NATIONAL QUALITY FORUM
+ + + + +
SURGERY ENDORSEMENT MAINTENANCE
2010 STEERING COMMITTEE
+ + + + +
MONDAY
FEBRUARY 28, 2011

+ + + + +

The Steering Committee met at the Washington Hilton, Lincoln West room, 1919 Connecticut Avenue, N.W., Washington, D.C., at 9:00 a.m., Arden Morris, Chair, presiding.

PRESENT:

ARDEN MORRIS, Chair, University of Michigan Health System
NASIM AFSAR-MANESH, UCLA Medical Center
JAMES CARPENTER, University of Michigan
ROBERT CIMA, Mayo Clinic

CURTIS COLLINS, University of Michigan Health System
PETER DILLON, Penn State Hershey Medical Center
RICHARD DUTTON, Anesthesia Quality Institute
STEVEN FINDLAY, Consumers Union
PAULA GRALING, Inova Fairfax Hospital

VIVIENNE HALPERN, Carl T. Hayden VA Medical Center
EILEEN KENNEDY, Pepco Holdings
RUTH KLEINPELL, Rush University Medical Center
JOHN MORTON, Stanford University
DENNIS RIVENBURGH, St. Anthonyþs
TERRY ROGERS, The Foundation for Health Care Quality
CHRISTOPHER SAIGAL, UCLA Medical Center
NICHOLAS SEARS, MedAssets

ALLAN SIPERSTEIN, Cleveland Clinic
RENAE STAFFORD, University of North Carolina
CONNIE STEED, Greenville Hospital System
CAROL WILHOIT, Blue Cross Blue Shield of Illinois
CHRISTINE ZAMBRICKI, American Association of Nurse Anesthetists

NQF STAFF:

HELEN BURSTIN
KRISTIN CHANDLER
ALEXIS FORMAN
ANN HAMMERSMITH
MELINDA MURPHY
JESSICA WEBER

ALSO PRESENT:

RICHARD PRAGER, The Society of Thoracic Surgeons
HARRIET GAMMON, The Joint Commission
SHARON SPRENGER, The Joint Commission
DAVID SHAHIAN, The Society of Thoracic Surgeons (via telephone)
JANE HAN, The Society of Thoracic Surgeons (via telephone)

JOHN BOTT, Agency for Healthcare Research and Quality (via telephone)
JEFFREY GEPPERT, Battelle Memorial Institute (via telephone)
PATRICK ROMANO, UC-Davis (via telephone)
ANNE SNOWDEN, MPH, CPHQ, Minnesota Community Measurement*

JOHN A. SPERTUS, MD, MPH, University of Washington School of Public Health*
SAMANTHA TIERNEY, MPH, American Medical Association
MANASI TIRODKAR, PhD, MS, National Committee for Quality Assurance

*Present via telephone

C-O-N-T-E-N-T-S

Page

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

9:02 a.m.

CHAIR MORRIS: Good morning. We're going to go ahead and get started. I'd like to thank everybody for coming to the meeting today on surgical endorsement and maintenance measures with the National Quality Forum.

We're going to start by going around the table and introducing ourselves, and then just mentioning whether we have any disclosures and what they are.

And I'll start. I'm Arden Morris. I'm an associate professor of Surgery and the University of Michigan.

MS. MURPHY: And let me just remind everyone that everything from this point forward will be recorded. So be sure that you're using the microphone, that you press the button and see the red light when you're speaking. When you are no longer speaking, be certain that you turn it off,

because we'll get a lot of noise if we have multiple speakers on. But we are recording from this point forward.

CHAIR MORRIS: And in addition to that, the transcripts will be posted online, so you'll be able to go back and review them if you desire. So I'm Arden Morris. I'm an associate professor of Surgery at the University of Michigan.

Today, I'm the Chief of General Surgery at the Ann Arbor VA and tomorrow I'll be the Chief of Colorectal Surgery at the University of Michigan. I have no disclosures besides that. Beg pardon? Day after? We'll see.

(Laughter.)

CHAIR MORRIS: Stay tuned. So let's go ahead and go around.

MS. HAMMERSMITH: Hi. I'm Ann Hammersmith and NQF's general counsel. Before you finish up with disclosures, I just want to remind you of a few things and make a few

statements. You all received the conflict of interest form from NQF, which you all filled out. We went through those. We eliminate people who we believe had conflicts or had apparent conflicts of interest.

What we ask you to do today in the spirit of openness and transparency, which NQF is known for, I'd just ask you to go around the table, introduce yourselves, as your chair did, and disclose anything that you think your fellow committee members should know.

One thing I want to remind you of is that you sit on the committee as an individual. We often have people say I'm representing the interests of or the views of fill in the blank organization. That's actually not the case.

Even if that organization nominated you, you sit as an individual, as an expert. So I'll ask you to go around the table.

DR. GRALING: Good morning. I'm

Paula Graling. I'm the clinical nurse specialist at Perioperative Services at Inova Fairfax Hospital here in D.C., and I have no known conflict of interest.

DR. STAFFORD: Good morning. I'm Renae Stafford. I'm an assistant professor of Surgery at the University of North Carolina in Chapel Hill.

I have no known conflicts; however, I am a member of a number of different surgical and trauma organizations that clearly would be benefit from this, and I also have an indirect conflict in that I have a family member who works for a biotech firm.

DR. CARPENTER: Good morning. I'm Jim Carpenter. I'm an orthopedic surgeon. I'm the chair of Orthopedic Surgery at the University of Michigan, and I have no conflicts regarding these topics.

DR. COLLINS: Hi, good morning. My name is Curtis Collins. I'm a clinical

pharmacist also at the University of Michigan, and no conflicts.

MR. RIVENBURGH: Good morning. My name is Dennis Rivenburgh. I'm a physician assistant practicing Orthopedics and Sports Medicine in St. Petersburg, Florida, and I have no conflicts.

DR. MORTON: I'm John Morton. I'm chief of Minimally Invasive and Bariatric Surgery at Stanford. I'm a director for Surgical Quality at Stanford. My one disclosure is I have an educational grant from Ethicon Endo-Surgery.

DR. KLEINPELL: Good morning. I'm Ruth Kleinpell from Chicago, Illinois, Rush University Medical Center. I serve as a director for Clinical Research there and I'm also a professor of Nursing and a nurse practitioner.

DR. CIMA: Good morning. My name is Robert Cima. I'm a colorectal surgeon and vice chair of the Department of Surgery for

Quality and Safety at Mayo Clinic in Rochester, and I have no disclosures.

DR. SIPERSTEIN: Hi. Allan Siperstein, a professor of Surgery at the Cleveland Clinic, chair of Endocrine Surgery there. I have no conflicts.

DR. HALPERN: Vivianne Halpern. I'm the chief of Vascular Surgery at the Carl T. Hayden Phoenix VA Medical Center, and associate professor of Surgery at the University of Arizona. I have no conflicts.

DR. DILLON: Good morning. I'm Peter Dillon. I'm chair of Surgery at Penn State-Hershey, and I have no conflicts other than contracts with Synthese.

MS. STEED: Hello, I'm Connie Steed with the Greenville Hospital System University Medical Center, and I have a research grant with Deb Rovai (ph), which is doing research on hand hygiene and surgical sepsis.

DR. SAIGAL: I'm Chris Saigal.

I'm an associate professor of Urology at UCLA. I did some consulting for American Medical Systems last year.

DR. ROGERS: Hi. I'm Terry Rogers. I'm a recovering pulmonologist who currently is a CEO at the Foundation for Health Care Quality in Seattle. We're a state-wide organization that looks at various surgical and medical procedures. I have no conflicts.

DR. DUTTON: I'm Rick Dutton. I'm a trauma anesthesiologist from Baltimore, and currently the executive director of the Anesthesia Quality Institute.

MS. ZAMBRICKI: Hello. I'm Christine Zambricki. I am as of two weeks ago the deputy executive director for the American Association of Nurse Anesthestists. Prior to that, I was chief operating officer and chief nursing officer for a hospital, and the conflict that I previously reported is that I sit on the Executive Advisory Board of

Surgical Information Systems, SIS, which is an information system technology company for the perioperative interval care.

DR. SEARS: I'm Nick Sears. I serve as the chief medical officer for MedAssets, Incorporated, and I have no conflicts.

DR. WILHOIT: I'm Carol Wilhoit. I'm Quality Improvement medical director for Blue Cross/Blue Shield of Illinois, and I have no conflicts to report.

MS. KENNEDY: Good morning. I'm Eileen Kennedy. I'm the manager of Benefits, Reporting and Compliance for PEPCO Holdings, and I have no known conflicts.

DR. AFSAR-MANESH: Hi. I'm Nasim Afsar. I'm an associate professor in Internal Medicine and Neurosurgery, and I'm the associate director of Quality at Ronald Reagan UCLA Medical Center, and I have no conflicts.

MS. HAMMERSMITH: Okay, thank you everyone. Is anyone on the phone? Anyone

participating on the phone? No. Okay. Do any of you have anything you want to discuss about what was disclosed, any questions for each other or for me?

(No response.)

MS. HAMMERSMITH: Okay, great. Have a good meeting. Thank you.

CHAIR MORRIS: All right, thank you. We're going to briefly run through some of the expectations in the process for the meeting next. I'm sorry. But first, Helen Burstin is going to say a few words, and have we been successful at all at reaching David Torchiana on the telephone?

MS. FORMAN: I've sent him the dial-in information, so he can call Donald to let us know.

CHAIR MORRIS: Okay. For those of you who came late, he's ill with the flu and so has been unable to make it today.

DR. BURSTIN: At least we're healthy and we're here, so we'll take it as a

benefit. Good morning, everybody. I'm Helen Burstin. I'm the senior vice president for Performance Measures at NQF. Thank you all for coming together.

I just wanted to add my welcome and also just to let you know a little bit about this process. Some of you who may have served before will recognize it's a bit different.

This past spring, the NQF Board approved a change, where we moved from doing endorsement of new measures separate from maintenance and instead brought them together in this process we called endorsement maintenance.

The idea is is that we are actually going to be looking with equal footing at measures that are newly-submitted, and measures that are endorsed and up for maintenance. The measures that are up for maintenance will actually now be subject to the full review of all the criteria of full

submission that you've already seen.

The idea here is it really allows us to achieve two important things. The first is we really want to allow the measure that we think is really best in class to move forward. It's very confusing, we feel, to have multiple competing measures on the same topic. The only way to do that is to allow you to see the two measures head to head.

The way we'll accomplish that is if there are measures that are related or competing, and we'll go over that with you as we get deeper into it, we will ask you to review each measure on its merit individually.

Then we'll have a process, usually at this meeting or to follow, we'll see how your timing works, to actually put those ratings head to head and actually assess the best in class and try to help make that determination.

The second thing it allows us to do is actually harmonize measures. There's a

lot of differences in measures that are used at the ambulatory level, the hospital level, the post-op, just the cacophony we've all seen of measures that change, depending on setting of care.

The only way for us to at least harmonize it on target, target condition, target surgery, patient population, whatever the case may be, is by bringing them head to head. So that's why this change in process.

So this is somewhat new for our developers as well. You're only the second committee that's done this. Cardiovascular met last week and my understanding is you don't have nearly as many competing measures as they did. It was kind of one big competing measures, aspirin beta blockers, ACE/ARBs. I mean it was just extraordinary.

So I think you're in a little bit better shape here, and again, I'm here to help if there's any questions about process or where we're going or just general questions

about direction. But we now have over 670 endorsed measures across multiple sites, settings, types of providers, specialties, etcetera.

Some of that growth is great, because it allows us to say yes, we have measures for a particular area where they're needed. Some of that growth is duplication, which we don't want.

So this is really our attempt to hone in on what's important, what's most useful at the end of the day for public reporting and accountability, and the things that are really not being used by anybody or not meeting the rigorous tests of reliability and validity could probably fall to the wayside.

So that's kind of our thinking, and again, I'll be here with you if you have any questions as we move forward. Arden's a veteran. I think this is your third steering committee.

CHAIR MORRIS: Yes, it is.

DR. BURSTIN: I'm sorry, we don't have David with us, but hopefully he'll get to call in. So thanks.

CHAIR MORRIS: And we can just go through a couple of things, just to give you an overview of the agenda. So as you heard, our goals here are to review the maintenance of measure issues and then also some new measures, and they all are going to be evaluated by the same criteria.

As I spoke before in our earlier executive meeting, one of my roles is to make sure that the discussion moves forward, but another role is to make sure that everybody really has an opportunity to talk about it if they have any issues or questions, particularly questions for the developers, who will be on the line or are present in person.

We'll start with a brief introduction of the measures by the developers, and then as we discussed

previously, we will go through the measures one by one and we'll evaluate them by each criteria.

I think that -- so one thing that I really want to underscore is that if you have any concerns about any of the measures, please do bring it up. You're here for that reason. Okay.

MS. MURPHY: So I'm going to give you a bit of introduction to the project, to the way in which the criteria will be approached before we get started, and we're actually running nicely ahead of schedule at this moment.

What you already know is much of what you're going to hear, so hopefully this will reinforce some of the things you've already looked at, as you did your preliminary evaluation of the measure. So we'll start with just the purpose of the project.

As Helen had said, we're looking at endorsing measures that address the care of

the surgical patient and surgical procedures, and at the same time you're considering new measures, you will also be conducting a review, a maintenance review of surgical measures that have been endorsed by NQF, and specifically to look at those that were endorsed prior to June of 2008. Measures from that, endorsed from that point forward, will be considered in a later project.

NQF endorses measures for public reporting and quality improvement, and so it's not "or." It's public reporting "and" quality improvement. So as you consider these, think about both those.

As you know, by this point, even if this is your first activity with NQF, there is a standardized consensus development process that is used, and it is that process that's both set out in terms of what the components are in law, and it's also one that's been used overtime with NQF to develop consensus through multiple iterations of

consideration of measures or other potential standards.

When NQF measures are endorsed, they are known as voluntary consensus standards, and as you should know by this point, they are widely used across government sector, the states, health plans and insurers and accrediting organizations, which makes it very important that any measures that you are recommending and considering be carefully considered in terms of the criteria.

This project, as I've already mentioned, will be looking at newly-submitted measures, and those measures that are being considered for maintenance. No measure gets a bye. Every measure that you're going to consider today you will consider based on each of the four major criteria.

So the measures that are here for maintenance will also be evaluated against each of those criteria additionally, and we'll talk about that in a minute. Each measure

that's being considered for maintenance must also be looked at in terms of information gained over the period of time that the measure's been in use.

We're doing this in two phases. So you will be involved in both those phases, and this first phase, looking at measures that are cardiac surgery-related.

We've got also esophageal resection, VTE prophylaxis and a set of newly-submitted blood transfusion measures. Phase II will pick up general surgery and a number of the other surgical specialties.

As Helen mentioned, we're also looking at what is the potential for harmonization of measures, whenever there are similar or related measures, in order for them to be harmonized in terms of the specifications in terms of the populations and any, again as Helen mentioned.

If a measure has become no longer relevant for whatever reason, including no

longer having a performance gap that is significant, and one other thing I would say about the gap is the fact that a measure is performing at a very high rate doesn't in and of itself suggest it should be retired.

There may be good and valid reasons why it should be continued. But there is the opportunity to look at measures. In fact, you should be looking at measures in terms of performance gaps, and then the opportunity to expand any related measures.

The orientation in looking at the measures is very much moved to a patient focus. So we're looking at the care coordination across settings; care transitions; hand-offs; shared accountability across individuals, teams, systems, organizations; shared decision-making with the patient sharing in the decision; and also looking at value.

So if you're looking at a measure that is an excellent measure in terms of what

it targets, what it gives you information about, but it is exorbitant in terms of what it costs to collect the information, to assess it, then you want to look at whether the value gained is significant enough to support the cost. This also feeds to supporting the payment reform approach.

So as measures have evolved, it is no longer looking at let's get some measures to look at various aspects of care; it's let's be certain that as those measures are applied and we're able to look back at them in terms of the performance, that they are in fact driving toward higher performance; that we are having the essential measures, that we're not having so many measures that it's impossible to deal with everything that's out there.

So looking at shifting towards composite measures, looking at harmonizing measures across sites and across providers, and measuring the largest possible group that's supported by the evidence. So if

you've got two very similar measures or two related measures, and they look at two populations, and you can harmonize those into a measure that looks at both populations with a single measure, maybe stratifying results, to look at what are the opportunities to do those things, and to promote shared accountability and measurement across the patient-focused episode of care.

Not focused on individual providers, not focused on hospitals or professionals, but focused on the continuum of care within which the patient receives care. So outcome measures, appropriateness measures and looking at the resource utilization balanced against the quality information that's gained.

Your role as a steering committee is to act as a proxy for the NQF's multi-stakeholder membership. You are the individuals with the expertise in the subject matter area. You can be expected to opine

upon the measures with full knowledge of what it takes to establish the evidence, to evaluate the evidence and to evaluate the value of the measure as a proxy for the NQF membership.

You'll work with the NQF staff and primarily those of us who are here in the room, to achieve the goals of the project. Our job is to try to facilitate and make your job as straightforward and easy and put things for you, to have you not have to do all of the searching and work.

