think the next one we want to go back to is 0528. And we are back to the CMS SCIP measures. And this is the antibiotic selection for surgical patients.

So, Dale, are you still with us?

MR. BRATZLER: Yes, I am. So, very briefly, this is a performance measure that looks at selection of antimicrobial, based on the type of operation being performed. The measure is continuously updated to be consistent with published guidelines on surgical antimicrobial prophylaxis.

There have been, actually recently,

quite a few additional studies that have shown that antibiotic choice probably is very important with respect to patient surgical outcomes, particularly infections. And in fact, I think the literature basis perhaps is stronger now than it was back when the measure was first put into place.

To a certain extent, this measure and the next one that you will discuss around discontinuation represent, to a certain extent, antimicrobial stewardship measures because what we found when we originally looked at performance on this metric was that a lot of people were using broad spectrum antimicrobials, which really weren't recommended in guidelines and really have not been shown to improve patient outcomes.

So, I will be happy to answer questions.

CHAIR GUNNAR: Who is the discussant? Barry.

MEMBER MARKMAN: As discussed, this was a process measure and it was originally endorsed in 2009 and we endorsed it in 2012. And I think it significantly is different than the previous measure that is based upon a finite concept of giving the antibiotics within the hour.

So, if you look at what -- it is well written. I am a great fan of these CMS measures. It says there was strong evidence on which operations need to get an antibiotic, not just strong evidence on the best antibiotics.

So this is, as Dale said, this is an ongoing process. And I am going to argue later that I think there is a performance gap because once we say it is low -- I mean the reporting is great. The evidence is there. And then I will bring up another discussion after we talk about the evidence.

But I think that should not go into a reserve status because it is an evolving

process. And it is specifically looking at antibiotic selection for specific operations.

So, we can talk about the evidence and vote on that but I really believe that even though there is great reporting and it is up to 99 percent, that it is an ongoing measure that needs to be continued and not put a reserve status.

So, will take it step-by-step and then I will make my argument. And then I have a question for the developer after we vote on the evidence.

MS. WINKLER: So, how would you summarize the evidence?

MEMBER MARKMAN: Strong. I mean --

MS. WINKLER: Do we have a systematic review or a clinical practice guideline?

MEMBER MARKMAN: Yes, we have Level 1 evidence-based medicine starting from a

historical article and going along with all of the other measures. I mean, their statement is true. You give antibiotics prior to a surgery and your SSI rate decreases.

And I misspoke. There is very strong evidence for that. So, I would rate it high.

MEMBER JARRETT: This is Mark because I was the secondary discussant on this. And I agree completely with the statement. I think because it represents antibiotic stewardship, which is really just rolling out across the country and, therefore, which antibiotic for which story, based on what is out in the community, may be going to change.

I think to leave this there is very important because, otherwise, my fear is that people will just keep giving the same things four years' from now when it is not appropriate.

CHAIR GUNNAR: Any other? Dr. Sawin.

MEMBER SAWIN: Question on how are the recommendations for a specific antibiotics updated? Is it based on literature or consensus?

MEMBER MARKMAN: Well, I mean, once you go through the measure they have specific antibiotics, starting with the cephalosporins, as well as alternatives.

But at this point, they are still evaluating each recommendation. But within the body of the measure, there is a table that explains or details which antibiotic for which operation. And that is the basis of their measure.

MEMBER SAWIN: But based on data or on consensus?

MEMBER MARKMAN: It started with the systemic review from the Bratzler article. And that was, it is -- and then there is an article in 2013 that was referenced, that Dale referenced as an update. And you can go

through it. In fact, she is rolling it on the screen here.

CHAIR FLEISHER: Barry, could we actually get Dale, --

MEMBER MARKMAN: Yes.

CHAIR FLEISHER: -- since he shares a lot of these comments, to comment on the process by which are these clinical guideline based and what is the process of developing the clinical guidelines, Dale?

MR. BRATZLER: Yes, to the performance measures -- CMS makes the point that they don't create guidelines. They develop performance metrics that are consistent with guidelines.

And so, Lee know as well as anybody, we have technical expert panels that meet quarterly to review measure specifications and when new guidelines are published by anyone, we evaluate those guidelines and update the performance metrics.

So typically, right now, the performance metrics are updated once or twice a year, depending on the frequency that CMS has the resources to do the updates. But they are continuously updated.

CHAIR GUNNAR: Any other discussion? Dr. Yates.

MEMBER YATES: I am looking at the chart of what the appropriate antibiotics are and what the criteria are and I have several comments to make that reflects practice in 2014. You might be able to help me with which one is reference D for orthopedic procedures.

And my principle, my primary problem is with not so much the cephalosporin or cefazolin but my primary problem is the alternative antibiotics. For instance, penicillin allergy, which creates hives or some benign allergic response as opposed to anaphylaxis, in our practice and in most people's practices, this is not a

contraindication to use a cephalosporin.

Number two, I'm not sure if this represents current practice in terms of MRSA screening, which has become prevalent in at least orthopedics and I think cardiac as well, but we routinely screen for MRSA. And if that is the case, we will administer both vancomycin and cephalosporin, at least in our practice.

And the third observation I would make is that diagrams for a lot of hospitals for the static drug, which is clindamycin, doesn't even support the use of clindamycin. And there was an abstract from one of the hospitals local to our area at our recent academy where they all of the guidelines, the old guidelines of using clindamycin as one of the alternative antibiotics, and they got burned and then they had a higher infection rate.

So given that, I just wonder what is Reference D and when was it written? And exactly what is the level of evidence for D,

other than perhaps a consensus opinion?

MR. BRATZLER: I don't have the document in front of me to tell you what Reference D is.

MEMBER YATES: Well, it is annotated D and I can't find D.

MS. WINKLER: If you scroll down to the bottom of the table, you will find A, B, C, and D laid out.

MEMBER YATES: Smaller than my eyes can see.

MS. WINKLER: D says for procedures in which pathogens other staphylococcus or streptococcus are likely, an additional agent with activity against those pathogens could be considered. For example --

MR. BRATZLER: Yes, so I know that statement well.

So first, I generally agree with everything that was said. Remember, this is a performance metric that is rolled out across

4,000 hospitals. And so, the antimicrobial pattern, the antibiogram in one hospital certainly may not look anything like the antibiogram on another one that is across the country.

So, the performance measure just simply reflects what is represented in guidelines. So, as you point out, cephalosporins tend to be the drugs of choice for most forms of surgical antimicrobial prophylaxis.

And we completely agree, and if you read the guidelines, we can't put it in a performance measure but if you read the guidelines, we make it explicit that even if a patient reports a beta-lactam allergy, but it was not a serious life-threatening one, they should still use the cephalosporin. And in fact, if you look at the algorithm for the performance measure, the way the algorithm works is if the patient, let's take a hip

arthroplasty, the patient comes in and says they are allergic to penicillin, they had a skin rash in the past. The surgeon elects to give cefazolin.

The case passes because the performance measure looks first at the antibiotic given and if a first generation cephalosporin was given to that particular patient, the case passes, regardless of how the hospital answered the question about beta-lactam allergy.

If they decided to use a drug such as vancomycin alone, then the algorithm does look at that beta-lactam allergy question to see if they documented a rationale for using vancomycin. So, beta-lactam allergy might be one. Positive MRSA screen might be another.

We don't look at combination antibiotics. My personal preference and recommendation when an orthopedic surgeon asks me, if they have an MRSA positive patient, my

recommendation is typically to give a dose of vancomycin and a dose of cefazolin, just as you discussed. We don't look at both of those because, if they documented the rationale for vancomycin or if they give cefazolin, the case will automatically pass the performance measure. So, we don't have to ask the additional questions and data collection.

CHAIR FLEISHER: So, just to be clear, and I am happy to be corrected by my colleagues to the left. These are from an evidence standpoint. That is actually specifications in some way, Dale, what you have just defined. These are updated once or twice a year, based upon the best available evidence with input from the specialty societies, sitting on the technical experts.

MR. BRATZLER: That is correct, with a review of guidelines.

CHAIR FLEISHER: We can all argue over evidence and there is an entire separate

evidence committees and others but we should really focus on whether or not -- the committee can spend a lot of time debating the evidence and that, I don't think is our primary role by really saying whether the evidence supports the development of the measure.

MEMBER YATES: And I am just saying in terms of Level 1, you are dealing with consensus statements for the most part. And I say that, having sat on the Periprosthetic Infectious Consensus Group.

CHAIR FLEISHER: Right. And so I think that is a good point and it will be part of your votes with regard to how you feel the evidence is there.

MEMBER YATES: And I only say that because it was raised as being the best antibiotic to give. This was the whole premise for this being a better measure than other measures.

CHAIR FLEISHER: Sure. There is a

paper today in JAMA I urge you to read about that.