You will make recommendations to the NQF membership for endorsement through the process you're engaged in today, which will go into a report that will go to the NQF membership and the public for review and comment.

You'll be able to look at the result of that review period and provide information about whatever improvements need to be made, and either going back and have

conversations with the measure developers, or in the report itself.

Then post-vote, you will have another opportunity to look at the report. So throughout the process, you are acting on behalf of the over 400 members of NQF.

The co-chairs of the Committee will represent you whenever the Consensus Standards Approval Committee meets to consider your recommendations. This is just a schematic of what I've just mentioned, in terms of the process.

So at this meeting, the objectives are that you will evaluate the measures that you have before you according to the NQF criteria, to determine if those measures are suitable for endorsement initially or continued endorsement if they're maintenance measures, as voluntary consensus standards.

Then once you have done that, to the extent that there are related measures or competing measures, you will look at those in

terms of are there opportunities to harmonize the measures. If there are measures that are clearly competing in terms of having the same, essentially the same numerator, the same denominator, the same specifications, the same population, is there one that is best of the two?

And then to identify gaps in the performance measures that are available for the care of the surgical patient and surgical procedures. That we really expect we will have an introduction for you at this meeting, and will consider more fully at during Phase II.

Okay. So basic consideration for any measure to be brought forward to you. If it is a non-government organization, there must be a measure steward agreement that, among other things, provides the information that everything that is available to utilize the measure in terms of all of the specifications and access to any tools that

are needed to apply the measure will be made fully available for any measure that's endorsed, and that includes any measure that has proprietary components.

There also on the part of the measure steward must be a commitment that they have in place the tools and a process to maintain and update the measure as needed, and at least every three years to provide the information that it is up to date, or to provide updates.

They must commit that the measure is available and is expected to be used for both quality improvement and public reporting, and the measure submission information must be complete. The measures in general must be fully developed and have been tested, so that all of the evaluation criteria that you're going to use have been addressed, and you can assess that.

The endorsement criteria that you're going to be looking at, and you've

looked at this already in your preliminary review, are the four that you see on the screen, and they are in an order for a reason.

First, Importance. Importance to measure the topic area and report the information. This would be measures that have the greatest opportunity to really drive improvement.

If the measure is not important, based on evidence; you're looking for evidence of the importance. If it's not important, nothing much else matters. Once it passes the threshold of Importance, and this is a yes/no question, yes it is or no it isn't, if the group determines that it is important to measure and report based on the evidence, the next consideration is Scientific Acceptability of the measure properties.

So you're looking for validity; you're looking for reliability. If the measure is scientifically acceptable, you can go on to consider Usability, and that is can

it be used to come to conclusions and make decisions.

If again it doesn't pass that threshold, then it probably doesn't matter if it's easy to reasonable to collect. But if it is usable as defined, then the next question is -- the next criterion is Feasibility.

The objective with Feasibility is the ability to collect it with as little burden as possible, and there will be people who say yes, right. But at this point, we really have had a lot of experience, and there are many efforts underway to improve the ability to collect information electronically.

So you're looking for as little burden as possible, and then if they're competing measures, you're looking for best in class.

For each of the criteria, there are subcriteria. For each of the criteria and subcriteria, there's rarely a time whenever it's all or nothing, apart from importance,

yes or no. But for the others, it's generally you've got to weigh a number of factors.

The rating scale that you will use and you're going to use the electronic voting mechanism, is what you see on the screen right now.

You will be looking at each and every one of the criteria and evaluating, after Importance, which is yes or no, you'll be evaluating them in terms of whether they completely meet the criteria, partially, minimally, not at all, or it doesn't apply, and there are very few that don't apply.

The steering committee has already had the opportunity to use the voting hand-held device. You'll see the information from the voting on the screen by title and number, okay, and if it is necessary to re-vote for whatever reason, that can be done.

So what you're going to be voting on, Importance. The extent to which there is the evidence that demonstrates importance, the

information that the use of the measure would have a significant impact, that there is a gap in performance to be addressed by the measure, and the evidence supports the focus of the measure.

In terms of Importance, there are a few things in the subcriteria that are for maintenance, and that's what you see with the second smaller font size information on each of these subcriteria.

So for the summary of data demonstrating the performance gap, you're going to be interested in knowing what has occurred over the time that this measure has been in use, in terms of the results of its application.

You're going to be looking for what is that performance gap. You're going to be looking for whether or not they've identified and what they have identified in terms of disparities by population group, and when looking at the information about

disparities, you're going to want to see specific information about the disparities.

Any of those things that you do not see, you're perfectly welcome, in fact, encouraged to ask questions of the developers, to collect that information, to provide you that information. Sorry. You will vote on importance. That will be one of those times you'll have your keypad and you'll say yes, it's important or no, it's not.

Scientific Acceptability, you're looking for validity and reliability. You're looking for is it precisely specified? Has reliability and validity been tested? Are the exclusions that have been identified justified?

In the risk adjustment method, if there is risk adjustment used, is it evidence-based, and are any factors that are risk-adjusted out certain that they are factors that were present at the start of care.

You don't want to see factors that

could be changed as a result of care to be risk-adjusted out. Statistically significant differences in performance.

If there are multiple data sources, do they provide comparable results? If there are disparities, are they stratified? Can you see what the performance is with the measure across different groups for which there are disparities.

Then at the end of the consideration of Scientific Acceptability, you will vote on each of those areas, about whether it completely meets, etcetera. Then you'll move on to Usability, looking for the extent to which audiences can understand the result of the measure, and can find them useful in decision-making, including consumers, including patients, and are they harmonized, and you'll have an opportunity to look at any that should be considered for harmonization, and do they add value to the current set of performance measures that are

available to look at that topic area.

Again, for maintenance measures, any of the measures that are not being publicly reported, the measure steward should be able to tell you what is their plan for public reporting.

If the measures are not being used for quality improvement, again, you should be able to hear from the developers about what is their plan for using them for quality improvement.

Because again, these are measures that are -- the maintenance measures have been in use. So there will be information related to the extent to which they've been used for quality improvement in public reporting. And at the end of the Usability discussion, you will vote on that criterion.

Feasibility, the extent to which the data that you need for the measure is readily available, without undue burden, and what you see A through E are potential sources

of information.

If there are exclusions that are there, you'll be looking for whether or not they have to go to some other data source to find the information for the exclusion, and whether or not that activity adds significant additional burden. So you're looking for can the data collection strategy be implemented, and you will vote.

So Step 1 that you see there is what gets us through each of the measures that you have to consider today. A full evaluation, based on each of the criteria.

Step 2 will be for any measures that are similar related, looking at those in terms of the potential for harmonization. As Helen said, in this particular project, particularly in Phase I, there are very few, but there are a few.

At the end of Day 2, there will be some discussion about those. The question right now that will resolve it before we get

there is whether or not we can take care of being clear that we know what those are, in order to ask developers to look at them for harmonization, and then if there are competing measures, and there may be a couple, the opportunity responsibility to take a look at those in terms of is there a best in class, and then a final recommendation for endorsement.

So each of you were assigned a measure or more than one measure to do a preliminary evaluation. At this meeting, everybody should participate in the discussion of each measure and vote on each measure.

Those of you who have the responsibility or had the responsibility for doing an indepth review of individual measures will kick off the discussion. So you can provide some summary information, but the whole group should then engage in the discussion, and the entire committee will discuss each of the four criteria and will

vote.

So format always. NQF member and public are invited to the meetings. Their input, their comment, their insights are always useful and are always provided for within the agenda. The measure developer, and today we will have measure developers in the room. We have some here now, and we're very happy to see them.

Each of the measure developers will have an opportunity to introduce the measures that they're bringing forward at the beginning of each day. After they have introduced the measure, then they will remain in the meeting, so that if you have questions or if there's additional information that you would like from them during the course of the day, they can provide that for you.

They will also be able to offer additional comments during the public and member comment period. We've already talked about your voting. Any questions about any of

what I mentioned?

DR. ROGERS: Linda, just a quick question about the life cycle of an approved measure. So that if approval is granted for a certain measure, it's likely to be in force, if you will, or present in that form for two years, three years?

MS. MURPHY: What we are asking is that there be an opportunity to take a look at them every three years. The expectation is that if the evidence changes, that the measure will be updated.

If the developer finds that, for example, there is a really high level of performance, they might come back and say there's something that we want to do about this because there's such a high level of performance, which could mean that they might want to make some adjustment in the measure, in order to make it more sensitive.

They might want to repilot. But the expectation is when there is new

information, new data, new evidence, that the developer would update the measure and provide the information, but that at least every three years we would have the information that it's either been updated or that there is no data to support a change.

DR. ROGERS: So it's then incumbent upon the developer, if you will, to keep an eye on this longitudinally, to guarantee that there's appropriateness as time goes on?

MS. MURPHY: That's correct.

DR. ROGERS: Good, thank you.

MS. MURPHY: Right, and my experience has been that they do, that they're very good about that, and that organizations, entities or stakeholders who have a particular interest in particular sets of measures around which they have knowledge and understanding about the evidence will take that forward to the developers, whom I have seen over time to be very sensitive and open to that.

If not anything else, than we could hear from developers.

CHAIR MORRIS: Okay. We're a little bit ahead of schedule here. Any other questions or any issues that anybody has?

(No response.)

CHAIR MORRIS: Do we have measure developers on the telephone as well as here in person? No, nobody yet? All right. Shall we just -- sorry?

DR. PRAGER: Hi. I'm Richard Prager, another person actually from Ann Arbor. I'm a cardiac surgeon. I am here as a chair of two task forces for the Society of Thoracic Surgeons, which is one of the measure maintenance groups that you will see today.

I believe Jane Han from the STS in Chicago is to be on the phone actually to present the history of these measures. Jane, are you on?

DR. HAN: I am here, Dr. Prager.

DR. PRAGER: Arden, would you like

that to start this?

CHAIR MORRIS: That sounds good.

DR. PRAGER: Okay. Jane, it's all yours.

DR. HAN: Sure. I just have a couple of brief sentences for you regarding the history of the measures. Currently, in Phase I of the Surgery Endorsement Maintenance project, we have 17 of 22 STS-built cardiac surgery measures in front of you today.

These measures were all endorsed in 2004 and received reendorsement in 2007, and as scheduled, in 2010 they were up for maintenance, and that's why they're being reviewed by you.

They have been in use by the STS database for years, and measures data are reported to STS Adult Cardiac Surgery database participants on a semi-annual basis. We'll be on the phone all day today and tomorrow, so if you have any questions, Dr. Prager will be there and I will be here as well.

CHAIR MORRIS: Thank you. Do we have anybody that wants to go next?

DR. GAMMON: Hello. I'm Harriet Gammon from the Joint Commission, and we're bringing forward the Patient Blood Management measures. This is really a new area for measurement. I think as a nation we've worked a lot on blood safety, but this is the first time we've really looked at patient transfusion safety.

We have five measures that are directly related to transfusion, and we have two measures that are related to surgical patients, the first one being pre-anemia screening, in that we wanted to optimize our patients before they go to surgery, because there is an association. If they're optimized before they go to surgery, they may not need as much blood during surgery or after surgery.

The other measure looks at type and cross and type and screen prior to the

procedure, and we want to make sure that it's done before the anesthesia start time. A lot of times with the patients coming in on the same day for surgery, this isn't always done in advance.

You know, we've heard of some difficult cases that have had some issues with this, and so we would like to bring this forward as a measure for patient safety.

CHAIR MORRIS: Thank you very much. I have a question for folks on the phone. We're looking specifically for John Bott from AHRQ, for Jeffrey Geppert from Battelle Memorial Institute and for Patrick Romana from UC-Davis. Are any of you on the telephone?

(No response.)

CHAIR MORRIS: Just silence. Okay. Anybody else want to go next, in terms of the measure developers introducing their measures?

(Off mic comment.)

CHAIR MORRIS: Okay. So that's what we have. All right, great. We're moving along at quite a rapid clip, which is not always so bad. Yes, I'm sure. Okay. So let's go ahead and get started then for Work Group A, Measure 0113, and that was Dr. Wilhoit.

DR. WILHOIT: Measure 0113 is titled "Participation in a Systematic Database for Cardiac Surgery." This measure assesses whether an entity is participating in a multi-center data collection and feedback program that provides benchmarking relative to peers and uses process and outcome measures.

Work Group A did review this measure and had a number of comments in response to it. First and foremost, the work group felt that there is value for an entity to participate in such a database, and that there is evidence that participation in a database leads to improved quality of care.

There were a lot of positive

comments related to the sense of the measure. However, there was also discussion within the work group about whether there was value for NQF to have this as a stand-alone measure, particularly given the other measures being considered today that require basically database participation. So you know, that was certainly one of the considerations of the work group.

Regarding the four NQF criteria, first of all, Importance. While the work group could see the value for public reporting, it was less clear that there was value for internal quality improvement purposes. Additionally, since such a high percentage of entities are already participating, it was not clear whether there was opportunity for a lot of improvement.

Second, in terms of Scientific Acceptability, the work group had some questions about the numerator specification, which requires participation but does not

define what it means to participate. Is submitting one case participation, or does participation require submitting 100 percent of cases.

Also of note, while the measure is about systematic databases for cardiac surgery, the numerator says absolutely nothing about the database being related to cardiac surgery.

Secondly, in terms of Scientific Acceptability, the measure is not a rate, but rather indicates that an entity either participates or does not. So there's no denominator. However, despite the lack of a denominator, the form gives details in terms of age and gender about the target population, which seemed a bit confusing.

We did feel that it would be of benefit to have a clear statement about what types of entities are eligible to report the measure, and that that might be an area where there could be improved clarity.

In terms of Usability, the primary issue identified here was the question as to whether the indicator is redundant, and whether it remains useful with the addition of the other indicators that are dependent upon participation, and in terms of Feasibility, there were not any specific issues identified.

CHAIR MORRIS: I have a question about that discussion around this. Suppose that, I think this is unlikely, but suppose that none of other related measures are endorsed. Would it be worthwhile to endorse this measure in that case?

DR. WILHOIT: Well, that's obviously a matter for the group to discuss. But that certainly, you know, could potentially change that, yes.

DR. ROGERS: Well, it wasn't clear to me that you are actually in favor of endorsing this one or not. In the nature of your comments, it was unclear.

DR. WILHOIT: Oh, me personally?

DR. ROGERS: Well, yes. I think we are going to queue a lot off of the presenter.

DR. WILHOIT: Right, and overall, I had trouble with this one, as an individual. Again, not anything negative in terms of -- I mean participation in a database seems like a really good idea.

The way this was written up seemed really confusing, lacked clarity and whether it, you know, and clearly this is dependent upon other measures being approved. But whether it -- its role in light of all the other measures just seemed really ambiguous as well.

DR. DUTTON: I was initially concerned about the redundancy of this measure, and we discussed this on the phone, with the fact that all the other measures are reported through a registry. So you would seem like you would get this.

On the other hand, registry

participation is a good marker for quality in a program, and even though it's already over 90 percent, as presented by the developers, I think it needs to be as close to 100 percent as you can possibly get.

Obviously, if only some practices participate in the registry, you have an inherent selection bias as to whose data you're capturing and it's then going to affect everything else you do. So I am in favor of this measure. I do think, from what Carol said and the technical points, they need to define what constitutes a qualifying registry.

I mean this is all presented from the STS point of view, but there are other cardiac surgical registries, as Dr. Torchiana mentioned on the phone, and Carol raises the very valid question. What makes a registry that would give you a yes on this measure?

CHAIR MORRIS: Any other comments or questions?

DR. HALPERN: I think also Carol's

comment about having a denominator to judge what, how, like she said, if you put one case in, does that count as participation.

MR. FINDLAY: Yes. So let me get this clear. There's no way who would say, who would acknowledge to this one single measure no position, no real practice, without also essentially acknowledging to others or fulfilling the others, right?

In other words, this is as a stand-alone measure, it's sort of, you know, not pointless to your point. But it's not really essential. Is that what I'm hearing?

DR. WILHOIT: Well and I think, you know, from looking at a group of measures, that if there were no other measures being reported, if you weren't looking at mortality, if you weren't looking at outcomes, if you weren't looking at complications, if you weren't looking at the output of the database, then the fact that somebody is participating or not becomes more meaningful if that's all

you have.

MR. FINDLAY: Okay, that clarifies --

DR. WILHOIT: But if you've got the outcomes, the complications, the mortality and so on, then the value of this as a yes or no stand-alone measure seems less clear. The yes or no would be presumably for a surgical practice for a hospital, for a state.

But even there it's confusing, because if you look at a hospital, a hospital could either participates or doesn't. A surgical practice either participates or doesn't. The measure refers to units, I think, such as states or counties.

But a state or county doesn't participate or not. You might be able to say that 80 percent of the facilities in a state participated. But this isn't a rate measure. It doesn't have a denominator defined. So it doesn't really allow you to report a percentage for a geographical area.

Again, that's why I had so much trouble with this measure, was trying to understand really what it was getting at. If it's a practice or a hospital, I understand it better. But it refers to larger entities as well.

CHAIR MORRIS: Can STS respond to that?

DR. PRAGER: Yes, I'm happy to respond. I think some of this needs a little bit of historical perspective. As we all know, the clinicians around the table, creating opportunities for our colleagues to submit data to a registry has not been an automatic or a given in anyone's practice, or at least in our lifetimes to date.

I think one of the things that helped the STS a great deal was NQF having this as a measure, obviously supported by the professional society, the STS. So I think I would share the comment that was made, that we're not at 100 percent.

We would like to be at 100 percent, and having -- while there is poor wording, there is no question, and this can be clarified, and neither Jane, who's on the phone, nor I were part of the wording. But we recognize it when we read it over that this is not clear what it means to participate.

Having said that though, this has helped the STS and frankly surgical practices regions, and we haven't defined it by regions. There are many states that mandate this participation, and perhaps that would clarify aspects for you, including Massachusetts and other states.

So having said that, understanding the wording needs clarification, I think from the STS' perspective, if I may be that broad, this is very important to us, to encourage others to participate, so frankly eventually we have a 100 percent capture of cardiac surgical cases that are done in every practice in the United States.

To clarify just three things. To participate means you submit all cases you do every year. You are audited. We are at about a five percent audit. We will be at a 20 percent.

So every fifth year, every site participating in the STS will be audited for completeness. So that is running parallel to the expansion of the database. I hope that answers some of the thoughts.

DR. SAIGAL: I had a question. So does this measure allow you to record whether a practice or hospital participating, and the question is the regional definition is not clear?

DR. WILHOIT: The wording, I believe, in the measure is generic, and --

DR. DUTTON: I think we eventually heard from STS that this is on the practice level, that it's the unit of -- your unit of collection or capture is a group of cardiac surgeons. Is that correct?

DR. SAIGAL: It actually can be both. It can be both, and it can -- it can be either/or or both.

DR. WILHOIT: Under the level of measurement analysis, which 2.A-32 through 35, it says "Check the levels for which the measure is specified and tested. Clinicians, group; facilities/agency; population, national; population, regional/network; population, states; population, counties or cities."

So it lists all those different entities, but there's no definition about how you get there. The only definition is the numerator; you participate or you don't.

DR. SAIGAL: Okay. Well, I hear what you're saying about that. I do think, though, that for the value of this in terms of practices in hospitals, you get information about people that aren't participating in a database.

So although there are other

measures that help you understand what's happening in the folks that are reporting, this measure would help you get information on people that are not reporting to a database in that sense.

DR. SEARS: If we consider this, do we need to take into account what this means for other databases and the implication of whether NQF supports those? Because I'm sure, for instance, the American College of Surgeons would probably love everyone to be in -- to have the implementer here and be in the NSQIP.

So what's the implication in terms of focusing on one specific database?

CHAIR MORRIS: I think that's a good question, what precedent does this set. I guess, you know, my response to that is that the STS cardiac surgeons have been recording their data and working together to improve. There are a limited number of operations for a long time.

In general surgery, obviously we have an enormous number of different kinds of operations that we do on many different kinds of people, and it's much more -- it's much harder to get your arms around it than it is around the operations in thoracic surgery and cardiac surgery.

So we've been -- so in many ways, the STS and cardiac surgeons have really led the way for us, and I think that rather than the implications really being negative, it's very likely that general surgeons will be learning from how thoracic surgeons organize it, and everything won't be applicable to us because we're in it and have to deal with it a lot more.

DR. BURSTIN: And just to add to that, and this has already come up. We've had other structural measures like this submitted by other surgical disciplines.

We also had a measure endorsed last year, we'll send the details out, or two

years ago, to all of you, which was a generic measure that came in through our Health IT Structural Measures project, which is participation by a hospital, physician or other clinician in a systematic clinical database registry that includes consensus-endorsed quality measures.

So there is now a more generic measure, and I guess one question for the group would be is there a need to continue to have the narrower measures, if in fact a more generic measure would allow, for example, the urologists to come forward, the various groups.

As an internist myself, it would be nice if some of my colleagues came forward as well, not just you guys in surgery. You're so far ahead. But just, it's just a thought and a consideration, and we'd be happy to share those detailed specs with the group.

You should still vote on this measure on its own, but again, in the next

phase, we can continue to think about whether there are opportunities to either improve this one or that one.

DR. MORTON: I was just going to mention that I think it's important to endorse the specific measure, because it's something that hospitals use to get support for the database.

A lot of these databases are supported by hospitals, and without ratification of this, then that support may not be there. So it's a theoretical consideration.

DR. CIMA: I have a question about that, to follow up, is attribution. If a hospital, if a cardiac surgical group doesn't want to participate, let's say in a hospital, that they work in a hospital and another one does, I mean how does this get attributed to the hospital?

I mean we have to realize that there are -- what we're doing here is people

are looking at it and making choices. But sometimes, it's beyond the control of those institutions or those individuals to do it. If the group wants to do it but the hospital won't pay for it, who's at fault, and who's net quality are you tracking?

So you know, if you're going to measure outcomes, measure outcomes. Is this really an outcome? This has so many people involved in it that have different stakes in it and different participation, that whether or not it really is going to help, that's my concern, is who are you going to attribute this to? Whose quality or whose measurement are you supporting?

That's the only concern with what Carol's saying. It seems very vague, you know. Is it the state? Is it the government? Who's being held responsible?

DR. DUTTON: My question about this measure is if we endorse it, do we want it to have a denominator or not, and if we --

because I think people are looking at it in two different ways right now.

If we have -- if it has a denominator, it's presumably something like the number of cardiac surgeries done in the United States. Who calculates that? I mean who is responsible then for maintaining or calculating the measure, and I'm not sure I understand that.

CHAIR MORRIS: I agree. I think that's an important structural issue, and I think that -- I really agree. When the denominator is unclear, it's hard to get to the answer for my question for the STS, which is after the initial endorsement of this measure, how do you know what the participation rate changed to, from and to, with sort of an unclear denominator here?

DR. PRAGER: I'm not sure I can answer the specifics from 2000 -- well, for the last eight years. What the STS has used is the calculation of the number of hospitals

in the United States that do cardiac surgery, and then looking at who is submitting to the STS, either groups or institutions.

Basically, as of January, it was felt that there was a 95 percent penetrance of all hospitals that do cardiac surgery were currently submitting either their group as a group submission or the institution as an institutional submission to the STS, understanding when they started this database in 1989, there were 50 groups, 200 groups, and it has continued to rise.

DR. STAFFORD: So does the STS know how many groups participate that also aren't associated with a hospital that participates? I think that's some of what we're trying to get at, and that's what muddies the waters a bit with this.

DR. PRAGER: Jane may have to help me on the phone, but this is -- it's an important question because actually we have, and I don't want to step into something to

confuse it, but we have public reporting now, two vehicles. Consumers Union, and that is out as institutional public reporting, but the STS has public reporting now on its own website, that is both by group and/or institution.

So that doesn't answer -- that doesn't really clarify it, but we have both of those sources available in the database. Jane, can you expand on that?

DR. HAN: I can certainly look into -- I don't have the number of groups not associated with hospitals that þ- I don't have that information right now off the top of my head, but I can certainly investigate that and get back to the steering committee.

DR. SHAHIAN: This is Dave Shahian. We have done pretty extensive mapping over the past year or so to look at this issue. We contract and have always contracted with participant groups. In most cases, a group tracks to a hospital, and most

hospitals have one major group. So there is typically a one to one mapping.

There are some instances where a group will travel to multiple hospitals, or a hospital will have multiple groups. But that is a distinct minority. So in most cases, it's a one to one mapping.

DR. STAFFORD: That's pretty much what I would have expected. I guess the other reason this is important to talk about is, as all of us who are clinicians know, now with maintenance of certification in various fields, you actually, as part of your maintenance of certification, have to participate in quality improvement and performance improvement projects, and submitting your data to a database and having that data evaluated is important. So in some sense, it does make this an important thing for clinicians who are out there.

DR. CARPENTER: I would just like to get back to one point Carol made, which was

whether this was redundant if you're participating in these other measures. I think participation in a registry is really fundamentally different than reporting outcome measures from your own internal database.

So that this really should be a stand-alone measure that we should keep. I would argue for that and that reporting all these other measures doesn't mean you're participating in a registry. I means you have a database, maybe an internal database, but it doesn't mean you're reporting these as a group with patient-specific level data at each point, which a registry can do and can be much more powerful potentially.

So I think this should be a stand-alone. We shouldn't assume that this is redundant if the other measures are being reported on.

DR. SEARS: The question I have is is this a measure that's for individual practicing physicians or institutions. Maybe

what we should do is fracture this measure, so it looks at either individual surgeons and then as a group the facility itself.

DR. AFSAR-MANESH: I was just going to add on to what Jane said as far as in a data-starved profession, I think what the STS has been able to do with getting the various groups to submit data has really been key. I think we've brought up a number of different challenges and barriers that really warrant us to not only discuss them but figure them out as we move forward because we do need other, not just surgical again, medical specialties creating databases like this as we move forward. So I also do think that this is a stand-alone on its own.

MS. ZAMBRICKI: My question is one of clarification. Looking at the numerator statement, whether or not the facility participates in a multi-center data collection feedback program that provides benchmarking, et cetera, then the numerator details

"participates in STS database."

Is this measure specific to STS database, or is it like the title the measure implies, that it is participating in a multi-center data collection feedback system?

CHAIR MORRIS: Can you guys clarify?

MS. ZAMBRICKI: The reason I ask that is because as a health care executive at a large hospital, I remember we were comparing outcomes for certain cardiac surgery procedures with other large medical center/academic centers, and there was one in particular that was reporting using different definition and different criteria for mortality. They were not using the STS definition of mortality. So I'm just wondering.

CHAIR MORRIS: Can you guys clarify if you're specifically referring to the STS registry?

DR. PRAGER: If it's not clear, it

will be. Yes, we are.

DR. WILHOIT: So just -- that is, actually, I had not noticed that. But there is a disconnect there between the numerator statement and the numerator details, where one refers to STS and the other one doesn't.

CHAIR MORRIS: Okay. So those are some -- okay.

MR. FINDLAY: Helen, just a clarification. The broader measure of participation in a registry, could you repeat that? How mature is that measure? When was it implemented? When is it going to be? I missed a little bit of that.

DR. BURSTIN: It was endorsed, I believe, at the -- let me just pull it up real quick -- it was endorsed at the end of August 2008.

MR. FINDLAY: So it's not in place -- it's not in the field.

DR. BURSTIN: It actually is being used. CMS adopted it as part of the hospital

program. They actually modified the measure to add hospital to it. It's being used, for example, as part of currently payment for hospitals around nurse-sensitive measures and stroke measures, as I recall. So it was intended to be generic enough to capture --

MR. FINDLAY: So it's not a broad-based --

DR. BURSTIN: Actually, it's fairly broad. I mean literally it says, the description is "Participation in a systematic qualified clinical database registry that involves hospital, physician or other clinicians submitting standardized elements to the registry. Data elements are applicable to endorsed quality measures. The registry must include at least two NQF-endorsed measures and report on all patients eligible for the selected measures.

"D. The registry provides calculated measures results, benchmarking, QI information on individual hospitals,

physicians and clinicians. The registry must receive data from more than five separate practices and may not be located at an individual hospital or practice.

"Participation in a national or state-wide registry is encouraged for this measure." So it specifically tries to get at that point. Then lastly, "The registry may provide feedback directly to the hospital provider's local registry if one exists."

So it's quite broad, and the question would really be, you know, is there still a need, if the STS measure would fit under this, that an STS stand-alone measure would need to persist. This has been an issue that's come up before, particularly when the NSQIP measure had come forward, which I believe did not get through, saying do we really want to go down this path of bringing, you know, a measure in for every stripe, to say yes, we have a registry, we're participating, as opposed to a more generic

measure that could be more encompassing.

DR. CIMA: But that would then brings the issue of, you know, what if you are participating in the NSQIP multi-specialty and has cardiac surgery in it? Then you're no longer participating in STS.

Or if your hospital's participating in UHC, which does collect cardiac surgical outcome data, maybe not to the specification, it uses administrative databases. So we're basically then saying if we endorse this, you have to pay STS to do it.

Which is, I think, is not what the purpose of the NQF is supposed to do. It's supposed to look at quality outcomes. You're almost mandating participation in a private entity's process.

DR. STAFFORD: Yes, and my understanding is, and correct me if I'm wrong, but the STS database is not the only cardiothoracic database that's out there that does this work.

DR. DUTTON: Dr. Torchiana mentioned two others on the phone when we were talking. One was the Northern New England collaborative, that's been looking at cardiac surgery outcomes for a very long time, and I don't know if that rolls into STS, and the other was the New York state mandated registry.

DR. PRAGER: The Northern New England does not automatically roll into STS, number one. The eight sites do not. Some of the eight sites currently are part of it, and others are now considering joining it. It certainly was the gold standard of early databases. New York state, there are many participants in New York state, although it's not 100 percent yet that are in the STS as well.

MS. ZAMBRICKI: I'd like to speak in favor of a measure that requires participation in a database. It does drive behavior. It does change how people look at

their work, and as far as the issue of whether it should be specific or not, I think that is -- one, it is very helpful having this discussion. The point that I want to make is I believe enrollment in a database, where you compare across institutions performance and outcomes, does have an impact on quality.

CHAIR MORRIS: Okay. So we've all made several points. Do you also have one Ruth?

DR. KLEINPELL: I did. I think, you know, Christine, you brought up an important point. As the measure reads and as the numerator statement reads, it's broad. But then in the numerator details is where you see it's specified for STS. So if we endorse that as it's written, we are saying STS.

Now at this point in time, can we ask for modification and clarification of that language in the numerator, or is that a whole separate process?

MS. MURPHY: May I comment on

that? What you need to do as a first step in the process is to vote on the measure as submitted, as specified. If the measure fails, as specified, then you have an opportunity to identify any conditions that you would want to have considered in order to find it as meeting the criteria.

CHAIR MORRIS: So specifically we could ask for -- if the measure fails, we could ask for a clearer definition of the registry throughout the measure, and then also a clearer definition of the denominator throughout the measure. Any other -- if there is anything else that we'd like a clearer definition on, then this is a good time to succinctly bring it up.

DR. ROGERS: Is there not a necessity to sort of look forward to what we have in front of us also because if in fact the rollout of a number of measures that we are yet to evaluate actually requires the kind of information that can actually only be

obtained through being in this registry, to the extent that will populate all of the other measures that we're going to look at, I think that needs to be considered also.

I'm not savvy enough to know the details, but the other cardiac surgical registries, I doubt, have the content and the complexity that STS has, that would allow us to actually move ahead with all the others that we have in front of us. So that's just something --

DR. SHAHIAN: This is Dave Shahian. If I could just make a brief response to that. The other clinical data registries, like NNE and New York, are superb databases. They suffer, however, from the fact that they're not nationally representative. I think that is the distinctive feature of the STS, that it permits national benchmark referencing.

CHAIR MORRIS: Okay. I think -- is there anything else anybody else wants to

say about this before we move on to a vote of the individual criteria?

DR. KLEINPELL: Yes. I just have one other question. This is not a new measure, so it's been in effect, and so this is not new language; correct? So it has been endorsed previously as it's stated? Okay.

CHAIR MORRIS: All right. Let's go ahead with the vote, then, on the individual criteria.

[COMMITTEE VOTING.]

MS. MURPHY: So you see, you're voting on the first criteria, importance of the measure, based on the evidence.

CHAIR MORRIS: Okay. So we have 18 responses for yes and 4 responses for no on the importance of the criteria. Do we move on to the next criteria or is there -- do we want to have more discussion now?

MS. MURPHY: We should move on to the next criteria, and if there is any additional discussion prior to vote on

Scientific Acceptability.

CHAIR MORRIS: Okay. So time to vote on Scientific Acceptability.

[COMMITTEE VOTING.]

CHAIR MORRIS: Okay. So in the summary of responses for Scientific Acceptability, we have 4 say completely meets the criteria, 15 say it partially meets the criteria, 1 says that it minimally meets the criteria, and 2 say not at all. Next, we're voting on usability.

[COMMITTEE VOTING.]

CHAIR MORRIS: I think we're waiting on one vote. If everybody puts their vote in one more time, then it won't record twice. But if for some reason it missed your vote, then it will be recorded.

[COMMITTEE VOTING.]

So summary of responses, we have 9 that say it completely meets the criteria for usability, 13 say partially, 3 say -- oh, nobody says minimally and nobody says not at

all. Next, we are voting on criteria for feasibility.

[COMMITTEE VOTING.]

CHAIR MORRIS: Okay, and in terms of the vote on feasibility, 17 say that it completely meets the feasibility criteria, 5 say it partially meets the feasibility criteria.

So then the next vote is whether this measure meets all of the NQF criteria for endorsement. We had several different votes on this. Now do we need to all say that yes/no or do we need to have a majority? What do we do? Do we have more discussion here to try and reach better consensus?

MS. MURPHY: If there is any other discussion points that need to be brought forward, then yes. Otherwise, the group can vote whether or not each individual sees the measure as meeting the criteria.