MR. BRATZLER: Yes, and I think we acknowledge them in the guideline that we published in 2013 on the antimicrobial prophylaxis. When you actually look for randomized controlled trials that are of one antibiotic compared to another, the data is not very rich. Lots of observational studies, but RCTs, there is not.

But for some of those things, there never will be an RCT.

CHAIR GUNNAR: All right. So, are we okay with moving on for a vote on evidence? Okay.

MR. SANCHEZ: Voting will not begin for subcriterion 1a, evidence. One is high, two is moderate, three is low, four is insufficient evidence.

The timer starts now.

(Voting.)

MR. SANCHEZ: We have eight high; 14 moderate; one low; zero insufficient evidence.

CHAIR GUNNAR: Dr. Markman.

MEMBER MARKMAN: In terms of the performance gap, I am going to reiterate what I said previously is that the reporting is up to 99 percent. It is the statistical data. But I am going to argue that the performance gap is still moderate to high because it is a continually evolving process and it is continually updated. Because if we say that the performance gap only because of statistically reporting, then we are going to end up in a reserve status.

So, my comment is that I think this is a great measure. It should be ongoing. And the performance gap really is based upon the continued updating of the information.

CHAIR GUNNAR: I guess the only argument to that is that we have just heard that

the recommendations are updated every six months and the compliance is viewed over a long period of time. No? Did I miss that?

And so, people are actually going to the reference and modifying practice to be compliant, whether we measure this as a performance or put it in reserve status. So, the field has accommodated the on the ongoing update and accommodation for new evidence.

For those on the phone, it has begun to rain. And that is an unusual thing when you are in D.C.

CHAIR FLEISHER: I would actually argue, Bill, that you could view it either way. You could view it that payment is driving people to make sure they stay updated or, and if payments stop, they wouldn't stay updated. That, I think, was what I heard Barry said. I would be curious what Dale has to say.

MR. BRATZLER: I am not sure I completely understand, Lee. But I think we

have documented over time that when measures go into the public domain, performance improves rapidly. And I think, to a certain extent, payment has driven part of that.

CHAIR GUNNAR: But the reverse of that is we don't know, and it is an assumption that when we take our eyes off the ball or pull one of these and put it in reserve status, that suddenly, there will be a slippage. But I don't if that -- we don't have any evidence to that effect.

So any other discussion? Collette.

MEMBER PITZEN: Collette Pitzen. I just wanted to make a comment. It probably isn't going to be very popular among all the surgeons here but when I am looking at a measure that I want to put forward to improve care and I have something that is 99.9 percent in compliance as it is right now, that measure is not demonstrating any gap to go anywhere.

Maybe in the future but does it really justify being used as a national measure and the resources that it takes to require it?

MEMBER MOYER: I was going to say something similar. I mean we have a pay-for-performance program and I wouldn't put this measure in it. I mean what would I pay for it?

CHAIR GUNNAR: You are paying everyone.

MEMBER PITZEN: Doing what is being done today. Exactly. So, it isn't really useful to me from that perspective.

CHAIR FLEISHER: So actually, I have one question for Dale and I don't know the answer. When there is a change in antibiotic choice for a procedure, does this go out of compliance until it goes back into high compliance? So that would be the only -- because if you are saying things change over time and it is not hardwired what the choice is,

do you have any data for that?

MEMBER CIMA: I would like to clarify what I mean by changing. What is happening is more and more antibiotics have been added as being considered appropriate as part of the choices. They are not taking any away and putting new ones in.

So for colorectal surgery, one paper out of Harvard, small study, said you could use ceftriaxone as a prophylaxis. And that, somehow, got put in as opposed to what standard prophylaxis means is skin organism. But they got it through, so they just added it to the possible choices.

MR. BRATZLER: Well, it was a little more complicated than that. There were institutions around the country reporting gram-negative surgical site infections resistant to all of the first and second generation cephalosporins. So, the hospital had a choice of using ertapenem or some other

option. And if you read the guidelines, we actually discuss that issue.

So, we have added antibiotics but I will tell you that if you look at the current literature, the second generation cephalosporins are the useful agent for colorectal surgery appears to be dropping fairly considerably in most of the studies that have been recently published. So, I don't know how much longer those agents will be recommended.

CHAIR GUNNAR: So, I think we can vote on performance gap, unless there is any further -- so, let's -- the reality is is that we have 99 percent compliance. So, shall we vote on performance gap?

MR. SANCHEZ: Voting will now being for subcriterion 1b, performance gap. One is high, two is moderate, three is low, four is insufficient.

Voting timer starts now.

(Voting.)

CHAIR GUNNAR: There are some undecideds out here still. I need a couple more. Okay.

MR. SANCHEZ: Zero for high; five for moderate; 15 for low; one for insufficient.

CHAIR GUNNAR: So, should we vote, a hand vote for those who would wish to carry this forward as a potential reserve measure? Hands up.

(A show of hands.)

CHAIR GUNNAR: Okay.

MEMBER JARRETT: Hands up on the phone.

CHAIR GUNNAR: So 16, a majority. So, we will carry through.

Dr. Markman.

MEMBER MARKMAN: In terms of high priority, I think that we discussed this in other issues. And in particular, I think this is a high priority.

CHAIR GUNNAR: Any other discussion? We will vote.

MR. SANCHEZ: Voting will now begin for subcriterion 1c, high priority. One is high, two is moderate, three is low, four is insufficient.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have 11 for high; nine for moderate; one for low; zero for insufficient.

CHAIR GUNNAR: Reliability.

MEMBER MARKMAN: Reliability based upon the data set points, I would say is high.

CHAIR GUNNAR: Any discussion? Hearing none, please vote.

MR. SANCHEZ: Voting will now begin for subcriterion 2a, reliability. One is for high, two is for moderate, three is for low, four is for insufficient.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have 11 for high; 10 for moderate; zero for low; and zero for insufficient.

CHAIR GUNNAR: Move on to validity, Dr. Markman.

MEMBER MARKMAN: My recommendation is that the validity is high.

CHAIR GUNNAR: Any discussion? Hearing none, move to vote.

MR. SANCHEZ: Voting will now begin for subcriterion 2b, validity. One is for high; two is for moderate; three is for low; four is for insufficient.

Voting timer starts now.

(Voting.)

MR. SANCHEZ: We have 12 for high; nine for moderate; zero for low; and zero for insufficient.

CHAIR GUNNAR: Move on to

feasibility.

MEMBER MARKMAN: Feasibility is high.

CHAIR GUNNAR: Any discussion? Okay, go to a vote.

MR. SANCHEZ: Voting will now begin for criterion 3, feasibility. One is for high, two is for moderate, three is for low, four is for insufficient.

The voting timer starts now.

(Voting.)

CHAIR GUNNAR: We have 13 for high; eight for moderate; one for low; zero for insufficient.

CHAIR GUNNAR: And usability.

MEMBER MARKMAN: Usability is high.

CHAIR GUNNAR: Any other discussion? Hearing none, we will take a vote.

MR. SANCHEZ: Voting will now begin for criterion 4, usability and use. One is

high, two is moderate, three is low, four is for insufficient information.

The voting timer starts now.

(Voting.)

CHAIR GUNNAR: Go ahead and vote again. Oh, now its picking them up.

MR. SANCHEZ: We have 15 for high; five for moderate; two for low; zero for insufficient evidence.

MEMBER ASHER: I just want to ask a question for reference in some other questions we are going to be looking at here. So, that second criterion, I don't know if you can go back to the last slide. So, the 4b, you know we hadn't really talked a lot around that point. So I am just wondering how heavily should that be weighted?

I mean there is usability here. But if the progress from year to year has been relatively low with these things, then how much does that weigh into this particular thing? I

just don't know that we have discussed that too much.

MS. WINKLER: Well, improvement, actually, for this criteria, is to really understand how effective this measure is at driving improvement. The problem is when you are 99 percent, you really can't expect to see much change.

MEMBER ASHER: See, that is the point I'm trying to make. So, how --

MS. WINKLER: Yes, so this kind of is a corollary to the gap. I mean the two are very much related. If when it was originally endorsed, performance was at 60 percent and it is now at 80 percent, that tells you something.

MEMBER ASHER: Yes, I guess what I am saying is I haven't seen as much a correlation between those two areas. So for example, in the gap, we have been voting things extremely low. But here, we have been voting them extremely high. And so if that is

relevant, it seems to me there should be more a correlation between those two areas.

MS. WINKLER: I think that is a correct thing. Is Karen here? No. We will certainly discuss that. But realize there are two other criteria. So, it kind of gets buried in there with the unintended consequences and accountability uses. So, it isn't as pure.

CHAIR GUNNAR: So, we will vote for placing this measure in reverse and reserve status. Any discussion? Hearing none, we will go to a vote.