DR. BURSTIN: And that's separate from a "do you recommend the measure move

forward," which you'll do after you've had a chance to look at competing measures. This is basically just the test on its own. Does it meet the criteria?

CHAIR MORRIS: This is whether it meets the criteria. If for some reason we decide that it does not meet the criteria for endorsement, then we have the opportunity -- am I correct in saying that we have the opportunity to ask the STS to make further clarifications?

DR. BURSTIN: Yes, you're voting as is.

CHAIR MORRIS: Okay.

DR. CIMA: So if we want them to take out, strike STS and make it a generic, is this the time to do it now?

CHAIR MORRIS: What we would do is vote no, if that's what you want, vote no on this, and if -- ultimately if this has a majority of no votes, then we would say what it is that we're looking for.

[COMMITTEE VOTING.]

CHAIR MORRIS: Still waiting for one more vote, so please hit your markers again and hit send. Okay.

[COMMITTEE VOTING.]

CHAIR MORRIS: So we have a dead tie. Why do we have an even number of people voting?

DR. BURSTIN: Because David's not here. You're not supposed to.

CHAIR MORRIS: Oh, yes. So let's clarify the conditions that we'd like from the STS for this measure.

DR. DUTTON: I'll start. I'm supportive of the measure. I think it's important. I would be perfectly in favor of a measure that required registry or submission to a registry that looked exactly like the STS registry, but didn't mention it by name.

CHAIR MORRIS: So one of the clarifications is that exactly which, what the registry is and what qualifies as a registry?

DR. DUTTON: Yes. I'm supportive of the concept, no question. I just think it's a mistake for NQF to endorse a particular registry by brand name.

DR. DILLON: So then how does the STS then become the steward of such a measure because now you're appealing to multiple databases that they may not have access to. How do they then measure? Can we -- I don't think we can ask them to measure, you know, what's going on in other databases.

MR. FINDLAY: Yes. Doesn't that fragment the world of this?

DR. DILLON: If -- it requires a larger governing body to be able to collect that data. STS isn't in the process of collecting who's in or who's out in NNE or the New York database. It's a higher level now.

CHAIR MORRIS: One thing that we could request is that there's uniformity of the descriptions throughout the measure. That was one of the sources of confusion here. But

that doesn't really speak to the point of the concern about branding.

MS. ZAMBRICKI: Could we request that the measure require that the other measures that are listed later be collected by the body, that that's the criteria, that they collect the other cardiac measures?

DR. HALPERN: How does the more broad one that you mentioned, how are they collecting the data from the various databases?

DR. BURSTIN: It's self-report by hospitals or groups to -- and again, it describes what's considered a registry with adequate numbers. You know, there are some issues that STS does that are in a special like auditing. So you know, it's not a one to one match, but you should certainly take a look at it.

But it would be self-report by hospital, physician or group, that yes, we participate in a registry that matches these

characteristics. I do think STS could not realistically get information on other people's -- in other people's registries.

It would still be very limited just to cardiac surgery. So it's, I think, more of a philosophical issue than a feasibility issue, actually.

DR. AFSAR-MANESH: Well, and I understand the concern with it being STS-owned, but I think at the end of the day, I don't really foresee another body that could ever come in, rather than a professional society, to collect this type of data.

So I think, again, looking at other subspecialties, at the end of the day it's likely going to be a professional society stepping in to do this.

So I think instead of being uncomfortable with how STS is doing, probably it's better just to make sure that we're defining it appropriately for the data that we want collected. But we are going to need to

have one body so that we can compare data nationally.

DR. HALPERN: I will just mention, coming from the VA system, that the VA does have a system that measures nationally that isn't a specific society's.

DR. AFSAR-MANESH: But it's just the VA, correct? Yes. So, I mean, that's still very limited compared to all the other hospitals.

DR. CIMA: So basically you're saying that if a society comes up with their own set of rules, that we're going to have to follow those rules. You know, you're picking a winner. I mean, I'm not arguing participation. But I'm saying if you're participating in the NSQIP multi-specialty, which has cardiac in it, doesn't that qualify?

DR. AFSAR-MANESH: Sure, and I understand that. I guess it's not that I think we should pick a winner. I think we should pick the criteria that we want for

quality and safety of patients.

But we do need to have one body doing that, versus five different ones, where we're not going to be able to compare all the different groups that are out there. That's what I --

DR. CIMA: But then you're saying we have to pick a winner, because if you say oh, I'll let you participate in NSQIP but you still also have to participate in STS to do this measure.

DR. AFSAR-MANESH: Well, how else would you propose that we could compare data nationally as we move forward in ensuring --

DR. CIMA: Well, you either have to have one national database, or you're going to have to figure out a better rule. But is this the rule that does it?

DR. AFSAR-MANESH: And I guess I don't know any other organization that at the end of the day can come in and have a national umbrella to cover that, and that's my concern.

DR. HALPERN: The American College of Surgeons is what he's saying, the NSQIP multi-specialties out of the American College of Surgeons.

DR. AFSAR-MANESH: I understand, but from my understanding of what STS does, is that their database is actually more in-depth than what NSQIP offers right now for cardiac surgery, correct? Not for cardiac?

DR. CIMA: I'm not arguing for or against it. I'm just saying this measure, which we're looking at, this one specific measure, says at the top "Participating in a cardiac registry nationally," and then throughout it says you have to participate in STS.

That's the only thing I'm talking about. I agree with everyone saying participation in a registry that provides you feedback is great, I think. But are we in the place of picking the winners? Are you saying you have to use STS?

Then the fundamental problem becomes, as Peter pointed out, for all the other cardiac ones that have been presented, it uses the STS database. So then are we saying that we have to endorse that everyone participates in STS? I'm just pointing that out there. As you go down this road, then you are picking a winner.

CHAIR MORRIS: At some point we will be picking a measure steward, and the only measure steward here is the STS. Helen, do you have a point?

DR. BURSTIN: I just want to make one clarification. So NQF has endorsed measures that certainly come from various registries. That's not an issue. The important thing to note that the specifications that STS promulgates are fully transparent and open to anyone.

You don't have to submit your data to STS, but you can go ahead and use these data, including the risk model. It's all

publicly available to do your patients. It's not easy to do, but it is doable, and that's the requirement from where NQF sits.

It's fully transparent, anybody can do it. I don't want to confuse the issue of this measure, which is a structural measure, yes/no, do you participate in a registry, which is, I think, getting more at your issue. But I think the overall issue of the quality of cardiac surgical care and the STS being the data source for that. Exactly, right.

DR. SAIGAL: And just to this point, I think pragmatically, I mean other entities could bring forward measures in a similar vein, and they'd be harmonized with this.

DR. BURSTIN: Or competing, or competing.

DR. SAIGAL: Or competing. So to the extent to which there are other measure stewards who have equivalent data sources, we

could eventually, when the time is right, harmonize them. Right now, STS is the one that's in front of us. I mean, it's certainly better than not having the measure, in my view, at least.

DR. WILHOIT: The other things that I think need clarification, one is the definition of participation. Is it one case? Is it 100 percent? Is it somewhere in between? But some kind of a clean definition of participation.

And also clarify with respect to the denominator. Is it a yes/no for whatever entity chooses to report? Is it by practice? Is it by hospital, is it by state? But adding some clarity so it's clean.

CHAIR MORRIS: Okay. So we have three things that we're requesting from STS. One is a clearer and more consistent definition of registry. Another is the definition of the denominator, and the third is definition of participation. Can we move

on? Okay.

DR. SHAHIAN: Hi, this is Dave Shahian. I could clarify that right now. I think in terms of the definition, I think if you are going to consider eliminating STS from the definition, it would be important, however, to put nationally representative or nationally inclusive in that statement, which I think we could certainly do.

In terms of what it means to be a participant, it means inclusion of all your cases, and in fact, that is one of the things that we audit when we go to programs. We actually compare the cases that have been submitted to STS with the hospital operative logs, to make sure that all cases have in fact been included.

In fact, we've also done a separate audit and have compared our results for inclusiveness with data from MEDPAR, and it's 98, 99 percent. So that's the answer to the second question, and I'm sorry, the third

question?

CHAIR MORRIS: Definition of participation, the denominator.

DR. SHAHIAN: It's a yes/no. I mean, for a particular participant, you know, you participated or not. I think the denominator in this sense is only of interests to get a sense of national variability in participation. But in terms of the individual institution, you did or you didn't.

CHAIR MORRIS: So it sounds like you're saying that the denominator is institution; is that correct?

DR. SHAHIAN: It's participant group, which in the vast majority of cases is an institution. But in some cases, no. You may have Lakewood Surgical Group, for example, that happens to be one of two groups that practice in a given hospital. You have a few situations like that nationally.

CHAIR MORRIS: Are you satisfied with those responses?

DR. CIMA: So it's 100 percent participation. So again, based on the NSQIP methodology, only 20 percent of cases are captured. So that would exclude them. UHC is based on a sampling provided by CMS, so that would exclude them.

So again, I'm getting -- I agree with the registries. I think it's important. I have concerns about this.

DR. WILHOIT: And I, at least, would want to see the numerator and denominator in writing, to be comfortable of just, you know. I don't necessarily disagree with the concepts, but would want to see what's there.

MS. MURPHY: And in the case of any of the measures in which there is a vote for a measure with conditions, we would always ask that they provide that information back to us from the developer, precisely what they would be able to do. So in all cases, we'd ask for that to come back, written

information.

CHAIR MORRIS: Okay. So we have questions that we'd like to have written confirmation of the answers on or written information on, and we have a dead tie here, and I think that we need more information from the developer before we proceed on this.

DR. ROGERS: I think that last piece of information for me is very useful, because if we are representing -- if this group represents the community and advice to the NQF, and we're looking at options that would include two other programs that only do sampling, there's no comfort in my mind about recommending then.

And although yes, there's some endorsement about picking a winner, if that's the only program that does 100 percent, we have no option in my mind. I mean, I am completely uncomfortable recommending an option that just does sampling. Makes no sense.

CHAIR MORRIS: I think that's important to keep in mind. For this measure itself, this is just about the STS. So those other things don't really apply here, but it will come up.

DR. DILLON: And we have to be careful with that statement, because we're not in a position to compare scientific validity of the other databases. You know, NSQIP, I'm very comfortable with NSQIP. So again, this is really stretching it in my mind what the STS is asking of the NQF.

I don't see this as solvable, because, as I said, I think if we want a participation in a database, which obviously I think that's what we all do agree on, STS cannot answer that, because by definition, it will be a forced or a limited report from them, because they won't have everyone. There's no way of them measuring the overarching denominator, which is what we're talking about.

CHAIR MORRIS: Okay. I think what we need here is more information, and then basically to hold another vote, all right. Anything else before we move on, Melinda?

Let's move on to the next measure. So that was a good discussion, and we'll probably touch on it in our upcoming measures.

DR. BURSTIN: The first measure -- just to make you feel better, the first measure, having now done this for four years, usually takes 90 minutes, so you're ahead of schedule. You're doing fine, and then it just goes much, much faster, because you've kind of gotten it out of your system.

CHAIR MORRIS: So we're all warmed up. The next measure is 0114, Postoperative Renal Failure, and Dr. Stafford is going to talk about this.

DR. STAFFORD: Thank you. Good morning, everybody. Well, at least on the conference call, this was much easier than the first one. So if that holds true today, I

think that will be a little bit better.

The title of this measure is Risk-Adjusted Postoperative Renal Failure, and the description of the measure is the percent of patients undergoing isolated coronary artery bypass grafts without pre-existing renal failure, who develop postoperative renal failure or require dialysis.

In general, on the conference call, I think all of us agreed that this was important to measure, and there really wasn't any controversy about that.

When it came to Scientific Acceptability, there were a number of concerns that we had with the measure, one of those being were there any exclusions for emergency cases, because we know those cases are much more susceptible to the development of renal failure for a whole lot of issues, including need for blood transfusions, as patients are often sicker to begin with.

The anesthesia isn't going to be

as well-thought-out as an elective-type case. Should there be a specified time window for the development of renal failure, and that was not found anywhere on the documentation that we had. So was this the development of renal failure at 30 days, at 60 days, at one week during the hospitalization?

Because it is linked to the process of actually having the bypass graft, the outcome should be measured and should be close in time. So we'd like some more information on that.

Let's see. Then, secondly, we had a question about how the definition of renal failure was actually defined, and why wasn't something like RIFLE criteria used in the development of this measure? Then also, how did you just arbitrarily pick a creatinine of two for an exclusion. So we had trouble with the exclusion criteria as well.

I'll stop there on the scientific assessment, because I think that was one of

the biggest things, and then we can -- the Feasibility was very straightforward.

There were some questions about yes, the costs are low for the database, but there was nothing in the documentation about the cost of actually hiring people to do your data abstraction. So that actually does cost groups and/or hospitals, and everybody felt that the Usability was not really an issue.

CHAIR MORRIS: Thank you. I have a question about what you said, and that is, I thought your first point was that the exclusions were not clearly defined. Is that correct?

DR. STAFFORD: The exclusions were defined. What we couldn't tell was why they chose a creatinine of two as an exclusion criteria. So they were defined, but we didn't know where that came from. There was no documentation for that.

DR. DUTTON: Just had the science point there, and this may be a suggestion more

for future measure development or expansion of this, but wouldn't we also be interested in renal failure worsening in patients who already have it?

DR. CIMA: I had one question about the denominator. When they say isolated CABG, in the risk adjustment from the STS, or whatever the listing, is there -- what about patients who've had a prior CABG, and you're doing a reop?

We know in almost all surgical practices, orthopedics, cardiac, that patients who have had prior surgery, you know, are at higher risk for complications or a sicker patient they're different.

I couldn't see anywhere in the denominator that there's anything about reoperative patients, because you know, there's another measure about using an inferior mammary artery.

Well, what if that's been used, and now you're coming back and doing two new

vein grafts? I mean, is that anywhere in there? I didn't see it anywhere in here.

DR. SHAHIAN: Rich, you want to take that, or do you want me to respond?

DR. PRAGER: David, you can respond. I was waiting for all the questions.

DR. SHAHIAN: First of all, we have in general tried to avoid exclusions. We feel it's much better to include significant factors such as reoperation or emergencies, the two that you've mentioned, to include them in the risk model. In fact, those sorts of things are included in the risk model. So that's how we deal with those particular things. We try to avoid exclusions.

Second, in terms of the specific cutoffs that were used, greater than two and two times preop, those really have historical roots. In discussions with a nephrologist at the time, I think there have been probably more sophisticated subsequent definitions.

But I don't think these are that

far different, and for the sake of consistency, so that we can actually look at this particular complication over years and decades literally, we basically decided to leave it as it is.

We had a lengthy discussion about this when we had our specification upgrade earlier this year. We discussed RIFLE and similar definitions.

But we decided for the sake of continuity and consistency, that it was a reasonable enough definition that we would just retain it as it is. And the question was also asked about the time frame.

DR. PRAGER: Right.

DR. SHAHIAN: We have limited the measurement of post-operative renal failure, and in fact all complications other than death and sternal infection are limited to the in-hospital, index hospitalization.

DR. STAFFORD: Great. That helps a little bit. So I would have two questions

for you. So you talked about consistency in terms of keeping your definition. I have a little bit of a problem with that, because the RIFLE criteria is now being used by everyone. It helps you define where your patients are, where they start, where they go.

We have to code our patients that way in the ICU for our coding purposes. So I'm a little concerned that you decided, after all of that discussion, to leave something like that out, just because it makes it easier to measure.

The fact of the matter is everybody measures creatinine. These patients usually get daily creatinines at least in the early part of the hospital stay. So I can't see that it would be that much more difficult for people to collect that data for you, and I think on a national level it would make the data, for those people who are going to look at it, a lot more robust.

So that would be one point about

that. But I do appreciate that you're measuring it within a 30-day hospitalization. However, what do you do with a patient who's been in the hospital for quite some time?

They sailed through their coronary artery bypass graft and at the end of two weeks, you know, they maybe stayed a little longer than usual and they're about ready to go home. Then they have a GI bleed or they perforate a tic in their colon, end up having emergency surgery and then developing renal failure then.

So there could be a problem with ascertainment bias, because that renal failure may not be related to your coronary artery bypass graft.

DR. SHAHIAN: Anything that happens in the hospitalization after the coronary bypass is something that we own. So you know, if the patient had a GI bleed that led to a cascade of complications including renal failure, we own it. That's the way

we've always done it, and I think that's the right way to do it.

DR. SEARS: David, one other question. What about patients who have pre-operative catheterizations? We all know that dye can be renal nephrotoxic. So do we take that into account?

DR. SHAHIAN: No. I think that's part of the game. I think you should, unless it's an emergency, you have some latitude in when you schedule a surgery.

If the patient has any evidence of pre-catheterization renal insufficiency, hopefully they've gotten appropriate measures during their cath to try to mitigate the possibility of post-operative renal or post-cath renal failure, and their renal function should be checked, and I think you have the responsibility to have the patient in the best possible shape.

On the other hand, if it's an emergency and they get a cath, yes, they are

going to be at higher risk. But I think there, the emergency status that's included in the risk model will help to account for that.