MR. SANCHEZ: Voting will now begin for potential for reserve status. One is for yes, two is for no.

The voting timer starts now.

(Voting.)

CHAIR FLEISHER: Do people want to go to the next measure or would they rather go to the companion measure on PCPI?

What is the companion measure?

MS. WINKLER: Well, it will be the clinician level for selection.

CHAIR FLEISHER: Companion, why don't we just do the same one? Can't we shorten it?

MR. SANCHEZ: We have 20 for yes and three for no.

MS. WINKLER: Okay, PCPI's measure for clinician level 0268 is sort of the same subject around selection of prophylactic. Would it be easier to do that right now because we have just done selection, you think? Okay, Sam, you are good with that?

CHAIR FLEISHER: So, I am going to disclose that I was, in the 2006 PCPI but have not been involved since. So, therefore, I am going to vote.

MR. BRATZLER: This is Dale Bratzler. I am here, also representing I was on the committee with PCPI. So, I can represent the PCPI measures also.

MS. WINKLER: Who wants to introduce the measure briefly?

MR. BRATZLER: I will do it very briefly.

The percentage of surgical patients aged 18 and older undergoing procedures with the indications for a first or second generation cephalosporin prophylactic antibiotic that have an order for -- this is the PCIPI measure of antibiotic selection.

The data sources, the clinicians themselves, the data can come from usually administrative claims with physicians submitting the data. This is a part of the PQRS program.

I know AMA representatives are on the call. They can give you a better idea of the actual number of physicians that are actually reporting this. But the measure applies both in ambulatory care hospital and acute inpatient surgeries.

So, again, it is a limited number of operations for which first or first or second generation cephalosporins would be recommended for antimicrobial prophylaxis and it is an order for the antibiotic to be given.

CHAIR GUNNAR: Who is the discussant? Dr. Sawin.

MEMBER SAWIN: So, this is an improvement of a measure that was first accepted in 2008 and expanded to include the second generation cephalosporins in the numerator. The evidence is pertinent to our prior discussions. It is pretty well documented about the importance of appropriate prophylactic antibiotics.

They also widened or expanded the denominator exceptions, allowing for documented medical reasons and patients who had been receiving antibiotics for other reasons.

So, I guess the evidence for this, as all the other antibiotic measures is fairly

strong with lots of Level 1 data to support it.

CHAIR FLEISHER: Any unique questions from what was not discussed during the previous discussion?

MEMBER SIPERSTEIN: I was the secondary reviewer on this and the only small caveat that was mentioned was that in order to qualify, you just need an order or you could show that you administered the antibiotic.

CHAIR FLEISHER: We can discuss that under specs when we get to that.

Can we vote?

MEMBER DUTTON: One question. Just to point out that if I read it right here, only about 29 percent of eligible people do this report. Only about 29 percent of eligible professionals reported in 2011.

CHAIR FLEISHER: Yes, I assume that that is the same issue that was discussed previously, that it is a function of -- it was a one percent to volunteer.

MS. WINKLER: Are we talking about participation of PQRS? Yes. So, that will be later down. It is not evidence.

MEMBER DUTTON: You want to hit that later?

CHAIR FLEISHER: Yes.

MEMBER DUTTON: Okay, I will be later.

CHAIR FLEISHER: Just evidence. Let's vote.

MR. SANCHEZ: Voting will now begin for subcriterion 1a, evidence. One is for high, two is for moderate, three is for low, four is for insufficient evidence.

Voting timer starts now.

(Voting.)

CHAIR GUNNAR: Let the record show that Dr. Ko will recuse himself from this series of voting.

MR. SANCHEZ: We have 15 high; eight moderate; zero low; zero for insufficient

evidence.

CHAIR FLEISHER: Next.

MEMBER SAWIN: In terms of performance gap, the PQRS data in 2008 showed that 62 percent were compliant with this measure and in 2010, it was up to 96.4 percent. I don't see any more current data, unless I am missing something. There was no data regarding disparities.

CHAIR FLEISHER: And this would be an appropriate time to discuss the percent of eligible reporting on this measure, if there is any comment.

MEMBER DUTTON: Yes. Just PQRS is a voluntary reporting system for physicians. It has been made physicians eligible for incentives up to now but does not involve penalties.

The highest reporting group of physicians, anesthesiologists and emergency medicine physicians are among the highest --

and I shouldn't say just physicians, by the way, it is all eligible providers -- is around 50 percent. And other specialties are all lower than that.

MS. TIERNEY: If I could just make a comment, too.

CHAIR FLEISHER: Please.

MS. TIERNEY: I know that someone had asked if there was more recent data available. And I just wanted to clarify. So the information that we presented, we get some confidential data from CMS at the decile level. So, that is what we have provided. But they also publish these experience reports. And they just published one for 2012 data that had the performance rate for this measure at 92.9 percent. That is the average. They don't break it down further than that.

And then also, I just wanted to emphasize so the 29 percent of eligible professionals reporting is for the program

itself. This measure, in 2012, had nine percent of eligible professionals reporting.

Again, that is a function of the program and whether or not, given that the incentive is so small and the additional burden in reporting and the requirements around that for the program. But I just wanted to highlight the newer information that we just recently got since the submission was submitted.

CHAIR FLEISHER: Fred, did you have a comment? No, I guess not. Any other comments? Are we prepared to vote?

MR. SANCHEZ: Voting will now begin for subcriterion 1b, performance gap. One for high, two for moderate, three for low, four for insufficient.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have two for high; 12 for moderate; nine for low; zero for

insufficient.

CHAIR FLEISHER: Okay, we keep going.

MEMBER SAWIN: In terms of reliability, the data source is administrative data. I did mention that the numerator had been expanded from the old measure, as had the denominator been clarified.

I have some reservations about how well-documented the medical exceptions are in administrative data but the liability testing was done and showed a pretty good signal-to-noise ratio. So, I think the reliability is moderate to high.

CHAIR FLEISHER: Any comments from the developer? Dale, any comments on the exceptions from your perspective?

MR. BRATZLER: No. I don't have -- let me see if I have got -- there was a question about -- let me make sure I understand the question before I open my mouth.

MEMBER SAWIN: When we had our phone call, there was some discussion about whether people would adequately document the medical exceptions so they would be removed from the denominator, using the administrative base, whether those data would be readily available.

MR. BRATZLER: Right. Certainly AMA probably can give a better thought about that. I mean, that is how the clinician essentially reports the measure, usually with claims. So, I can't imagine them very often reporting that they failed the measure if there was an exception or an exclusion. But I don't have any specific data.

CHAIR FLEISHER: Any other comments? Okay, let's vote.

MR. SANCHEZ: Voting will now begin for subcriterion 2a, reliability. One is for high, two is for moderate, three is for low, four is for insufficient.

The voting timer starts now.

(Voting.

MR. SANCHEZ: We have seven for high; 14 for moderate; two for low; and zero for insufficient.

CHAIR FLEISHER: Okay, next.

MEMBER SAWIN: As far as validity, the data or the measures, rather, were validated with multiple specialty organizations and societies. And a face validity test was done, which was good. And also the same societies and organizations were asked about whether or not this measure would adequately discriminate poor or good quality. And there was a high degree of concurrence.

The one concern that was previously mentioned by Alan was that the numerator is actually whether the order was written not that the antibiotic was actually administered. So, that was my concern with validity.

Otherwise, it is moderate validity.

MEMBER SAIGAL: Can I comment? This is for the physician level. So, I guess the physician can't do more than write the order. Right?

MS. WINKLER: Just to review the criteria, if face validity is the only assessment of validity. The highest rating possible is moderate.

CHAIR FLEISHER: Are we ready to vote?

MR. SANCHEZ: Voting will now begin for subcriterion 2b, validity. One is for high, two is for moderate, three is for low, four is for insufficient.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: Are we still waiting on one committee member?

(Voting.)

MR. SANCHEZ: We have three for high; 15 for moderate; five for low; zero for

insufficient.

MEMBER SAWIN: Feasibility. The data source is administrative. Other than the concerns mentioned about the documentation and the order versus administration, feasibility seems moderate to high.

CHAIR FLEISHER: Comments? Vote.

MR. SANCHEZ: Voting will now begin for criterion 3, feasibility. One is for high, two is for moderate, three is for low, four is for insufficient.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have six for high; 12 for moderate; three for low; and zero for insufficient.

MEMBER SAWIN: Usability and use. It is currently used by, as was mentioned, a relatively small number of providers currently. And so the high compliance with the measure might be skewed but there was a survey

of those participating and if 55 percent reported that they found the study to be satisfactory, I'm not sure how that compares to other PQRI measures.

I don't know whether the developer wants to comment on that.

MS. TIERNEY: I'm not following where -- could you point me to what you are referencing? I'm not following. Sorry.

(Pause.)

MS. TIERNEY: I know there is a lot of pages. Maybe you are referencing the validity testing results.

MEMBER SAWIN: No, there was also satisfaction of those who participated.

MS. TIERNEY: Oh, I do know what you are referencing. I apologize.

So, yes in the section, the testing attachment, we included information from the PQRS program about those who satisfactorily report the measure. That, again, is sort of a

function of the PQRS program and the requirements around it. So, you know, the program has changed over the years but I think currently the requirements are 50 percent. For individual measures, you have to report on 50 percent of your eligible patients. So, that would factor into satisfactorily reporting.

So, it is really a function of the PQRS program. It is probably not the best thing for us to include because I think it is kind of confusing and it really is, again, more of a function of the PQRS program than an indication of the measure properties or properties of the measure.

So, sorry if that threw you off a bit. I'm sorry I didn't follow initially.

MEMBER SAWIN: Misinterpreted.

MEMBER HANDY: Well, let me ask a question regarding that because it is pertinent to the next one.

I took that to mean this is the N

that you are actually evaluating. So, of the large number of eligible professionals, you really get down to where you are evaluating a single digit percentage point. Am I understanding that correctly?

MS. TIERNEY: So, as it relates -- so, I guess on the testing attachment 2b5.1, this is the section that I think we are referring to. And so, the N of the performance rate is on the number of professionals satisfactorily reporting -- actually no. It is on the ones reporting at least one valid QDC, quality data code, which is 6175, I believe. Does that sound right? Yes.

But it does -- this does sort of walk you through. I mean there is quite a number of eligible professionals but very few actually even tried reporting, 6100, which represented 8.9 percent of the total eligible professionals reporting.

And then of those, of the 6175, 3415

satisfactorily reported, which is 55 percent of the total -- of the 6175.

So again, it is a function of the PQRS program and the various requirements around what is considered satisfactory reporting.

Some of it has to do with putting the right code on the claim and so some of it might -- some of the unsatisfactory reporting might be a function of that. But it also might be related to, again, how many, if you were reporting on the number of patients you are supposed to report on, at least 50 percent of your eligible population, things like that.

MEMBER SAWIN: So, if I understand it, so only nine percent of the participants completed more than one report.

MEMBER HANDY: And only half of them did it satisfactorily. So, you are really talking about 3400 people, not 69,000 people.

MS. TIERNEY: Let me clarify, too.

I think it is obvious but I want to make sure. That is of the eligible professionals. So, that is not patients. That is the physicians who could have reported on this measure because they had the relevant codes. They were doing those procedures that are included in the denominator of the measure.

MEMBER GROVER: I think this kind of really bothers me because it is such a small percentage. And in the material we got, it said they can really select what patients they report, if that is really true.

So, I mean is this something that is likely to mislead us? That really worries me if you only have nine percent. What does this mean?

MS. TIERNEY: So, I think to Dr. Fleisher's point earlier, it is because of the incentives being relatively small and the program right now is voluntary.

I imagine over time, this, again, I

really feel is related to the implementation of the measure in the PQRS program, I imagine that over time you would see, given that they are moving to a penalty phase now for physicians who do not report, there will be a penalty applied, I think you will see reporting rates jump significantly and maybe be more representative, like what you see at the hospital level about how many eligible hospitals participate in the public reporting programs.

CHAIR FLEISHER: Larissa.

MEMBER TEMPLE: So just to follow up on that, I am struck by the fact that this is 2010 data and we are in 2014. And I know the program has definitely gained speed since then. So, could you give us a better sense of what the current rates are?

MS. TIERNEY: Sure. So, as I said earlier, we get confidential reports from CMS. The data is slow coming to us. So the most

current set of full complete data we have is from 2010 and that includes information of the performance rate at decile levels, which is what is requested in the NQF form.

Since we submitted these measures, CMS has come out with their experience report for 2012, which includes more recent information. And that is what I had alluded to earlier with nine percent of eligible professionals reporting. It is actually -- it went down from 2010 and 2011. It was 9.9 in 2010, 9.8 in 2011. It was 9.9 percent in 2012.

But again, I wouldn't necessarily -- I think that that is the CMS PQRS program and not specific to the measure. The measure is in use in that program. It is one of the examples of its uses. But the rate of -- we have no control over the rate of reporting. It is a matter of whether or not professionals choose to report on those measures.

Yes, please?

MS. KAYE: And I just wanted to highlight when you are looking through the experience report, that that low, that nine percent participation rate isn't unique to this particular measure.

You will find like we referenced earlier, some of the emergency medicine measures, they are really the high fliers there with 50 percent reporting. But there are quite a few. I would venture to say most of the measures have similar low reporting percentages. So, it isn't unique to this particular measure.

MEMBER PITZEN: I have a clarifying question. So, the low rate of the QBC that is being reported, is that reflective of the actual use of the CPT-2 codes that report the numerator? Because if that is true, then I have a concern about the reliability of the measure.

If people are saying I have these

patients I my population but I am choosing not to even tell you if I am compliant with the numerator. But maybe I am not interpreting --

MEMBER SAIGAL: Basically, though, what we are trying to look at is the usability here and how it could be used, how different organizations use these measures and care to comment under implications. Is it up to us, necessarily? This is like a CMS program that didn't encourage participation by providers.

So, if the measure is valid and it is being gamed by providers, that is a separate issue?

MS. WINKLER: Well, I think it is related. Because one of the things that we want to know, particularly for a measure that has been around for a long time is how usable it is for various stakeholders. Is this information meaningful? I mean it costs something to collect the information, calculate the data, and report it. So, is that

usable and useful? And how is it being used?

I think it is fair to ask why do we see a relatively low uptake in the PQRS program. So, I think they are related to how useable you perceive the measure to be because you do have a bit of a track record for this measure, unlike say a brand new measure.

CHAIR FLEISHER: Okay, shall we vote?

MR. SANCHEZ: Voting will now begin for criterion 4, usability and use. One is high, two is moderate, three is low, four is insufficient information.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have one high; ten moderate; 12 low; zero insufficient information.

CHAIR FLEISHER: So, it's gray. It is a gray area. Okay, which means we can go on to vote for endorsement. Okay? Unless

anybody has any objection, we are going to vote. Okay.

MR. SANCHEZ: Voting will now begin for overall suitability for endorsement. One is yes, two is no.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have ten yes; 13 no.

CHAIR FLEISHER: So, we can have two choices. One, to stand up for like five minutes, if people think that is necessary, or I can start an interesting discussion. Because we have got an hour left. Are people okay to continue or do they want to stand up?

Okay. So, the question is, and can I do it now? So keep doing the measures and then have the discussion. So, think in your mind what would have happened if we didn't have reserve status. That will be a question we want to have either at the end of today or

tomorrow. So maybe we have to sleep on that one or we can do it over dinner. But we do want to give some insight.

Okay, we are now back to prophylactic discontinuation.

Dale, are you still on the line?

MR. BRATZLER: I am.

CHAIR FLEISHER: Measure 0529.

MR. BRATZLER: All right. So, this particular performance measure also introduced with the first set of antimicrobial prophylaxis performance measures back in 2002 and rolled out nationwide in about 2005, looks at discontinuation of antibiotics after surgery. The current performance measure looks for most operations, whether the antibiotics are stopped within 24 hours after surgery, 48 hours for cardiac surgery.

There is a lot of misperception that runs around about this measure and many papers have been published that have been trying to

link performance on this measure to surgical infection rates, but that is not what this measure is designed for.

There has never been a study that showed that stopping antibiotics at any particular time frame after surgery impacts the surgical infection rate. This is a measure of antimicrobial stewardship. And when we first started this measure, the national performance rate on the measure was about 41 percent.

So, this is a measure of stewardship. I would argue it is, perhaps, the most effective antimicrobial stewardship measure that has ever been rolled out nationally and focuses on stopping antibiotics after surgery.

Well, the point I wanted to make, we read through the comments of the committee before the call and there was a misperception about patients who have a second procedure done after the first operation.

So the classic scenario that came up frequently was a patient that had coronary artery bypass surgery. And we looked at all the antibiotics given to that patient for 72 hours after surgery to see if they were stopped within 48 hours. And what was happening periodically with a patient would maybe go into third degree heart block and require a pacemaker placement on the third postoperative day after their bypass surgery. Well, of course, a single antimicrobial dose is recommended for that pacemaker placement.

So, we have an exclusion in the measure for patients who have an operation with an incision and a general or regional anesthetic agent within 48 hours from most operations, 72 hours for cardiac surgery to address that specific circumstance where a patient may have to have a second procedure that also requires the single antimicrobial prophylaxis.

And then the last I will make is remember that patients are excluded from this measure if they have a documented infection pre- or intraoperatively or in the first 24 to 48 hours after surgery.