DR. HALPERN: I have a question. Before you said it was only in-hospital changes in creatinine. So let's say the creatinine was sort of trending up but didn't hit your threshold prior to them leaving the hospital, and then two weeks later, after discharge, come up with a creatinine that meets your threshold. Do those get captured?

DR. SHAHIAN: No.

DR. PRAGER: No, no.

CHAIR MORRIS: Any other questions? Did you feel that your question about the source of exclusions was answered, Dr. Stafford?

DR. STAFFORD: Not really. I mean it clearly was listed, but I didn't really -- I mean, they talked about it, but I didn't get where that actually came from. Where did that data come from? How did you come up with that

definition?

DR. PRAGER: You know, I'm not sure. Unless David or Jane can be more specific, I'm not sure I can answer that either, how the definition originally became part of the measure.

DR. SHAHIAN: This goes back probably a decade or more.

DR. CIMA: So this is a historical definition used by just STS?

DR. PRAGER: I'm not sure we invented it, frankly, so as David said, this came from discussions. I'm not sure we're privy to it at this point.

DR. CIMA: No, but what I'm saying is that this is something that the STS has built into their criteria --

DR. PRAGER: Yes.

DR. CIMA: And have been using it?

DR. PRAGER: Correct.

DR. CIMA: So for consistency purposes, they're continuing to use it?

DR. PRAGER: Correct.

DR. CIMA: But if there's new data or something out there, if you want to use your risk model, you have to continue to use this?

DR. PRAGER: At this point, correct.

DR. STAFFORD: But the risk model could be adjusted for RIFLE criteria?

DR. PRAGER: David, do you want to take that?

DR. SHAHIAN: I'd have to go back and look at the criteria. But don't the RIFLE criteria require an assessment at a longer period of time?

DR. STAFFORD: No, actually they don't. You can assess it during the hospitalization.

DR. DUTTON: I guess this is going to be a common issue with a lot of different measures that we talk about. I mean, this measure was created 25 years ago. They took

their best guess at what the criteria is, and they've stuck with it, which has made it very consistent for repetitive use and reporting.

Now we have a better understanding of the disease and we want different criteria. One reason not to change would be to allow continued comparisons with the past. Obviously, there's a science reason to change, because we think we defined the disease. I think we're going to see that tension in a lot of these measures.

Then there is the additional burden of data collection, and the real question about RIFLE is does that require data that the STS is not capturing now?

CHAIR MORRIS: So I think that that's, you know, that's really why we're talking about maintenance here. Maintenance is for upkeep, and if upkeep includes upgrading the way that criteria are defined and collected, then that's what we're here to do. Just because it -- oh, I'm sorry.

DR. BURSTIN: And if there is a national standard, and I just looked it up. The RIFLE criteria certainly looks like that is the national standard, I assume STS already has creatinine and urine output, which is really all that's required to compute it.

There's a long history of measures that do change based on changes in evidence, changes in process. So certainly, I think, if the Committee thinks that's important, it would be an important consideration for STS to consider.

DR. STAFFORD: And actually the RIFLE criteria requires either/or. So you can make the definition based on creatinine, which you're already capturing, or urine output change. So either one, and if you're already calculating creatinines, then it shouldn't be that much more in terms of the database.

DR. WILHOIT: A couple of other things, and this is, you know, throughout the STS measures, but a couple of things that were

a little bit troubling to me. One is like in 2.A-8, the denominator details. The denominator is defined in terms of the STS database, but there's not a specification that you need to be using the STS database.

And yet you can't get to the details without using the STS database. So that troubled me throughout the measures, that it seemed like if it's contingent upon STS, I can live with that. I don't have a problem. But it would be better for me to be upfront about that.

But the way the denominator details are written, you can't read the details and understand what it is without going through all of the fields in the database and trying to understand what they mean. There's no clear specification as to what it means to be in the denominator. It's just all based on fields, which you need to understand the database for.

The second thing that's troubling

to me, and again, this goes throughout, is in 2.C-3, the testing agreement rate, the -- you know, there's a number here that for renal failure there was a 98.5 percent agreement rate.

Well, if you reviewed 200 cases and only three people had renal failure, there might have been disagreement, I think; I'm not sure how that's defined, but I think you might have disagreed on all three cases. But yet it comes up to a high number because most people didn't have renal failure and it wasn't relevant.

So I'm not sure how that's defined, but I never knew quite how to interpret those numbers for the agreement rate, because I'm not sure how many of the cases reviewed had the outcome of interest. And there's no numerator, there's no denominator.

So again, it's left to me, the reader, to assume that the data are good, but

I don't really have any basis for agreeing or disagreeing.

CHAIR MORRIS: One of the other issues that was raised, I think by Dr. Stafford, was sort of the feasibility question, and that was the cost of the data abstraction. Is that something that we tend to get into here, talking about that aspect of feasibility? Does the STS have any response to that?

DR. PRAGER: The cost of data abstraction? We do in the abstract, in that we have -- certain of us have calculated the cost for data management, if you will, and abstraction, and at least in certain areas of the country done it on an FTE basis of to enter 500 cases into the database, and the database is getting, is more robust now. So there are more variables to enter and will be more in July.

Having said that, it is one FTE ballpark for 500 cases. That FTE, though, can

range, frankly, from a data manager who's a nurse practitioner with a great deal of experience, to someone who is looking at the chart and is a coder. So salary would range then between those two. David, do you have any other thoughts from Massachusetts?

DR. SHAHIAN: No. I think that's right. I think there are many institutions that do this with FTE. There are small programs that do it with part of an FTE, and it's the data abstraction costs that are the major costs. Our actual costs to be a participant are really quite small. So it's data abstraction that is the cost.

DR. STAFFORD: And I think that's important, and just you don't address that in the application. So if you had -- the application just says the cost is minimal and it's really just related to the database.

But we all know that it isn't, and so I think to be open and transparent about the process, you should at least say it may be

one half to one FTE per 500 cases, and then institutions can figure out how they would want to spend that money and what kind of FTE they would want to have.

But I think in the application, that needs to be clear. It's not just the cost of the database itself.

CHAIR MORRIS: Okay. Any other questions about this measure? So just to very briefly summarize, the source of the exclusions is a historical decision-making process by the STS. The time window is the hospitalization itself and not after the hospitalization has concluded.

The definition of renal failure is based on a previous definition which the STS decided to continue for comparison purposes over time, and does not correspond to what sounds to be the current standard, the RIFLE criteria. The cost of data abstraction includes one FTE for 500 cases, and that cost, of course, would vary, depending on whose FTE

we're talking about.

And then the exclusions are minimal in order to capture all cases, but they are -- but some of the potential confounders would be adjusted for in the model, including things like previous operation, previous use of the internal mammary artery potentially, or other potential influences on the outcome.

There is a desire for a clearer specification of the numerator and denominator, and ultimately for a clearer definition of renal failure, which I guess is that last one I already said. Anything else that anybody's concerned about with this measure before we vote?

DR. WILHOIT: One question I have is with something like the time window, which is during the hospitalization is what we were told, but it's not in the write-up, it seems like that's just really, really key to have in the documentation, and it's not there at all

right now.

So what's the process? Do we have to basically turn down this measure in order to request that that be added, or what's the process for that?

MS. MURPHY: The process is that you vote on the measure as specified, and if it is voted down, then you have the opportunity to ask that certain conditions be met, be reconsidered.

DR. WILHOIT: So then just to clarify, the only way to get, to request that the time frame be added to the measure is to turn it down?

MS. MURPHY: If there's information that can be provided to clarify what is in fact the case, then that information could be brought forward, yes, without having to vote it down. So if it's something that is known and just was omitted from the documentation, then that could be clarified here and now even, and be able to

vote on that, yes.

DR. HALPERN: I think what she's asking, though, is -- I think what you're saying is you feel uncomfortable voting on it affirmatively if it's not in writing, or at least going to be added?

DR. WILHOIT: And I think -- right. But maybe not, I mean, I trust it. That's not the issue on that, and that's pretty clear. But in terms of the measure that goes forward, the way it's documented for the public, the way it's documented for public review, the way it's documented for the next time it's reviewed, it just seems like it's really, really key that it be there.

And the fact that we understand it in this room today is one thing; whether we'll remember it in six months when we look at it or whether somebody else reading the measure has the opportunity to have that information is different. And it seems to me like that's really important to have in the document

itself.

DR. BURSTIN: And that's certainly something you can request as part of the follow-up from the measure developers, to clarify that whatever needs to be in the documentation gets added to the documentation.

I think the issue that you raised though, about whether they would shift to the RIFLE criteria, I think, is something more substantive that you would need to potentially make a condition if you thought that was appropriate.

DR. CIMA: Just one question to clarify. This has penetrance of about 93 percent, they said, STS. So what about the organizations -- this is a separate measure from the one we discussed previously. What about organizations that don't participate in STS? How are they going to report? This is a national quality initiative. They only have five percent of the practices that don't participate.

DR. WILHOIT: And the denominator definition requires the STS fields. So the denominator definition is not really flexible, at least as I read it, to take data from some other sources, or would require a ton of work to map and create the STS fields in some other database.

DR. CIMA: And to further that, my concern about, you know, emergency case versus a reop CABG was mentioned that that's in the risk adjustment model. But that means then you have to be using that model to adjust your patients, which therefore mandates that you participate in STS.

DR. BURSTIN: The model is fully transparent. They give you the actual intercepts, the whole thing, published annually. So somebody could conceivably take it and run it. They would unlikely be submitting it then to STS. It would be something they might do on their own, so they could then -- to STS?

DR. CIMA: So that's an additional burden that they would have to do, a reporting burden? No, it's not potentially. It is, right?

DR. BURSTIN: If they choose to do it, yes.

DR. CIMA: Yes.

DR. DUTTON: I think the fundamental question is about the difference between a measure steward and somebody reporting on a measure. So that the steward's job is to put out a rational measure that makes sense of quality, that if we were going to define quality, this is how we would do it in this area.

Presumably, in most cases, the steward is also going to be reporting on the measure, because they have some expertise in it. But it doesn't, as long as it's transparent, it doesn't preclude anybody else from reporting on that also; correct?

CHAIR MORRIS: Anything else

before we vote?

DR. HAN: This is Jane Han from STS. If I may add just regarding how the specifications and the field names are presented. As the measure developer and steward, we are instructed to provide detailed specifications that we can measure and report upon and maintain.

So we use data field names that are used in the STS database for that purpose. But data specifications, definitions, code names and what they stand for, they're all provided in the supplemental documentation that we provide, and it's also publicly available on the STS website.

So it's not that we are mandating that STS be used. As Dr. Burstin had stated earlier, everything is transparent and available online. So it can be used by, theoretically by organizations that don't participate in the database. It's a preference that they do, but it's not

necessary.

CHAIR MORRIS: Okay. Did you have one more thing to say before we vote?

MS. ZAMBRICKI: Just one more thing. It seems that there are going to be a number of measures where this issue is going to come up. So I think it's really important for us to talk about the content of the measures, and to recognize that this same fundamental discussion about the source and the stewardship and the options is going to be there.

I wonder if it would be possible for our learning, for staff to do some type of a summary of the different potential measure, I don't know what they're called, like STS and the others that have been named.

Not stewards, because the stewards are the one that brings the measure forward, but the VA, the different systems that are to collect this information on a multi-clinical site basis, and whether or not they do collect

the specific measures that we are asking for down the line, like renal failure, like mortality within 30 days, etcetera.

Because I personally don't have any knowledge of the other ones on a basis, and is it possible to know what the enrollment is in those other systems, so we know what the scope is?

DR. BURSTIN: Just to try to set some benchmarks, just to be clear, you know, we have endorsed measures from STS in the past. We have in fact brought in a couple of NSQIP measures last year. We have numerous measures that emerge out of the American College of Cardiology registry database.

CMS has adopted several of the NSQIP/ACC measures, and has actually put it forward that you can either submit to NSQIP or ACC, or here, they're going to be developing an alternative data platform for you to submit the data individually.

So I don't think in and of itself

we view the registry as being something that would hold you back from saying the measure is well-defined, all -- it meets the criteria. One of the issues is when you get to Usability, and we've talked about this a little bit in the past, is this issue of does the data -- since they are both the steward as well as the holder of the data, what's the transparency of the data, which has been certainly something that's come up in the past.

But the bottom line is we have felt very comfortable endorsing those measures because the steward is fully transparent. Everything is available. Again, I understand it's clearly a burden, but it's a burden if you're in or you're out, and you're collecting your data some way.

So that those data are available -- the measure specifications are fully transparent is our requirement. We can't force hospitals or use or not use it. We

don't make the ultimate decision. The end users do, whether that's CMS or other payors that says you must use X, you must use Y.

But the NQF endorsement is there to indicate that the measure is important, reliable, valid, precise specifications to compare, usable and feasible. We can't make the ultimate end decisions. That's for others to make.

But how the measure gets used if it's picked up for payment, if it's picked up for public reporting, your job is to say do you think the measures meet the standards that we've set forth.

I do think the issues around feasibility are fair play, the amount of burden in terms of collecting these data are things that are going to come up on all, on many of these measures, because they are such rich outcome measures.

It's hard to do them quickly off of claims data, for example. But I don't

know, as was pointed out. You don't have the information in front of you of the other competing, more regional registries. I think it would be difficult for you to really make that assessment today.

DR. STAFFORD: Helen, I think your point about trying to separate the two in terms of we put the endorsements out there and other groups, whether it's CMS or even JCAHO may adopt them, or AHRQ or whoever wants to look at them, it's up to them.

But I think what everybody, what I'm hearing underneath all of this is that everybody realizes that they aren't totally separate, that if NQF puts something out there, the likelihood of CMS adopting something is actually probably going to be pretty good.

So they are linked, and I think that's the importance that you're hearing everybody really struggling with thinking about this, about you know, putting a winner

out there, because that's exactly what does happen. We know that's what happens, and while it may not be NQF's goal to have that happen, it's the reality.

DR. BURSTIN: And that's true, and I think you can only evaluate what's before you today, I guess would be my last comment. You've only got this one on the table. We don't have something else for you to look at that you think is superior or not superior.

So I think you have to weigh the criteria and think about whether it's worth having a measure like this out there, to drive public reporting and quality improvement in the field of cardiac surgery.

CHAIR MORRIS: Let's move on to the vote. So the first criteria is Importance to Measure and Report.

[COMMITTEE VOTING.]

CHAIR MORRIS: Our results are 100 percent of the people said yes. 22 responders said it's important to report. Second item is

Scientific Acceptability of the Measure Properties.

[COMMITTEE VOTING.]

CHAIR MORRIS: And we have 3 say completely meets criteria, 18 say it partially meets criteria and 1 says that it minimally meets criteria. Next is Usability.

[COMMITTEE VOTING.]

CHAIR MORRIS: And we have 12 responders who say that it completely meets criteria for Usability, 9 say it partially meets criteria and 1 says not at all. Next is Feasibility.

[COMMITTEE VOTING.]

CHAIR MORRIS: We have 14 said it completely meets criteria for Feasibility, 8 say that it partially meets the criteria for Feasibility. Then our last vote is does the measure meet all of the NQF criteria for endorsement, and we raised a couple of issues here.

One was the definition of renal

failure. Another issue was clarification of the time window in the language of the measure itself not within the supporting documents. Another issue that was raised was the cost of data abstraction.

Clearly, if data is to be abstracted, somebody bears the cost. Then request for clearer specification of the numerator and denominator in the language of the measure itself, again not in supporting documents. Let's go ahead, and if there's any more discussion, please feel free to go ahead and bring it up.

DR. ROGERS: It's still not clear to me what it takes for the NQF to endorse it, with respect to complete or partial. I realize this is a yes/no, but after we're done all our work today and tomorrow, what happens next? I mean what does it take for the stamp to actually be put on? Perhaps others understand it; I don't.

DR. BURSTIN: This is still fairly

early in the process. You'll have a chance to have the responses back from the developers. You'll then ultimately compare to other measures.

Potentially, if there are competing measures in the portfolio, for example, you will then make a recommendation that will go forward for public comment, public and member comment. We get lots of those.

You'll have a chance to wade through those, see if any of those public comments sway your opinions, and it will then ultimately go out for member vote and to our Consensus Standards Approval Committee, which is a board-level committee that reviews it and the Board, and then an appeals process. So you're still fairly early overall.

But your question is a really important one, in terms of what's the threshold of passing, and I think that, you know, this is -- if it was a simple

mathematical formula, we wouldn't need all you guys to sit here since we have your votes. It's not.

Other than having a must pass criteria of Importance to Measure and Report, and a hierarchy for Scientific Acceptability as being the next one, this really does get into your expert opinion as to whether or not at the end of the day, seeing how you voted, do you think the measure is appropriate to move forward.

DR. ROGERS: So potentially then a no vote at this point in time can easily be changed to a yes vote, if some of the questions that are brought up in this conversations are satisfactorily answered?

DR. BURSTIN: Yes.

DR. ROGERS: Okay.

DR. KLEINPELL: And then conversely, Helen, if we do have a yes vote, will all of the comments that Arden identified, will they be addressed per our

request, or do we really have to have a no for those to really go forward?

DR. BURSTIN: I think we would still send them to the developers for their response. It doesn't necessarily mean that you would say, for example, going back to the point that was raised earlier, what you wouldn't necessarily say is that you're voting it with conditions "I will only take this measure if you make the following change."

Which I think it would be very reasonable to pass the comments that Renae made about, for example, the RIFLE criteria aren't here and get their response, and then you would weigh that in your final decision.

CHAIR MORRIS: Okay. So let's go ahead and vote on this final one. On the question of whether the measure meets all criteria for endorsement, we're again very close. 12 voted yes, 10 voted no.

[COMMITTEE VOTING.]

CHAIR MORRIS: So the majority

voted yes, but we have -- obviously we have some issues that a substantial proportion of the group need to have answers on. Okay. Can we move on to the next measure? What time is it? Oh, it's 11:15.

Do you want to take a short break? Okay. Let's take a short break. I think we're scheduled for 15 minutes from 10:00 to 10:15. So our break is over now. Just kidding. Let's just take a ten minute break, if you would, and come back to the room in ten minutes.

(Whereupon, the above-entitled matter went off the record at 10:20 a.m. and resumed at 10:42 a.m.)

CHAIR MORRIS: Dr. John Martin was going to go ahead and describe the next measure, Surgical Reexploration. That's 0115.

DR. MORTON: Thank you. This is Measure 0115, and it refers to risk-adjusted surgical reexploration, and the steward of the

measure is STS.

The description is percent of patients aged 18 years or old undergoing isolated CABG who require a return to the OR for bleeding, with or without tamponade, graft occlusion, valve dysfunction or other cardiac reason.

Numerator is, as mentioned before, those number of patients who had the descriptor event, and the denominator is all patients undergoing isolated CABG.

So in terms of looking at the criterion, this is clearly in the public domain and interest. It's actually one of the PQRI measures from CMS as of 2009. It's Measure 168. I'm going to resist the temptation of saying that this is an easy measure to endorse, given our previous discussion.

But this is actually an interesting measure, in the sense that the database doesn't have to be STS. Many, many

hospitals around the country maintain some sort of reop data collection through their OR systems like Midas and things like that. So theoretically, that data can be obtained through a variety of measures.

In terms of validity, it's got very strong correlation to complications, certainly with bleeding. If there is a reexploration, there's an increased risk of other downstream complications such as mediastenitis, which is a hospital-acquired condition.

There's a definite correlation to cost. There are some implications around transitions from the pump team to the ICU, and in a way, this was presented as kind of a mini-readmission, if you will, within the episode of care.

With all that being said about its validity, there were some questions that came up in looking at the measure as written. The first question is for a return to the OR, why

is it only for cardiac reasons? In looking at the literature, about 80 percent of the reasons for a return to the OR because of bleeding or graft occlusion.

But there are circumstances for things like infection or a retained foreign body. The other issue that came up is why is there risk adjustment at all. In some ways, does this obscure opportunities for quality improvement? If it's risk-adjusted, we don't find out exactly which specific conditions or procedures will lead to this return to the OR.

Just like the previous measure about timing, this was left open-ended. Some of the concern about the timing definition is mitigated by the fact that it's strictly about cardiac. If we do open it up for other reasons, then perhaps that timing issue should be better-defined.

One kind of procedural issue is there appears to be a little bit of a conflict between the denominator statement saying all

patients undergoing isolated CABG, which interprets to mean strictly CABG, no valve. But if you look at the descriptor, it says "valve dysfunction."

So some way or another, that needs to be resolved. Those were essentially the main questions about the measure itself.

CHAIR MORRIS: Thank you. Was there anything else from the group or anybody else want to bring up any issues or concerns, questions about the measure?

DR. HALPERN: I would say I actually agree with just looking at cardiac complications, since that's why you're doing the surgery, and I think the valvia dysfunction probably relates to possibly clotting up one of your coronary vessels, which can lead to acute valve dysfunction.

CHAIR MORRIS: Any other issues or concerns? Dr. Martin.

DR. MORTON: I guess, you know, the only thing about the risk adjustment,

again, the risk adjustment is made through the case mix adjustment model that's in the attachment from STS. Again, it's open, so anybody can access it.

Just more of a philosophical point about, you know, do we really need to continue to do the risk adjustment. If you dig into it, as to reasons why the risk factors for reexploration, they tend to be about non-modifiable risk factors, age and things like that.

So risk adjustment looks like it's appropriate to maintain, but just out of philosophy it would be nice if we moved beyond risk adjustment and just looked at the measure itself.

CHAIR MORRIS: I have a question for the STS. Did you all discuss this amongst yourselves, and think about changing risk adjustment, or having arguments for keeping versus changing it?

DR. PRAGER: I don't know if

David's on the line. David, are you on?

DR. SHAHIAN: Yes I am, and I would respectfully disagree with the issue on risk adjustment. We know from national data that there is very substantial variability across institutions in the prevalence of high risk characteristics.

For example, a small community hospital that does 150 cases a year has a much different patient population, or hopefully has, than the Cleveland Clinic. That's cases from all over the country that nobody else wants to do.

To account for the differential risk in the patients between those two institutions, that really imposes an unfair burden on the institutions that are taking those sort of last resort sorts of cases. So we believe strongly that risk adjustment is essential.

In terms of the question on valve dysfunction, you're right. Acute coronary

occlusion can lead to that, or you can have a patient where you have a two plus microregurgitation. You're not sure whether you want to do something or not, and you end up not doing the mitral valve repair.

Then the patient develops heart failure secondary to what in fact has now become three or four plus microregurgitation. So that's the valve dysfunction.

And you know, we -- this was intended to be a cardiac reoperation measure. There are other reasons that patients come back to the OR, but this is specifically designed as a cardiac reop measure.

DR. MORTON: I guess the only question I had about cardiac reop, right now it's essentially for bleeding, graft occlusion, valve dysfunction. There can be kind of a cardiac-related reason in terms of, you know, mediastenitis and things like that, where you reopen.

Or a routine foreign body was the

other one. Even though it's rare, it looks like the vast majority of returns to the OR fulfill the criteria. But those were other considerations.

DR. SHAHIAN: To capture the reexploration for mediastenitis, we capture that separately, and I think that's another measure that will be coming up at some point, captured separately.

CHAIR MORRIS: And I think the other -- I'm sorry, go ahead.

DR. MORTON: No. I was just going to add what David said, as we do for retained foreign bodies and things like that, a separate capture.

DR. DUTTON: I wanted to raise the question about this measure, and it might apply to a couple of the others. Are we creating an unintended consequence here? We would like to think that the decision to reoperate on a patient would be cut and dried and absolute, and applied the same way by

every surgeon every day.

But I've certainly stood around the ICU looking at people bleed, and trying to decide do we transfuse them or do we give them Factor 7, do we go back into the OR? While we have the subject matter experts here on and on the phone, how do you do that in gathering the data? How do you prevent gaming it by, for something that's publicly reported like this?

So it has some --there's some paranoia around it. Can the surgeon make his rate lower by choosing not to reoperate on people, and you know, how do audit for that? How do you see, make sure that doesn't happen?

DR. SEARS: Can I ask one other question as well, sort of a corollary to that. We don't have any exclusions here, and have you taken into account some of the newer therapies that are being used by cardiology, the anti-platelets in particular?

Are they part of the risk adjustment, or do we need to be considering

those, because those certainly increase the level of bleeding post-operatively?

CHAIR MORRIS: Dr. Prager.

DR. PRAGER: As far as the gaming, obviously certainly in our field, there are criteria to be considered when someone should be reexplored, and at training institutions they become a mantra for the residents as well as the faculty. If you're asking me could this be gamed by having a surgeon give platelets, plasma, wait it out, and the answer is if that's gaming, the answer's yes.

There are other surgeons, frankly because we see this when we look at rates, whose threshold for going back to the operating room is a very different threshold. They'll go back earlier, because they don't want to give blood, they don't want to give plasma.

So I think on balance, we actually balance out. But can we measure the gaming? The answer to that's no. Yes, we know if

patients are going to the operating room on active anti-platelet therapy, and can you tease that out eventually? Yes. If someone did a study, we could tease it out eventually.

DR. SEARS: So for right now, we don't list those as exclusionary?

DR. PRAGER: No, they are not exclusionary.

CHAIR MORRIS: Are they included in the model?

DR. PRAGER: I don't believe -- David, are they included in the risk model? I don't believe so.

DR. SHAHIAN: There were -- that did not fall out as a, actually as a significant predicter. The question, though, about the gaming, I think, would apply to every single measure that I'm aware of, not just in cardiac surgery but, you know, every measure that's out there, I think, is potentially gameable, potentially can have adverse influence on physician behavior.

I think at some level, we have to trust the ethics and the good medical practice of our colleagues. We cannot police gaming, nor can we eliminate any adverse consequences that might result from attempts to make one's score better.

This is true in mortality reporting, where risk aversion is such a concern. It's very hard to police that.

CHAIR MORRIS: Okay. I think it's a good issue to bring up and air, just because it always comes up in private conversations, and better to address it explicitly. I agree with what you said, Dr. Shahian. So in terms of exclusions, sounds like really not, no exclusions here.

But that you already examined for what would fall out as predicters, in order to adjust for those in the risk model. Then the last question I think Dr. Morton brought up was the timing of collecting this data. Should we assume that this is during the index

hospitalization, or is the timing otherwise specified?

DR. PRAGER: It's during the whole hospitalization.

CHAIR MORRIS: Not a 30-day window or --

DR. PRAGER: Not for this.

CHAIR MORRIS: Any other discussion on that before we move to a vote on the criteria?

(No response.)

CHAIR MORRIS: All right. Well, let's go ahead. So the first criteria, Importance to Measure and Report. I'd like everybody to go ahead and cast your vote.

[COMMITTEE VOTING.]

CHAIR MORRIS: And the result is that 22 out of 22 said yes, this is important. The second criteria, Scientific Acceptability of Measure Properties.

[COMMITTEE VOTING.]

CHAIR MORRIS: We have 19 that

said completely meets criteria for scientific acceptability, and 3 said it partially meets criteria. Next, Usability.

[COMMITTEE VOTING.]

CHAIR MORRIS: 20 of us said that it completely meets Usability criteria, and 2 of us said that it partially meets Usability criteria. Next is the Feasibility criteria.

[COMMITTEE VOTING.]

CHAIR MORRIS: 21 of us said that it completely meets the Feasibility criteria, and 1 of us said that it partially meets the Feasibility criteria. Then the last vote is does the measure all of the NQF criteria for endorsement?

Is there any other comment or any issues that anybody wants to bring up before we go to this one? Dr. Morton, your light is on, but I'm not sure if that means you want to speak.

DR. MORTON: No. I'm sorry about the light.

CHAIR MORRIS: All right. Let's go ahead and vote then.

[COMMITTEE VOTING.]

CHAIR MORRIS: We have 22 out of 22 say that yes, it does meet the criteria for endorsement. I think this is going to be our fastest measure. The first one is our slowest, and this is our fastest.

PARTICIPANT: You were right, John. Maybe you were right.

DR. WILHOIT: And the one piece of feedback that John had raised was the time frame, and I think it's worth feedback that that should be incorporated into the document.

CHAIR MORRIS: I agree. Okay. So the next measure is 0116, Anti-Platelet Medication at Discharge, and this will be introduced by Dr. Stafford.

(Off mic comment.)

CHAIR MORRIS: Oh, I'm sorry. You're right. I'm sorry about that. Yes. We're at 0129, Prolonged Intubation. Dr.

Stafford.

DR. STAFFORD: So on page nine of the handout they gave us today at the bottom. So this is Measure 0129, Risk-Adjusted Prolonged Intubation (Ventilation). The topic is the percent of patients aged 18 years and older undergoing isolated coronary artery bypass grafting, who require intubation for more than 24 hours.

Overall, the group felt that this was important. There really weren't any large issues when it came to Importance, and in fact we all felt this was actually very important to measure.

When it came to Scientific Acceptability, there were several issues, one of which has been addressed, and that has to do with the time window for the numerator, and we've heard that it's now within the hospitalization. So that's really no issue.

Dr. Dutton actually brought up a really good point about some potential

confounders, and the largest one being how do you capture, what do you do with the patient who is clinically ready to extubate by all the criteria that we all use in the ICU every day, but is kept intubated longer than that 24 hours for some logistic reason, needing some other procedure, say needing to travel for a CAT scan because they're not waking up post-CABG and they need to go for a head CT?

So how do you deal with that? It's probably a small effect, but it's something that I think is worth discussing. That was the major issue with that. When it came to Usability, again, Dr. Wilhoit had some issues with the denominator details, more discussion about confounders, and then finally in terms of Feasibility, there really weren't any large issues with the measure.

I think one of the larger questions, and it similarly related to the renal failure question and why did STS choose 24 hours as a gold standard, as opposed to

maybe 48 hours. That's more just a discussion. When we talk about ventilator-associated pneumonias, if you use the CDC criteria it would be a 48-hour window.

One of the problems with leaving patients intubated longer is the development of EAP. So that was a question that we had in terms of the definition.

CHAIR MORRIS: Any other questions or issues?

DR. ROGERS: With respect to the duration of ventilation, I would actually wonder about a shorter time period than a longer time period.

I think the vast majority of patients actually are off of ventilator sooner, and I just wonder if this is sensitive enough to be useful, to lead a discussion about how do you actually get people off ventilators, or extubated a lot quicker.

DR. STAFFORD: I think that's actually a really good point, and I think --

but I think that's why this measure is important, because you stated the vast majority are, and at least in the number of places I've seen, I think there's actually a large gap here in that.

You're right. Something less than 24 hours might be appropriate. But I think at least it gets to the importance of this measure in picking some time frame.

DR. ROGERS: I think in the state of Washington, I think we're at about six hours average, and I -- so, from a discriminatory standpoint, I understand your point. But does it help us really guide the fine-tuning of, I mean any minute with a tube left in is an unhappy minute, as you know.

DR. STAFFORD: I would absolutely agree with that, and I think you're coming from a background where you're lucky to be in some place where you have a robust database and a group that catches all of their data. Some states do, some states don't. I think

that's where participation in the database is helpful.

Perhaps, perhaps somebody in the group could come up with a proposal measure that's actually shorter. I think that would be a valuable discussion point to have.

CHAIR MORRIS: Let me ask the STS a question about this, the sensitivity of that measure. Can you speak to whether it is adequately discriminative, or how discriminative it is, how sensitive it is?

DR. PRAGER: Yes. I think I need Jane on the phone, but it is reported out, and you can ask that it be reported out in less than 24 hour time frames, and I'm not sure it is in everyone's report. so I think the discussion is --

I think the discussion is a very good discussion about this, and I would guess that David will answer that 24 hours was picked years ago, unrelated to -- before the definition of VAP and things like that, and it

has been a great target, if you will, for quality initiatives and improvement.

Although we can be more sensitive, and certain states do pull their data in six-hour time frames purposely to look at this, and yes, I think it's fair to say the majority of coronary bypass patients in most states that have quality initiatives are extubated within 12 hours.

CHAIR MORRIS: As a great target, can you tell us what percentage of people are extubated within 24 hours, if it's -- are we topped out --

DR. PRAGER: Yes. I can't tell you STS, unless Jane can look it up right now. I don't have that number. Jane, do you?

DR. HAN: I'm scrolling through looking for it right now, but not at this moment.

CHAIR MORRIS: How about this? Instead of giving me the exact number, can you tell me is this measure actually topped out at

the 24 hour mark? Are we topped out?

DR. SHAHIAN: This is Dave. It's not topped out, and there in fact is still a fairly substantial distribution, and in fact, I believe in the past, this measure was at 48 hours, and we actually have come down to 24.

It's a compromise. I think if we were -- you know, ideally, I think six hours to 12 hours would be ideal. But I think you have to reach a compromise between what's acceptable practice and what is the absolute optimal practice, and that's why we settled on 24 hours.

We have the data on what percentage of patients are extubated by 24 hours. If Jane doesn't have it right now, we can get that for you very easily. But it is a very substantial proportion.

DR. WILHOIT: What's in the article that was attached to the submission was 9.7 percent. I think, I assume it did not comply with the 24 hours. But what's in

the table is 9.7.

DR. SHAHIAN: No, I think that's correct.

DR. DUTTON: On the sensitivity and specificity, if you look at the 2008 performance compared to the 2009 performance, it's on page 17 in the measure submission. It's very linear, with a very high row value. So this is a system issue, in other words, places that are bad one year tend to be bad the next year as well or good. So I think this is a very discriminatory measure.

DR. STAFFORD: The only other comment I'd say, and I'm going to play the devil's advocate. If we, and then we'll get to Dr. Dutton's questions earlier about unintended consequences. If we set a time frame that's very short, then we ought to be measuring reintubation, and other morbidity associated with the measure.

So I think picking some reasonable time frame is appropriate, and I think we have

to be careful not to go too far on either end of the spectrum.

DR. SHAHIAN: And we do in fact measure reintubation. That is one of our measures.

CHAIR MORRIS: And would you please also speak to Dr. Stafford's question about adjustment, risk adjustment or potential exclusions?