CHAIR GUNNAR: Who is the discussant? Dr. Asher.

MEMBER ASHER: So, the level of analysis for this particular measure is the facility level with respect to evidence. Much of the evidence to support the measure is found in a 2013 systematic review. There was an update of a previous therapeutic guidelines effort. Specifically in this particular review, the authors give a Level 1 recommendation to stop all antimicrobials at the end of surgery, based on review of about 39 RCTs.

And in studies we are also excited to support the concept that prolonged prophylaxis was associated with increased risk

of acquired antimicrobial resistance. And there is no data to support the continuation of antimicrobial prophylaxis until all indwelling drains were removed.

And so, based on the information provided, I would rate the evidence of the highest score.

CHAIR GUNNAR: Reva.

MS. WINKLER: Yes. Dale, I just want to point out that in your submission, under the evidence, what is listed is the diagram relationship between this process and outcome is a relationship of decreased risk of surgical site infection. And that seems to be the focus of the information submitted around evidence.

And you have just stated that that is not accurate. So, it makes it quite difficult for the committee to evaluate this measure.

MR. BRATZLER: So, that is definitely not accurate. When you look -- I

don't remember seeing that in the submission form but there is no relationship.

So, and we are currently updating the HICPAC guidelines with the surgical site infections. And we reviewed the literature on all the studies, RCTs only, that looked at various durations of antibiotics after surgery. And what you see consistently in those studies is the duration of antimicrobials after the operation has no impact on surgical site infection rates.

So in other words, there is no benefit from continuing to give doses after wound closure.

So our draft recommendation, it is not final yet, in the HICPAC guidelines is a 1a recommendation to stop all antibiotics at the time of incision closure.

But is a broad misperception and I am sorry that got into the submission form, but stopping antibiotics or continuing them

doesn't change the postoperative surgical site infection rate.

So, this is a measure of stopping unnecessary antimicrobials.

But that outcome is much, much more difficult to measure. And that is the whole concept of can you reduce resistance and reduce the risk of c. difficile infections by stopping unnecessary antibiotics after surgery.

MEMBER ASHER: Having been involved in a lot of neurosurgery, --

MR. BRATZLER: This is a measure -- sorry to keep coming, but this is a measure where I feel very strongly that it has been a struggle to get to where we are. And I have a huge concern that if we stopped measuring performance on stopping antibiotics, it would be really easy to slip back into old habits.

CHAIR GUNNAR: Dr. Asher, additional comments?

MEMBER ASHER: Just that the way I

read this was it was a level 1 recommendation to not do something. There was strong evidence that there was no benefit to continuing antibiotics past that point. So, I agree with what is being stated.

MEMBER CIMA: I agree with Reva. There is some evidence later on about that. But if you look at the rationale, it really is about antibiotic stewardship. It talks about c. difficile infection and antibiotic resistance.

So yes, I think there was a cut and paste issue here in the details but I agree with what Dale was just saying.

But based on the recent recommendations that are going to be coming out from the CDC and HICPAC and stuff, why does this measure not go with those recommendations?

DR. HUDSON: Well, I want to make it clear that the HICPAC guideline is not final yet. It has still been in public comment and

the literature review is being updated. And it is going back to the committee in July. So, I want to make it very, very clear it is draft only. It is not final at this point.

And there was some push back because there is not a RCT for every single type of operation. There are a bunch of RCTs but not for every type of surgery. So, we always get push back from some specialty society that says wait a second, there is no RCT for my particular special operation that I do.

So, I just have to be cautious. What we put in the ASHP guidelines was that antibiotics should be stopped within 24 hours for all operations. We left it at that.

CHAIR GUNNAR: This is Bill Gunnar. So, the concept then is that this is best practice with a much broader view than the surgical site infections. It is actually about antimicrobial stewardship.

And then the evidence, so we are

just trying to drill down on the evidence for this particular measure is somewhat -- if we looked at it strictly from SSI, there would be very little evidence. We are just having -- I am having difficulty, I will speak for myself, framing this in relationship to overall antimicrobial stewardship. And it is the evidence to support that in relationship to this particular measure. I hope that was clear.

MEMBER ASHER: I think that is correct. I mean I didn't seen anything in this that really was looking at SSI. This really is just a stewardship issue.

CHAIR GUNNAR: Any other discussion? Dr. Yates.

MEMBER YATES: In addition to the clinical evidence for stewardship versus surgical site infection rates, you do have a cost issue. There is a lot of savings to be had across the country for not giving prolonged

antibiotics longer than necessary.

I would argue, and as he said, there is push back from specialty societies. But in particular those of us that put sterile implants in held our breath as went from a traditional 48 hours to 24 hours in a national experiment, which did not increase the rate of infection. But there was no data to show that that wasn't going to happen in terms of power of evidence. Which really in terms of pulling back would have been a standard of care, should have done with more power than just saying we don't see it happening.

That is irrelevant to the rest of this conversation in terms of what they are doing in terms of going to one dose.

But I would say that the cost is part of the impact of this, that there is some savings to be had.

MS. WINKLER: Dr. Gunnar, in response to your difficulties in trying to deal

with the information presented and rating it according to the criteria, let me point out to you that you should look and see, and I would suggest that there really is no evidence presented for what they describe as the real benefit of the measure in terms of antibiotic stewardship. So, that should drive your rating on evidence.

However, we haven't had to deal with this today but if you do rate the evidence low or insufficient, maybe more appropriate, the committee has an option of then saying that we will make an exception to the evidence requirement and say that we will let it go, even though we haven't documented the evidence.

So, there is a way through this if you like the measure but we truly don't have any of the evidence laid out here.

CHAIR GUNNAR: What you are saying, let me reframe that, Reva. If we fail to pass this evidence component, we can go on to

evaluate the other components of this evaluation.

MS. WINKLER: Only if you invoke the exception of evidence.

CHAIR GUNNAR: But if we get to the next validity or reliability and we vote -- I mean at some point, we can get past evidence but it may or may not stand on then another --

MS. WINKLER: Absolutely.

CHAIR GUNNAR: Okay.

MEMBER KO: Is there an exception to every one of these?

MS. WINKLER: No, just evidence.

CHAIR FLEISHER: I was on the evidence committee for CSAC and we spent a lot of time discussing this. And it really has to be a -- it is a lack of evidence in something that you think is important enough.

MEMBER KO: I mean all of us are scientists. And there is a point where the science ends and whatever expertise we have

collectively or individually takes over. And then there is just guessing. But there is a piece in there where there is some expertise that nobody has done the trial because it is too expensive, you can't accrue, or whatever and that is why it seems we get together.

MS. WINKLER: And again, I think that is exactly what the exception allows is going into that, not as strong evidence as we would like to see but there may be a very good reason that the committee agrees to make that exception.

MEMBER KO: But that would also be true for statistics. I mean we have a statistician. We have a lot of statistical experts in the room. But it is going to be the same thing for statistics where just because of the data we have that only goes so far and there is still going to be some guessing but there is going to be some good leaps of faith that we can take statistically if we had an exception to

reliability.

MS. WINKLER: Well the evidence criteria and the testing criteria were created with expert panel task forces. And so the Evidence Task Force went there and the Testing Task Force did not.

CHAIR GUNNAR: Well, let me ask the developers. Would they like to table this and resubmit? Would they like to provide -- now that it is clear from the documents that we see here that surgical site infections was the force of their evidence and the purpose for the measurement to begin with. Should that -- would they like to review that?

MR. BRATZLER: Again, I apologize I don't have that document in front of me. I think the issue was about when surgical site infections occur, they are less likely to occur with a resistant organism. I think that may have been the way it was stated.

But again, I mean you can look at

multiple, multiple RCTs. There is just no benefit of prolonging antibiotics. It doesn't change your infection rate.

MEMBER ASHER: Maybe I am confused on that. Maybe I missed something. So, is this an issue -- as I was reading this and maybe I was just looking for the bigger picture, it seemed to me that this was really about just the idea that there is not good evidence to support continued use of antibiotics beyond this time period.

So, is this an SSI versus stewardship thing or is this a this group has not yet finalized this Level 1 recommendation? In other words, is the issue that the evidence that is being put forward here really just is not good enough to support the idea that not going past 24 hours should be a Level 1 recommendation?

CHAIR GUNNAR: I don't want to be presumptive but I will take that even further.

I think the whole way of perceiving this measurement is to actually say there is no evidence that carrying on antibiotics beyond 24 hours is beneficial, supporting the measurement, and there is actually evidence that beyond -- that prolonged antibiotics actually the longer you extend the antibiotic course needlessly without purpose actually is shown to have a rise in c. difficile and resistant organisms and et cetera.

So, I think there is evidence here that is just not in the documents or supporting the measurement the way the committee would like to have it framed.

So, I am going to stop because I am not here to argue that. That is really for the developer.