DR. SHAHIAN: Yes. Once again, I think prolonged ventilation is the common final pathway for a lot of different problems. It could be preexisting lung disease; it could be massive fluid overload, you know. It could be a complication like a stroke that develops. It could be, you know, some other complication.

Rather than, you know, rather than try to sort out all those various things, we've simply accepted the fact that it is kind of a good general metric for a patient that has had a problem, that might not necessarily

be captured by other factors. Now the issue of a systems problems, you keep the patient intubated because you need to move them somewhere to get a test.

I guess I would say that's a systems problem within a hospital, and it's one of the things they ought to be able to deal with.

DR. CIMA: So if that's the case, do we even really need the risk adjustment model? I mean what role does it play in this?

DR. SHAHIAN: Well, I think the role -- I think if a patient, for example, has severe chronic lung disease preoperatively, meaning by our definition they're on home mode 2, that's going to be a very significant predicter for prolonged post-operative ventilation. I think that sort of thing needs to be adjusted for as it is in our model.

CHAIR MORRIS: Does that respond, satisfy your question? Okay. Any other issues or any other comments anyone wants to

raise?

(No response.)

CHAIR MORRIS: Let's go ahead and move onto the vote then. That first vote is on the Importance to Measure and Report this item.

[COMMITTEE VOTING.]

CHAIR MORRIS: Let me ask for everybody to put in their responses and hit the send one more time.

[COMMITTEE VOTING.]

CHAIR MORRIS: And 22 out of 22 agree yes, this is important. Next, Scientific Acceptability of the Measure Properties.

[COMMITTEE VOTING.]

CHAIR MORRIS: 17 said that the measurement properties are completely acceptable, and then 5 said partially. Next is Usability.

[COMMITTEE VOTING.]

CHAIR MORRIS: And 20 of us said

it completely meets the criteria for Usability, 2 said it partially meets the criteria for Usability. Next is Feasibility.

[COMMITTEE VOTING.]

CHAIR MORRIS: 20 of us said it that it completely meets the criteria for Feasibility. One of us said that it partially meets the criteria and one said it minimally meets the criteria for Feasibility. Then we're back at the yes/no, and that is does the measure meet all the NQF criteria for endorsement.

So the issues that I heard raised were questions about the duration of ventilation. There was a reasonable discussion of that, and some acknowledgment that it's been a bit of a moving target over time, but that it sounds like 24 hours is pretty reasonable right now, and that close to ten percent are not extubated within 24 hours, 9.7 percent or something, among those measured while this measure has been place. Please

correct me if I'm wrong about that.

Then there was a question about adjustment for potential confounders, and it sounds like Dr. Stafford was satisfied with that discussion. Dr. Cima raised the issue of, you know, why are we adjusting at all.

DR. CIMA: I just wanted to clarify that. It's like your point about topped out on this. If there's risk adjustment for seven or eight or nine percent of patients, are they being kept intubated because they meet all these other criteria. So are we tapped out? I mean that's the question.

If you're going to risk adjust it, then do we know how many of these people are chronic lung disease, how many people had a stroke in that 24-hour period? If that's the criteria, then do we really know if we have much more room to move. That was what I'm talking about.

CHAIR MORRIS: Okay. Thank you

for bringing that up. It sounds like the STS has told us that there is a -- although it could appear to be close to tapped out, there's still a wide distribution. So although more than 90 percent of people are extubated at the 24 hour point --

(Off mic comment.)

CHAIR MORRIS: How can the distribution -- that's just kind of where places fall out on the, you know, on the curve. So room for improvement among some centers.

DR. SHAHIAN: That's the issue. It is very hospital-specific, and there are some hospitals that are doing much more poorly than that, and then there are some superstar hospitals that are getting everybody extubated within a few hours.

CHAIR MORRIS: Is everybody comfortable with going ahead and voting on the last item, do we meet NQF criteria for endorsement?

(No response.)

CHAIR MORRIS: Let's go ahead and vote.

[COMMITTEE VOTING.]

CHAIR MORRIS: And 21 of 22 said yes, we should go ahead and endorse this. One said no. We're really picking up speed here. So of our next measure, 0131, Strokes, Cerebrovascular Accident is Dr. Dutton.

DR. DUTTON: Yes. Sorry. I'm trying to get the static out of it. This Measure 131, risk-adjusted stroke after isolated CABG surgery. The numerator is defined as a new neurologic deficit persisting for at least 24 hours. The deficit has to be associated with a structural abnormality in the brain.

The denominator is all patients greater than 18 having an isolated CABG operation. There's an exclusion for prior CVA. It's not specified how the exclusion is calculated. I believe this measure is

important. It has a great deal of face validity.

Obviously, if you're a consumer, this is an outcome that matters to you very greatly. So I think this is an important measure. I think there are some harder issues around how it is defined and how it is measured.

I note, and I don't know if Alexis was able to get my picture up, but if you look at page 18 in the submission, you have that same scattergram of 2008 to 2009 data. You discover that the reproducibility of this from reporting centers year on year is actually fairly low. It's only, the row is .26 on this.

In other words, a center can be good one year and bad the next, which I think is a consequence of this being a very low number event. It doesn't happen very often. These are anecdotes rather than something that it's easy to make a rate out of. So using

this on an individual hospital basis for performance improvement would be hard.

On the other hand, I think it has a great deal of credibility as a national measure and aggregated nationally, and I know that the SDS has reported a decline in this rate over time, and this is very useful. So I think it's a very supportable measure.

The other thing I'll say about definitions, again some potential for gamesmanship in this, because there is an exclusion for prior CVA. You need to understand how that exclusion is made.

If it's by ICD-9 codes, for example, from the hospital record, versus specific testing, versus if you want to look at the most expensive alternative, do a CT scan every patient before surgery, to rule out events and establish a baseline for subsequent CTs.

Obviously, not feasible in that respect, but that's the slippery slope that

you're on with this kind of measure. I think you just have to pick what is the most rational definition for cost and a reasonable criteria and go with that, and I think STS has done that.

Then the same question has come up with renal patients, is if a patient does have a prior deficit, shouldn't we be interested in the ones that are getting worse, and why not measure everybody, include everybody, have no exclusion for a prior event. But instead, have a measure based on worsening of a neurologic state. That's all I have to say.

CHAIR MORRIS: Any other issues anyone wants to raise? Questions?

(No response.)

CHAIR MORRIS: So it sounds -- let me just recap that for you. Thank you for that discussion. First of all, it sounds like you had a question is how is this exclusion of a prior CVA calculated.

Secondly, you're concerned about

the level of analysis, whether it's at individual hospitals, which shows a lot of variability from year to year, versus in the aggregate, which may be more reflective of a national trend.

And then thirdly, related back to the first one, should we be including all patients, so that we can look at folks who have had a prior stroke and whether or not their deficits are worsening with a CABG, after a CABG. Does that capture it? Okay. Does the STS want to respond to these?

DR. PRAGER: Yes I can start, and then David and Jane, feel free to chime in. I think on a national basis, it has proven to be a very good target to look at, at the individual hospital.

In other words, the individual hospital that does 150 isolated coronaries and has no strokes one year looks great, and the next year they could have five, because they were older, had a prior history of stroke or

whatever the issue was, had an atherosclerotic aorta based on aging.

So I think that point is well-taken, and we would agree with it. We have found that it is very useful on the local level, and prompted by both national levels and comparing yourself to that.

The issue of prior stroke being ruled out. I think historically, we have done that. I think that also is a fair issue. Do we want a bigger number to look at? I think when this was initially selected, it may have been based on the assumption that people with prior stroke, which to my understanding is simply by history, not by CT scans, not because the relative said that grandma had the stroke. It is someone took a history. Frankly, that's how we do it.

Having said that, we haven't broadened it, and I'm not sure we have really entertained that discussion in any robust fashion at this time. David, any other

thoughts or Jane?

DR. SHAHIAN: Yes. First of all, in terms of the sample sizes, you're exactly right. This is a fairly rare event. But if you think about it, the incidence of stroke after coronary bypass surgery is roughly in the same ballpark as the incidence of mortality, you know. It's a percent to a percent and a half, in that ballpark, in most recent studies.

So we're dealing with the same issue that we deal with in assessing mortality as a --, and there are ways to mitigate these problems to some extent using the statistical techniques that we do, and one can also look at data year over year.

I think a program that's bouncing around from zero to four to one to two says one thing. I think a program that is consistently 0.3 percent year after year tells you something else, and a program that's consistently four percent year after year

tells you still another story.

So I think there is information to be gained from looking at trended data, and of course we have that capability. In terms of this issue of previous stroke, Rich is right, and that we simply accept that a history has been taken, that there was a previous stroke. Now why is this important?

In cardiac surgery, most of us have seen this syndrome where a patient has a previous stroke, seemingly recovers from it or does recover from it functionally, undergo cardiopulmonary bypass and then perhaps because of transient cerebral edema in the area of the previous scar, or interruption of some new collateral pathway that developed, patients that have completely cleared their previous neurologic event may develop some of those same symptoms that they had with their previous strokes, even without evidence of new stroke.

So that becomes very difficult to

sort out. So that's why we've tried to exclude that particular group of patients, because it does get a little bit difficult to sort out what's new, what's old. So just try to make it a cleaner measure.

DR. HALPERN: I have a question about how the model works, because I noticed in your risk-adjusted model, part of that risk adjustment was a previous history of stroke. So is that somehow like double-counting it? Does it get double-counted that way?

DR. SHAHIAN: Well, those patients would not be included in the measure, in this particular measure, right? I think they would be excluded. It is, you know, let's see. I have to go back and look at that in terms of how that's calculated.

DR. DUTTON: Prior stroke is prior stroke?

DR. SHAHIAN: That would not -- those patients, that actually would not end up being -- although our general risk adjustment

model for stroke includes that, that particular group of patients -- measure.

DR. DUTTON: Yes. It's not part of this measure, but prior stroke is part of the mortality risk adjustment, which is the sort or generic risk adjustment they're showing us for all of these measures.

DR. HALPERN: So is that a typo then in this particular measure. Dr. Halpern. It's down in their risk adjustment methodology. It's included.

CHAIR MORRIS: Okay. So potentially a typo on the part of the measure developer. Dr. Afsar-Manesh, do you have anything to add to this discussion in terms of your clinical expertise?

DR. MORTON: I had a question about at what point are some of the preop characteristics obtained in data collection? Is this all done after the episode of care is done? Do you get some of these data before the surgery or more contemporaneous with the

surgery?

For the issue about stroke, is it correlated to the prior to admission ICD-9 codes that might be present?

DR. SHAHIAN: We recommend that data are collected contemporaneously within the usual processes of care. So these data would be collected by admission or a PA preoperatively, and then recorded in the database by the data manager. That's our preferred methodology.

The data are collected preoperatively; it's just a question of whether the data manager abstracts them down the line, or whether they're doing it contemporaneously.

We, you know, hospitals would have the option of using, of going back to previous ICD-9 codes for a history of stroke, or going to a letter from a referring doctor. There are many ways. We haven't been proscriptive about how that stroke, previous stroke is

defined.

CHAIR MORRIS: Let me ask you another question about that. We're getting some static, I think also. Is that on his line? Okay. It just stopped, which is nice. We'll see if it comes back.

MS. MURPHY: But if you're not speaking, please, if you're on the phone, mute your line when you're not speaking. Don't put it on hold, mute it. Thanks.

CHAIR MORRIS: Okay. So my question is does this require that the data abstractor then go through letters from referring physicians or go through the physician's written notes in the chart, in order to identify a previous -- and I think that's really what Dr. Morton was getting at, a previous episode of stroke.

If it's not reliant on ICD-9 codes, does this require that the abstractor spend what could be a lot of time and potentially not available in the hospital

chart?

DR. SHAHIAN: Well -- go ahead, Rich.

DR. PRAGER: Okay. The most recent to the time frame of the operation history and physical, certainly for coronary artery bypass patients and cardiac patients frankly should include whether there is a history. So it does not require going back. We don't use ICD-9 codes, and it should be frankly the most recent history that was done.

CHAIR MORRIS: Does that answer your question?

DR. MORTON: Yes, I guess so. Is it part of the -- I heard there was going to be validation of the centers collecting the data. Is that going to be part of the validation, to see how this data's collected and how contemporaneous the data collection is? Is that a plan for the STS?

DR. PRAGER: Yes. The STS is actually doing that now, and this is one of

the variables of the 73 that are audited, and as part of the Iowa Foundation for Medical Care, which is our auditing agency, they actually then tabulated were the data put together prospectively by the site, which is obviously what we recommend, or is it retrospectively or how is it done?

Is it electronically, contemporaneously? So we have some of those data starting to emerge, but this is an audited variable.

CHAIR MORRIS: Okay, and one further question. Does this apply only to the index hospitalization, or is it over a 30-day period?

DR. PRAGER: Index, index.

DR. STAFFORD: Arden, I have one more question, and this gets to the development of stroke in particular, but may affect some of the other end points. Has STS considered looking at, or do they discriminate between on pump cases and off pump cases?

DR. PRAGER: Yes, we do.

DR. STAFFORD: I didn't see that in your risk model. Have you considered doing that?

DR. PRAGER: Yes. David, help us with that a little bit, because I'm not quite sure that's in the risk model, but we now have data points for it.

DR. SHAHIAN: No, you wouldn't want to put that in the risk model, because then you might be, it's kind of like socioeconomic status in a way.

That's one -- the type of procedure performed within isolated CABG is something that you might want to do a stratified analysis for, but you might conceivably adjust away the salutary effect of off-pump CABG, if you believe that off-pump CABG is less likely to produce a stroke.

You might adjust that away if you include it in the model. So I think stratification would be the better approach.

CHAIR MORRIS: Any other issues before we move on to the vote? Anybody want to bring any other questions up?

(No response.)

CHAIR MORRIS: Let's go ahead and move on to the vote. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: 22 out of 22 said yes. Next, does the measure meet NQF criteria for Scientific Acceptability of the Measure Properties?

[COMMITTEE VOTING.]

CHAIR MORRIS: 12 said completely, 10 said partially. Next, does the measure meet NQF criteria for Usability?

[COMMITTEE VOTING.]

CHAIR MORRIS: I'll just ask everybody to hit their response one more time and hit send.

[COMMITTEE VOTING.]

CHAIR MORRIS: Okay, and the vote was that 17 out of 22 said yes, it completely meets the criteria, and 5 said partially. And then does the measure meet NQF criteria for Feasibility?

[COMMITTEE VOTING.]

CHAIR MORRIS: 18 said completely, 4 said partially. Then lastly, does the measure meet all the NQF criteria for endorsement, and I just want to recap some of the concerns that were voiced.

There was a question about how the exclusion is calculated, whether this is from -- well, it could be calculated in a number of ways. But it sounds like Dr. Prager said that primarily it's determined based on most proximal history and physical done prior to the operation. So not by prior ICD-9 codes.

Secondly, there was an issue about whether or not prior strokes that had worsened as a result of the CABG operation would be captured, and the answer is really no, because

there's so much noise from prior strokes and potentially cerebral scar tissue or residual effects of the prior stroke.

Then there was a concern about the level of analysis, whether this truly represents the experience at individual hospitals, or whether aggregate data are more useful. For low yield events, for fortunately low yield events like this, mortality and other events, there probably will be a lot of potential noise in there for some hospitals or institutions.

I believe it was Dr. Shahian, but it might have been Dr. Prager brought up that this might be most useful, dealing with that particular question, by measuring trends, which is not part of this measure, as I understand it, but could potentially be explored in the future in individual hospitals.

Then there are some issues in the language of the measure that sounded almost as

though potentially editing of the language of the measure would have been appropriate by the STS. For example, including in the model that it adjusts for prior stroke when those patients would have been excluded.

Then lastly, there was a decision, a purposeful decision by the STS, as I understand it, not to adjust for on or off pump, in order to be able to compare those operations more appropriately. Does this -- anybody have any issues besides what I've brought up to address? Any other questions?

DR. WILHOIT: Specification of time frame.

CHAIR MORRIS: Yes, thank you. The specification of time frame -- I'm sorry. I don't recall the answer to this. It was during the index hospitalization, you said? Okay. So stroke during the index hospitalization and not after. All right.

DR. SHAHIAN: Could I just clarify one issue that was raised, the issue of the

inclusion in the risk model of patients who had a previous stroke, whereas they're excluded from the measure. They're really, they're related but separate issues.

We devised risk models, generic risk models, to be used for many different purposes, not just for the purpose of this measure. So that generic risk model includes patients who have had a previous stroke. For the purposes of this measure, of course, there are no such patients.

So all the patients in the particular cohort being evaluated for this measure would have previous stroke, no. So that would not have any impact on the risk adjustment model. So it's not a typographical error; it's just two slightly different things.

DR. WILHOIT: Can I raise a question? In the document, it does not list prior stroke as being an exclusion, I don't think. Am I looking at the wrong one? Oh,

I'm sorry. I am looking at the wrong one. I'm sorry.

CHAIR MORRIS: Okay.

DR. MORTON: I did have one question about the risk adjustment issue, about off and on pump. I guess I'm still not clear, because one advantage of the off pump is to prevent stroke, is that right? So I'm wondering why I understand the stratification, but I don't know. I guess I'd like a little more clarification around that.