MEMBER YATES: Let's vote on that. I would argue that there may be a consensus with what you just said and we would be following 10, 11, and 12 on the guideline. And it may be

there is insufficient evidence in the document but we know where they are going.

MR. BRATZLER: And it would be very easy to, since I am an author on the HICPAC deadlines, to pull out the evidence table that has already been created for all those RCTs. It is well documented.

CHAIR GUNNAR: So, Reva, I don't think there is -- to answer your -- I don't think there is a lack of evidence. I think it has just not been provided to us. And the question is, should we just table this or ask the developers to resubmit based on this dialogue?

MEMBER HANDY: When I am looking at the document, I see a lot of evidence in here and it is primarily not about SSIs except for to say that extension doesn't prevent it. But it talks about the complications.

And so if you look in the evidence sheet 1a7.1, I mean that is where they start with a whole line of references after an

introductory paragraph.

MS. WINKLER: So if you all feel comfortable you have got the information you need to evaluate it fairly, by all means, please do. Delaying is not -- there is no advantage in delaying.

MEMBER ASHER: Okay. I think we should we vote. I mean it seems like at some people think that there is a reasonable amount of evidence to support this as is.

If the developer is okay with that.

CHAIR FLEISHER: So just to be clear, you should vote on whether you think there is sufficient evidence. If there is insufficient evidence, we can take the exception rule after that.

Okay? So, if it fails, you can go to the next question of should it be an exception.

CHAIR GUNNAR: I didn't review this. So, I appreciate Dr. Handy's sort of

just summary of -- do you think there is -- you have heard this conversation.

MEMBER HANDY: This is much more heavily referenced than the one that I am going to talk about right after this, its companion measure.

CHAIR GUNNAR: Any further dialogue? I think we are ready to vote, then.

MR. SANCHEZ: Voting will now begin for subcriterion 1a evidence.

CHAIR FLEISHER: I'm going to abstain.

CHAIR GUNNAR: Let the record show that Dr. Fleisher is going to abstain.

MR. SANCHEZ: Subcriterion 1a, evidence. One is high, two is moderate, three is low, four is insufficient evidence.

Voting timer starts now.

(Voting.)

MS. WINKLER: Folks on the phone, might want to put your cell phone on mute. We

are enjoying your typing.

MR. SANCHEZ: We have seven for high; ten for moderate; one for low; and five for insufficient evidence.

CHAIR GUNNAR: So, we can carry on. So, performance gap, Dr. Asher.

MEMBER ASHER: With respect to the opportunity for improvement, the national rate for performance for the second quarter 2013, which is the most recent data that I saw relevant to that point was at 98.1 percent with a denominator of 248,000 cases a numerator 244,000 in around 3500 hospitals.

And so, I had no particular concerns regarding disparities. And so it appears to me that it is similar to some of these other measures. This may be topped out. I mean 98.1 percent performance rate at least deserves discussion.

CHAIR GUNNAR: Any further discussion? Dr. Dutton?

MEMBER DUTTON: This is generic to all of the topped out discussions but I will bring it up now. Why should we be satisfied with 98 percent? That doesn't work for nuclear power plants or airplanes and I don't think the public would expect that for healthcare either. If we are talking about publicly reported measures, I think there is room to improve from 98 percent.

MS. WINKLER: I think the other factor you have to put in is to weigh it against burden and resource allocation and appropriate resources uses. So, the question is the benefit on the margin.

And again, it is absolutely a topic of conversation that is held on a recurring basis. But the question of added value and opportunity costs are all things that need to be factored into that topped out conversation.

MEMBER SAIGAL: And I think to be consistent, we have been pretty looking at

above 97 or 96 as topped out. And I do agree that it is probably, for most facilities, it is one patient every few months that they would be failing on. So, to that one patient, could they use those resources better to give him more impactful measures?

CHAIR GUNNAR: So, any other discussion? So, I think we are ready to vote for performance gap.

MR. SANCHEZ: Voting will now begin for criterion 1b for performance gap. One is high, two is moderate, three is low, four is insufficient.

Voting timer starts now.

(Voting.)

MR. SANCHEZ: We have zero for high; six for moderate; 17 for low; and zero for insufficient.

CHAIR GUNNAR: So the measure fails but -- so, the measure does not pass for endorsement but raises the question regarding

reserve status.

Any discussion about placing this on reserve? Well, I thought I would -- I know we are running late but the question is, is this an oddball or is this -- this is a bit of a funny one. Right? I mean I guess the question is would optics on going forward with NQF's endorsement or even as a reserve status on a measure that is, quite frankly, almost 100 percent compliant and two, isn't connected to a solid bit of SSI, which sort of goes along with all the other SCIP measures, et cetera.

I am just raising the point. So, Dr. Handy.

MEMBER HANDY: Well, I am concerned about our strategy of sticking everything into reserve status. I guess I don't understand what is the strategy of that. If by being a reserve status you have it on the shelf and you can pull it off with short notice because you start to see performance decline, that is one

thing. But if we are just sort of doing it because we don't want to kill it, I don't really understand it.

MS. WINKLER: The intent is the former but we understand the challenge committees have with the latter.

MEMBER KO: What is the disadvantage of putting it into reserve, do you think?

MEMBER HANDY: You know, I don't know that there is one per se. Just sort of you have got this portfolio of this inactive stuff that you are not really intellectually investigating or investing in. It is like having a car that you never crank in your garage. What good is it?

MEMBER KO: Well I think it is just -- to a point of maybe -- I don't know how the reserve status would work but if it could come off the shelf when -- and I suspect that the rates will go down after it is no longer, nobody

is going to watch anymore and one day somebody will measure it and it will be at 70 percent, and if it is something that is on reserve, rather than going through this whole process again, that might be advantage.

MEMBER SAIGAL: I think that if they are just topped out and they are good measures, they should all go into reserve and come back if they are needed, unless there is something really wrong with the measure.

MEMBER McCARTY: I think where our conversation is running a little short, too, is that we just don't have any data on this. So the same way we are kind of judging our metrics, we are not having good data. I mean until we know what actually happens to people's actions when you put something on reserve, that seems like the safest course of action.

And maybe in two years' from now when we can see what actually does happen, then we will be able to be more comfortable with not

putting something in reserve and go from there.

MEMBER CIMA: Just to operationalize this, I oversee 14 abstractors for this process. If we don't take it off -- if we put it in reserve, what do I go back and tell them to start doing eventually?

So, if we don't measure it -- you keep on thinking that like we are going to continue to measure it and then we will watch it. If we don't measure it, then we are not watching it.

MEMBER McCARTY: But can't you measure retrospectively? I mean you are still collect -- the data is still being held somewhere.

MEMBER CIMA: No.

MEMBER McCARTY: No?

MEMBER CIMA: No.

MEMBER McCARTY: Okay.

MEMBER CIMA: It doesn't work that way.

CHAIR FLEISHER: One important point is CMS can choose to continue it, independent of NQF endorsement. So, that is an important point.

Number two, I actually think the question, Cliff, you asked is an important one to feed up into the decision of should there be a reserve status. Because part of the question is why are you doing this. Why is this committee choosing to put these things in reserve status? Is it because they don't want to kill it? And that -- I don't know if we have time today. I am a little worried.

MEMBER CIMA: But I thought, yes, they can continue to do it but they can't tie it to any payment, unless it is NQF endorsed.

CHAIR FLEISHER: That's not true.

MS. WINKLER: I think that the information is telling potential end users, including CMS, where we think the measure is in terms of the criteria. And clearly, when they

are topped out, that is a message.

In terms of how CMS uses their measures going forward is totally up to them. However, I can tell you as we are starting to see the impact of measures that have been put on reserve status, they actually tend to -- and we haven't done the analysis. It is actually something we are going to do real soon is to see the measures we put in reserve status and what has happened with the developers. At least a few that I am aware of have been retired by CMS. So, it is not a one on one yet but there may be an association.

So, I don't think you can say just because what this vote is you should do something different if you are participating in CMS's projects.

But again, none of these things happen overnight. So the fact that you all have expressed a collective concern that this is not terribly useful going forward because

they are topped out, is a strong message to send to potential end users.

CHAIR FLEISHER: So, I actually wanted to ask one of the questions. In some of the measures you actually said they are so baked into the process, it is not going to fall off the radar screen because it is actually part of the process of care.

In others, which we have just discussed, it may. It is topped out now but it may not be baked into the processes. That may be an interesting discussion to have of what does reserve -- because if reserve means that you may want to look in the future to make sure that because it is not baked in -- I am making this up as I go, realize. But if it is not baked into the process, it is worth looking again, then maybe that is a reserve process.

I am just trying to get people to think of what is the definition of why you are keeping it in reserve. If you think it is in

the process but you just don't want to kill it.

MEMBER SIPERSTEIN: I mean obviously, the discussion is clear is that the reserved measures are perfectly valid. They are simply topped off, as opposed to saying a measure is no longer valid or useful. We are not saying that at all.

So, it is clear that we are now grading the measures, as opposed to just giving them a yes/no status. And the recommendation then is as opposed to shopping on that reserve shelf to pick your measures, you are going to go to the active ones, where you have got more bang for the buck.

And I think as an institution, as opposed to simply forgetting them, but if your outcome measures continue to be strong, then you may not need to pay attention to some of those process measures that are being reserved.

If you find slippage in your SSI rates or increase in your c. diff rates, then

that may be a call to action from your institution to go back, on an auditing purpose, and start paying attention to these processes.

MS. WINKLER: I think that reflects a great deal of the rationale for the establishment of the reserve status. But that is now three years' ongoing. And so it is time to take a good look.

So, your feedback in terms of how you are thinking about it and how you perceive the impact of your evaluation is very useful for us as we try to determine the utility of the status.

MEMBER HANDY: Well, I wanted to ask one thing and expound a little bit. Dr. Ko implied that by being in reserve we shorten the whole process if it becomes live again. I mean that is really true? You don't have to go back to the beginning and you just can pull it off the shelf and say we are going to start this up this calendar year. So, pretty huge.

And the other thing is there is a bit of a negative connotation to our retiring these things. But as what Alan said, I mean, this is a massive healthcare improvement victory that NQF ought to be blowing the horn about. And the measure I am talking about was 40 percent adherence ten years' ago and now, it is 98 percent. I mean that is huge. This is victory all over the place here we are talking about.

MEMBER SIPERSTEIN: But I agree that the term reserve could be misunderstood. And maybe if we used the term Hall of Fame, it would be -- the correct message would be out there and that these are successes.

MS. WINKLER: Trust me, Hall of Fame was one of the contenders for the naming convention.

MEMBER CIMA: But I am just going to say from an operational point of view that monitoring 400 surgeons like I do, my team, I can tell you there is a strong group of

surgeons, orthopedic surgeons, with all due respect, who if I find out and if they find out that at some point in time CMS or NQF says we are no longer going to track 24 hours, I can tell you within a matter of 24 hours, the order sets will be changed back to 24 hours, 48 hours, 72 hours. There is going to be variability all over the place.

Now, I can try as an institution we could say it is best practice but cardiac surgery, it will happen. There is going to be vanco. There is going to be this. I am just telling you the reality on the street.

MEMBER SIPERSTEIN: But that may be a misinterpretation of what the reserve status means.

MEMBER CIMA: They don't care reserve from anything. They are going to want to know can the institution go and say this is a government rule. If we cannot say that -- now, I know what you are going to say that CMS

--

CHAIR FLEISHER: What I am going to say is there is actually too many academicians in the room and quality improvement people. I do think we need to go forward but I think we can continue this over dinner, wine, and maybe a few short pithy comments for tomorrow of how we think about reserve, just going quickly around the room so that they could be recorded by NQF may be an ideal approach tomorrow morning.

CHAIR GUNNAR: So question. Should this go to reserve status? Yes or no?

Should we vote on keep going with --

(A show of hands.)

CHAIR GUNNAR: All right, there you go. I think anyone -- any negatives?

(A show of hands.)

CHAIR GUNNAR: We have one. Any abstain? Okay. Very good. So noted.

All right, next, Dr. Asher, high

priority.

MEMBER ASHER: So, I think that you could argue that there is a large number of patients impacted by the measure. The measure has significant potential implication for public health. So, high frequency medical care episode I think it is at least a moderate, if not a high score on priority.

CHAIR GUNNAR: Any other discussion? Shall we vote?

MR. SANCHEZ: Voting now will begin for subcriterion 1c, high priority. One is high, two is moderate, three is low, four is for insufficiency.

Voting timer starts now.

(Voting.)

MR. SANCHEZ: We have 14 high; six moderate; one low; zero insufficient.

MEMBER ASHER: So, this is another measure in which the reliability and validity were essentially looked at through the same

process. The numerator is the number of surgical patients who had prophylactic antibiotics were discontinued within 24 hours after anesthesia end time, 48 hours for CABG and other cardiac surgery. Denominator is all selected surgical patients with no evidence of prior infection.

Data source is administrative claims data but basically medical chart abstraction was also significant.

Denominator exclusion, nothing particularly stood out, as I reviewed this. Clinical trials infection prior to anesthesia, other surgeries, perioperative death, and no antibiotics or other procedure within three to four days of the index procedure.

So, I have no concerns with respect to specifications, definitions or coding. As I mentioned, the reliability testing, they deferred to the validity testing. The data was tested for validity at the data element level.

In summary, the sampling, the final sample included 6400 cases out of the original 1.5 million cases submitted in their data warehouse. They had determined that this sample was a fair representation of the original population. Validity tests were conducted on all 21 critical data elements. And they basically saw that for each selected data element there was reasonable -- and actually not reasonable -- very good agreement rate between the data from the hospital chart abstractor and the re-abstractor. And that information is in your folders.

And so, I would give this at least a moderate rating for reliability/validity.

CHAIR GUNNAR: Any other discussion? Hearing none -- oh, Amy. I'm sorry.

MEMBER MOYER: I guess I have seen this before and I just wanted to question it. So, on 2a2, they submitted comments received on

6/10/13, no longer necessary to report the results of reliability testing when the results of validity testing of individual data elements are reported.

And I concur that that would then mean the most they could earn on either is a moderate when they do that.

MS. WINKLER: Yes, that is correct. Essentially, the test applies to both reliability and validity. But because it is only at the data element level, moderate is your highest rating.

CO-CHAIR GUNNAR: Any other discussion? Hearing none, it is for a vote.

MR. SANCHEZ: Voting will now being for subcriterion 2a, reliability. One is for high, two is for moderate, there is for low, four is for insufficient.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have four for

high; 17 for moderate; one for low; and zero for insufficient.

MS. WINKLER: Validity, any comments?

MEMBER ASHER: So, I just stated that they wrapped it -- I would imagine it would just be the same vote. We were essentially voting on the same process.

MS. WINKLER: Okay. Is everybody okay with using the same vote for reliability and validity? So done.

Feasibility.

MEMBER ASHER: With respect to feasibility, like many of these measures, the data is generally available in the medical record. There were really no either feasibility or implementation issues identified. Some of the data, once they were defined in the EHR, I thought this deserved at least a moderate rating on feasibility.

MS. WINKLER: Further comments or

discussion? No? Ready to vote?

MR. SANCHEZ: Voting will now begin for criterion 3, feasibility. One is for high, two is for moderate, three is for low, four is for insufficient.

Voting timer starts now.

(Voting.)

MR. SANCHEZ: We have nine for high; 11 for moderate; one for low; zero for insufficient.

MEMBER McCARTY: Can I make a comment before we keep going?

CO-CHAIR GUNNAR: Yes, of course.

MEMBER McCARTY: I do share the concerns of if we are moving everything into reserve status. And I think this maybe is dinner discussion but we need to look at the QI processes and that whole cycle of what has happened with these measures that have started off with very low performance and now have achieved very high performance.

So, I am concerned that even though they were good measures, they suited their time. It perhaps might be time to go on to different measures. But if we take every measure and put that in reserve status, what are we going to end up with?

Thank you.

CO-CHAIR GUNNAR: And we go on. It should be the final -- oh, yes. Usability next.

MEMBER ASHER: So this measure is being used for public reporting, specifically in hospital inpatient quality reporting. The data is posted on Hospital Compare. It is also used in their payment programs. It is used for quality improvement with benchmarking. The rates, as we have seen, remain in hospitals across the United States.

We already talked about this issue, about it being topped out. And so, again, if I am looking at we are really -- we have to

average out 4b here. And again, I don't know how weighted that should be. But I would say at least it deserved a moderate rating based on 4a and 4c, although I see no real way that this can be significantly improved upon, particularly if we look over the last several quarters. I mean there really has just been no evidence that this thing has budged from around 98 percent.

CO-CHAIR GUNNAR: Further discussion? Time for a vote.

MR. SANCHEZ: Voting will now begin for criterion 4, usability and use. One is for high, two is for moderate, three is for low, four is for insufficient information.

Voting timer starts now.

(Voting.)

MR. SANCHEZ: We have got ten for high; eight for moderate; three for low; and zero for insufficient information.

CO-CHAIR GUNNAR: So, voting for

this measure to put it in reserve status. And you can decide what that is at dinner.

MR. SANCHEZ: Voting will now begin for potential for reserve status. One is for yes, two is for no.

Voting timer starts now.

(Voting.)

CO-CHAIR FLEISHER: Okay.

CO-CHAIR GUNNAR: Hang on. Let's just finish this voting and we will be done in 44 seconds or less.

MR. SANCHEZ: We waiting on one more. There it is.

MEMBER SAIGAL: We have one more to go and we are done? No.

MR. SANCHEZ: We have eighteen yes; three no.

CO-CHAIR GUNNAR: Very good. So, the vote is to reserve.

CO-CHAIR FLEISHER: So, I want to thank our colleagues from SDS. They will defer

to tomorrow, since they have to be here anyway. So, our final one -- well, let's try to quickly do the companion AMA PCPI discontinuation.

Dale?

MR. BRATZLER: So, essentially the same performance metric here again. It is the physician order to stop the antimicrobial which is within the control of the physician. And that data is different. It is the administrative data. But otherwise, the concepts of the measure are the same.

CO-CHAIR FLEISHER: Okay, who is discussing it? Okay, John.

MEMBER HANDY: It is a maintenance submission for a process measure. And so the evidence is primarily the same guidelines that was referenced a couple of times ago. The guideline, when you go to it is protean and it has to do with agent selection and timing. But this particular one is on duration. And so the paragraph on duration, which is one of the

shorter parts of it, is referenced quite well and has three clinical practice guidelines, four consensus statements, three retrospective studies, two surveys, five clinical trials, et cetera. And that is really the evidence that they have for it, noting that the evidence is really kind in the negative. In other words, the prolongation of antibiotics doesn't do anything. It is not clear how short the antibiotics can be.

And one of the things that these guys talk about in contrast to the foregoing discussion, notwithstanding, is that there is really, they note that there is really no recent work on this particular subject, which I thought was interesting.

So, it is a process measure and it does have a guideline that is heavily referenced with systematic reviews.

CO-CHAIR GUNNAR: So, any new comments compared to the previous measure?

Hearing none, do you want to vote on evidence?

MR. SANCHEZ: Voting will now begin for subcriterion 1a, evidence. One is high, two is moderate, three is low, four is insufficient evidence.

The voting timer starts now.

(Voting.)

MEMBER KO: I have to recuse myself for this one.

MR. SANCHEZ: So we have six for high; 15 for moderate; zero for low; and zero for insufficient evidence.

CO-CHAIR GUNNAR: Next.

MEMBER HANDY: So the next thing is the performance gap. And this has some of the same issues that the prior PCPI had, is that there are two data sets that are submitted, one from 2008, where there was a performance of 44 percent with regard to the successful cessation of antibiotics within 24 hours. And then 2010

data, which is the PQRS data, which has 29 percent of the people that are available reporting. And so that gets down into a -- I remain a little bit confused as to whether the -- this is a smaller number. Where the original one we were talking about was eight percent, this is five percent of the people ultimately are the ones reporting. So, I am not sure if this where the performance gap comes from.

But anyway, in 2008, there was a 44 percent adherence to cessation of antibiotics within 24 hours and in 2010, the most recent data that we have submitted, it is 98.2 percent. So, there has been a complete obliteration of the performance gap. The performance gap now is tiny.

CO-CHAIR GUNNAR: Comments? Shall we vote?

MR. SANCHEZ: Voting will now begin for subcriterion 1b, performance gap. One is

high, two is moderate, three is low, four is insufficient.

Voting timer starts now.

(Voting.)

MR. SANCHEZ: We have zero for high, four for moderate, 18 for low, zero for insufficient.

CO-CHAIR FLEISHER: So, the next question. Reserve status. All who wish to put this into reserve status?

(A show of hands.)

CO-CHAIR FLEISHER: Okay. So, if we can quickly go through the other issues, really if there is anything new.

MEMBER HANDY: Well, there is one thing new. And that is when you look at the reliability in the numerator and denominator. And Dr. Saigal changed my thinking on this a little bit. This is really measuring whether or not you had an order, not whether or not you discontinued the antibiotics. I guess I had

never framed in what can the physician be responsible for, other than writing of the order. So, I viewed that as a criticism before but that is really what is being measured here is whether there is an order for discontinuance, not the actual discontinuance of the antibiotics.

CO-CHAIR FLEISHER: So you want to get to the votes? Does anyone feel they need to comment on this or are we prepared to vote? Any comments? Does anyone want to comment on high priority? Okay, let's vote.

MR. SANCHEZ: Voting will now begin for subcriterion 1c, high priority. One is for high, two is for moderate, three is for low, and four is for insufficient evidence.

The timer starts now.

(Voting.)

MR. SANCHEZ: We have 11 for high; six moderate; four low; zero insufficient.

CO-CHAIR FLEISHER: Okay, we heard

the comments on reliability. Any comments? Vote.

MR. SANCHEZ: Voting will now begin for subcriterion 2a, reliability. One is for high, two is for moderate, three is for low, and four is for insufficient.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have six for high; 14 for moderate; one for low; zero for insufficient.

CO-CHAIR FLEISHER: Next, validity. Any comments?

MEMBER HANDY: So validity testing was done the same way as the prior measure. It was basically an expert consensus panel, which 80 percent of the 21 people polled said they thought it was valid. So, it is moderate, at best.

MR. SANCHEZ: Voting will now begin for subcriterion 2b, validity. One is for high, two is for

moderate, three is for low, four is for insufficient.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have zero for high; 16 for moderate; five for low; zero for insufficient.

CO-CHAIR FLEISHER: Okay.

MEMBER HANDY: The next thing is feasibility. These are from administrative claims and it has been done for a long time so it is very feasible. It is a data element versus a measure score.

CO-CHAIR FLEISHER: Comments?

MR. SANCHEZ: Voting will now begin for subcriterion 3, feasibility. One is for high, two is for moderate, three is for low, and four is for insufficient.

The voting timer starts now.

(Voting.)

MR. SANCHEZ: We have ten for high;

ten for moderate; one for low; zero for insufficient.

CO-CHAIR FLEISHER: Usability.

MEMBER HANDY: The usability has been awesome. It has been so useful that it made itself obsolete.

(Laughter.)

CO-CHAIR FLEISHER: Comments? Vote.

MR. SANCHEZ: Voting will now begin for subcriterion 4, usability and use. One is for high, two is for moderate, three is for low, four is for insufficient information.

Voting starts now.

(Voting.)

MR. SANCHEZ: We have eight for high; 11 for moderate; three for low; and zero for insufficient information.

CO-CHAIR FLEISHER: Okay, vote on reserve status.

MR. SANCHEZ: Voting will now begin

for a potential for reserve status. One is for yes, two is for no.

Voting timer starts now.

(Voting.)

MR. SANCHEZ: We have 17 for yes; four for no.

CO-CHAIR FLEISHER: That being said, we have actually gotten permission from the two developers to defer to tomorrow, which means we are done for going over the measures today. We are all in the surgical arena. So, we will be starting on time to get this done by 3:30. In fact, does anyone have to leave before 3:30 tomorrow?

Okay, so now we are open for public comment from the room.

MS. WINKLER: And also on the phone. Operator, could you see if there is anyone who has any public comment?

OPERATOR: Yes, ma'am. At this time, if you would like to make a public

comments, please press *, then the number 1.

At this time, there are no public comments.

MS. WINKLER: Anybody in the room? Okay.

Just some final thoughts before you leave. We pushed three measures until tomorrow -- actually four. So tomorrow's schedule is going to be really tight. So, for those of you who are presenting and for everybody in terms of discussion, we don't want to limit discussion but by the same token, we really have to stay focused and aware of time. So, that is on all of our responsibility to get through this and get all of our work done tomorrow. Long days. Thank you very much.

Just I have been emailing back and forth with one of my colleagues in terms of this whole business about reserve status and what is going on and what does it mean. And she just sent me a quick brief outline of CMS's proposed

rule for the IQR, the Hospital Inpatient Quality Reporting Program that is out. It was issued at the end of April. Now remember, this is the proposed rule. But they are proposing to remove 15 chart abstracted measures that include the SCIP 1, SCIP 2, and SCIP 3, which are the three CMS SCIP measures you guys just voted on reserve status.

Okay, so these things kind of do track together. Now, that isn't the final rule but that is the proposed rule. So, CMS is thinking somewhat similarly in terms of where they may be going with these measures also.

CO-CHAIR FLEISHER: Great. So now, housekeeping. Where are we going?

MR. SANCHEZ: You should have just received an email from me with the address, as well as with a Google Map link to where. It is literally two blocks from here, a place called Mio. It is right on Vermont Avenue. The reservation is for 7:00 p.m. It is both under

my name Amaru Sanchez, as well as NQF. So, whichever one you feel.

And again, the allotment for it is $36.

CO-CHAIR FLEISHER: Thank you for an amazing and thoughtful day.

(Whereupon, at 5:50 p.m., the foregoing meeting was adjourned to reconvene at 8:00 a.m. on Thursday, May 29, 2014.)