DR. SHAHIAN: Sure. Risk models used for profiling should include characteristics of the patient that are present when they first encounter the provider. Discretionary decisions, such as on pump, off pump, repair, replacement, whatever, all those sorts of things are, I think, not appropriate for use in risk models designed for profiling purposes.

They might be useful if they were being used specifically for patient

counseling, for example. If you wanted to devise a model that, to enable you to discuss with a patient what's your likelihood of stroke after this procedure, and, oh by the way, I've planned to do it using the off-pump technique, that's fair game.

But in terms of profiling models, data collection or patient characteristics are collected up to the time you encounter the provider, but not beyond that. At least that's my understanding. That's the way we've done it.

DR. CIMA: Just one question about this. If this is not anything modifiable, according to the data that we know of, and not really a system issue that we know of, not like whether or not you extubate patients early, this is more like just a rate.

Can you drive quality improvement if there's not a modifiable system or risk factor issue here? I mean that's one point, and the second one I'm just going to make

again the plea, to understand, you know, the usability of this model.

The model as -- the more and more complex, the modeling is going to be important, and institutions that don't participate are going to be excluded, and we're going to have to come up with another system.

So again, those are the two points. But in this one in and of itself, there's no quality improvement aspect that I can actually see.

DR. DUTTON: I'll tackle that, because I started this with a discussion about the rate and the impact of a very low rate on using this. It is certainly a system-modifiable event. How you manage a fib, how you, you know, where you put your holes in the aorta, you know, are you using ultrasound?

I mean there's a lot of variables that contribute to this that are well understood, how you manage perfusion on pump.

There are a lot of variables. The problem is that the end result, the stroke, occurs at such a low rate that it's hard to measure in an individual institution.

But certainly if you came up three years in a row with a high odds ratio in this, you would have a very good incentive to go look at your practice and figure out what you could do to make it better.

DR. STAFFORD: Yes. I think the measure becomes a trigger tool. So you see a change in practice over time or in outcomes over time, and that should trigger the institution to look for the reason that that occurred. Sometimes in many places, these are sentinel events.

So depending on how your institution looks at outcomes, that's -- but that's the reason to have that.

DR. SHAHIAN: I completely agree with all those comments.

DR. PRAGER: And I would just add

that it is a modifiable event in the institution that has a four or five percent. If you look at maybe the use intimate and cross-clamp, one of the techniques.

Maybe they run their perfusion pressures at 40 or 45. So we have learned that you can actually modify this. Now it's not going to be zero, but you can modify it, and it is a trigger for that.

CHAIR MORRIS: Okay. So we're going to go ahead and vote on, does the measure meet all the NQF criteria for endorsement.

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty-two out of 22 of us said yes, it does meet the NQF criteria for endorsement. We have two more measures here in consideration of candidate measures for the pre-lunch session. But we also have a caterer who's brought lunch.

So I think what we'll do now is we'll break for lunch. It's a half-hour lunch

break, so it's relatively rapid. Be sure you chew your food, and then we'll come back and go over the two measures that are left over, and then proceed.

(Off mic comment.)

CHAIR MORRIS: Right now? Oh sure. Yes, okay. I'm sorry everybody. We're going to take a brief break, I'm sorry. We're going to take a brief -- before we take our lunch break, we're going to take a brief moment for NQF member and public comment. I think this is from folks in the room and then folks on the phone.

NQF Member/Public Comment

(No response)

CHAIR MORRIS: Okay, wonderful. It's lunch time.

(Whereupon, the above-entitled matter went off the record at 12:26 p.m. and resumed at 1:02 p.m.)

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

1:02 p.m.

CHAIR MORRIS: Okay. We're going to go ahead and get started, and we have a couple more measures to finish up from the morning. Fist, we're going to Ms. Graling talk about Measure 0134, Coronary Artery Bypass Graft using Internal Mammary Artery.

DR. GRALING: Right. This was another one of the measures that Work Group A took a look at. It is in process. It's been reported since 2007. We felt that it was important to measure. It certainly is well tied to improved outcomes because of the high patency rates of the IMA.

One question we had is that there was no information that really addressed the disparities, and there is some literature there that discusses certainly women and regional use of the IMA. In terms of Scientific Acceptability, the numerator is the number of patients undergoing CABG with IMA.

The denominator is all patients undergoing isolated CABG, which really leads us to the exclusions, which is those patients who have had a previous CABG are not eligible for the IMA.

We had some discussion related to some of the other exclusions. Those you see listed in terms of subclavian stenosis, Mediastinal radiation, no LAD disease, and I think it was to Dr. Dutton's point about is this someplace where the surgeon can game the system if he chooses not to use the IMA.

Usability is well established as part of the CABG composite scores, and certainly Feasibility has the same issues in relation to acknowledging the fact that you do need a data abstractor. But otherwise, we felt it was quite feasible.

CHAIR MORRIS: Any questions or comments from the group?

(No response.)

CHAIR MORRIS: I have a question

for the STS around disparities, and I'm glad that you brought that up. That's something that we are really trying to discuss with each of these measures that have been present for some time. What's the data on disparities, and new measures as well? What's the information about disparities? We need to attend to that.

We have measures that have been present that are up for maintenance. We need to know a little bit more about what the measure developers have learned since the measures were first endorsed.

So we need to know how things have changed, since this is really for maintenance measures, they're being held to a different standard than when they were initially introduced. But also they've been present for some time, and hopefully have made some changes. We need to know if these quality measures are actually having any impact.

Would the STS like to respond to

the comments?

DR. PRAGER: Yes. I'll start the response. The answer for the IMA is we are having an impact. There is significant regional variation in utilization of the IMA, between -- in various states that look at it regionally, the variation ranges between 65 percent and 100 percent at different sites.

So utilizing these data allow us to create opportunities for quality improvement. I think that's the -- we well know, as the Committee knows, that certainly utilization of the IMA is a gold standard.

The other aspects of the learning are now the exclusion issues that have come after a great deal of debate as well, but a recognition of certain realities that can be encountered. So I think the STS has learned a great deal. I think the clinicians, frankly, have had an impetus to utilize it more frequently when appropriate.

CHAIR MORRIS: Have you seen a

reduction in the variation in IMA use?

DR. PRAGER: Yes.

CHAIR MORRIS: Can you give us any parameters around it?

DR. PRAGER: Parameters meaning I can -- there is a recently published article in the Annals of Thoracic Surgery about three months ago from the state of Michigan, showing increased usage via a quality collaborative approach, to up over the STS average now, from an average of in the 80s.

CHAIR MORRIS: Great. Any other questions or issues that you have? Does anybody else have any questions or issues that they want to raise, to be discussed by the measure developers, or questions for the measure developers?

DR. ROGERS: I just want to be certain I understand the disparity issue. This is not social disparities in any way. We're talking medical disparities with post-radiation, that sort of thing?

DR. GRALING: Well, some of the recent literature actually points at the disparity in the use in women and non-white race groups. So I think that's really important in terms of the learning, in terms of how that's looked at, and then the other big piece was the region.

CHAIR MORRIS: So regional and then socially vulnerable populations, and we're talking about disparities. That means disparities among socially vulnerable populations?

DR. ROGERS: Dr. Prager, do you have any comment about that?

DR. PRAGER: Yes. While regions have perhaps looked at that, the STS has not really asked DCRI or of itself to look at that, but we are more than willing to address it.

DR. DUTTON: I'd just expand for one second on the gaming issue, because it comes up because this is a publicly reported

measure, and the obvious gaming in this is there's an exclusion for IMA not suitable, and that determination is made by the person who's being judged by the measure. So I mean there's an obvious concern there.

I think this is a very important measure for private reporting and for improving quality, and there's no question the STS has moved this indicator to better outcomes, no question. But if it's going to be publicly reported, it raises the concern of, you know, who decides whether the vessel's unsuitable or not?

DR. HALPERN: Are there specific criteria for suitability?

DR. PRAGER: I was waiting for that question. I was waiting for both of those questions, and I think that question really addresses a significant issue, and perhaps this group ought to discuss this briefly, as to whether that exclusion criteria is a fair and reliable exclusion criteria,

because you're absolutely right. That is specific to the person performing the operation.

Now we would like to believe that 100 percent of people performing the operation would make the same decision. Having said that, we also know that that doesn't happen.

MS. WEBER: Are there any criteria out there by which somebody can judge?

DR. PRAGER: There are historical flow criteria for the mammary artery. If you take this down off the chest wall and then people in the 60s and 70s frankly would measure flow into any kind of basin and see how much flow was in a minute, and people would say yes, you needed 50 cc of flow or 80 cc of flow would be wonderful. If you had 20, you probably wouldn't use it.

But we also know in this era of putting Papaverine or something like that around an artery to reduce spasm may change the flow. So even using historical notations,

if you will, not truly evidence-based, as we say today, but historical notations about it, those can, one, also be gamed and, two, we can, with experience, if you will, improve the flow in many of the mammaries that you wonder whether they're questionable.

But the other side of gaming is hypoperfusion with a mammary. I don't want to get overly technical, but if you use the mammary and you put it to a dominant vessel on the heart, and it's not really flowing well, you have set up a potential problem.

DR. DUTTON: Yes, and I assume you capture the negative side of the indicator, which is objective, which is how often does an IMA fail.

DR. PRAGER: Well, you don't capture that unless they're recatheterized. Or they go back to the operating room, and that's a small number. So we won't, we don't know that number.

DR. HALPERN: Are women, somebody

had mentioned about women being underrepresented. Is that because like with many other vessels in women, that they're smaller than in men?

DR. PRAGER: The answer to that is that's not consistent, and it is still, if you will, under disparities, can certainly be -- there have been many papers in the literature, and frankly I think there's an increased utilization in everyone: men and women.

We use it in children. I mean the fact is children who have complex operations not for anomalous coronaries, but other complex operations where the LAD can be injured, redo aortic roots, things like that, Ross procedures, if you have a 12 year-old, we'll take the mammary down on pump and use it in a 12 year-old. So it's not just size-related; it's experience-related.

CHAIR MORRIS: So I just want to make sure that I understand this correctly. So there are, let me just make sure that I get

this. If the IMA is deemed not a suitable conduit, those cases are excluded. But there are no explicit criteria for whether the IMA is a suitable conduit.

There is a historical precedent that was developed the way many historical things are. But there isn't really any explicit, and particularly no evidence-based criteria for a suitability.

DR. PRAGER: Unless we either added to that or said there was no flow in the mammary or the flow was less than X. I don't know if David's on the line. He may have other thoughts.

DR. SHAHIAN: Yes. This is Dave. This is one that we have struggled with. We have not used this before as an exclusion. We want it in our current -- in our new data specs, we wanted to include granular detail on why an IMA was not used.

But I would not be at all disappointed and would be willing to stipulate

right now that we would remove that particular exclusion, the suitability, because of its subjectivity. I have no problem with that.

CHAIR MORRIS: Does anybody in the group want to talk about that?

DR. HALPERN: I think that's a great idea. Then are you going to have some kind of analysis, because on the negative side of things, is like you were mentioning. If people are going to be judged by how many IMAs they're using, will people start using IMAs that may not be suitable?

DR. SHAHIAN: You know, we don't expect that -- even with the remaining exclusions on this list, we don't expect that every patient is going to get an IMA. Everybody is going to have a certain number of cases where the IMA will not be of adequate size or flow. That's why our median and mean usage are, you know, in the 94 percent or so range.

Is there the potential for

incentivizing surgeons to do the wrong thing? Yes. Like everything else that we've discussed, every process measure out there has that potential. But I don't think there are many surgeons that are going to use an IMA that is clearly inadequate because it comes back to bite you very quickly.

DR. HALPERN: Is there a target number that you're looking for then? Is there a target usage of IMAs that you're looking for?

DR. SHAHIAN: I'd say the best programs are up in the 95 percent range right now. Rich, do you agree?

DR. PRAGER: No, I agree. I agree.

CHAIR MORRIS: Any other questions or issues with regard to this particular measure?

DR. MORTON: I had a question about, is obesity a contraindication to using the IMA?

DR. PRAGER: Historically, yes. Currently, absolutely not.

DR. HALPERN: How about diabetes? I remember diabetes used to be an issue too.

DR. PRAGER: No, diabetes is not a contraindication.

DR. CARPENTER: So we're having this discussion about how, about the exclusion criteria. Do we have any idea how common that is used as an exclusion criteria? Is that something you capture? How often -- is it deemed unusable?

DR. PRAGER: We have not captured it on a national level. Regional groups have, and that's what has fostered including it in the national database.

CHAIR MORRIS: And then can you say a few words about the distribution of this, of use of the IMA. If it's used in 95 percent of centers, is there actually -- are there places that could do better, and how much better could they do?

DR. PRAGER: Well, experience at several of our states, Washington, Virginia, the Northern New England all saw after initiatives to increase utilization, regional improvement.

DR. SHAHIAN: But there are still programs, there are still isolated programs that are down in the 80s, 80 percent range. So there is definitely room for improvement.

CHAIR MORRIS: I have a question for you, Melinda. If we would like, as a group, we decided that we wanted to move forward, that we wanted to vote in an approving way for this measure, but we wanted to put on the condition that this exclusion be removed, how would we do that? How would we add the condition to that?

MS. MURPHY: Well, I'm looking at Helen. One of two ways. One would be that it would be voted down, and you would say with that condition of removal of it, that it would be acceptable. The other one, which would be

the question, is if STS committed now, if they're saying we can strike that, then it would be a vote on it with that struck from the exclusions at this point, contingent on getting that back in the measure documentation.

DR. PRAGER: I believe David and I would both agree with that, that we could strike that.

DR. SHAHIAN: Yes.

CHAIR MORRIS: Any thoughts or comments among the group with regard to that?

MS. STEED: So when we vote, we vote with striking that from the exclusions.

CHAIR MORRIS: Okay. Is there anybody here that wishes to vote on the exclusion as it stands without striking that?

(No response.)

CHAIR MORRIS: Okay, great. Let's go ahead and proceed with the vote then. So first, as you recall, does the measure meet NQF criteria for Importance to Measure and

Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: I'll ask everybody to once more press their vote and then press the send, aiming at Jessica.

[COMMITTEE VOTING.]

CHAIR MORRIS: We have 20 responses of yes and 2 of -- I'm sorry 1 of no. How did that happen? Oh. Second vote, does the measure meet NQF criteria for Scientific Acceptability of Measure Properties? This is with that exclusion struck.

[COMMITTEE VOTING.]

CHAIR MORRIS: 14 said completely, 7 said partially. Thirdly, we're voting on does the measure meet NQF criteria for Usability?

[COMMITTEE VOTING.]

CHAIR MORRIS: 20 said completely and 1 said partially. Next, does the measure meet NQF criteria for Feasibility?

[COMMITTEE VOTING.]

CHAIR MORRIS: 20 said yes completely, and 1 said partially. So with our last vote, does the measure meet all the NQF criteria for endorsement? Let's just recap what some of the issues were.

First of all, in terms of disparities and measuring disparities, the STS says that they haven't looked at this at a national level, but that they're willing to. It sounds like there are disparities in use among some regions that have been examined.

There was a question about the exclusion, which the STS agreed to strike. The exclusion specifically is that the IMA is not a suitable conduit due to size or flow, since at this point there aren't explicit evidence-based criteria for determining that, and it really is the judgment of the individual provider, based on potentially their experience, their training. But not something that somebody external to them could

necessarily understand.

There was a question about whether or not this particular measure would have an impact, since it has -- since it's applied, estimated, in about 95 percent of cases. But the STS thought that there was still substantial variation, in that since the introduction of this measure in particular regions, that there had been an improvement in IMA use.

Then there was a mention of concerns regarding gaming the system. This speaks back to that particular exclusion that was struck. Any other issues that anybody has with regard to this measure, before we do our last vote? Any comments?

(No response.)

CHAIR MORRIS: Okay. So the last vote, does the measure meet all the NQF criteria for endorsement?

[COMMITTEE VOTING.]

CHAIR MORRIS: And please once

more hit your vote and then hit the send button.

[COMMITTEE VOTING.]

CHAIR MORRIS: We had 21 votes for yes, no votes for no, which is pretty good. So moving on, next we have Dr. Wilhoit talking about 0119, Risk-Adjusted Operative Mortality for CABG.

DR. WILHOIT: Okay. So this measure our work group reviewed. I think many of the things we've talked about with respect to the other measures apply to this measure as well. So there's not a lot of new discussion.

For Importance, the outcome speaks for itself. I think it certainly has face validity, and mortality is certainly what patients and families care about a great deal.

In terms of Scientific Acceptability, the measure does not require participation in the STS database, but it is defined in terms of the STS database fields.

So again, that's similar to what we saw with some of the other measures.

Usability, we identified no issues, and Feasibility, we identified no issues.

CHAIR MORRIS: All right. Any issues that anybody wants to bring up? Questions or comments?

MS. STEED: I was wondering if there had been any consideration to changing the measure from 30 days to 100 days.

DR. SHAHIAN: I can speak to that. The answer is no, although I think in the future, perhaps the next time we're coming back to you, that might be possible. Right now, it is difficult, costly and time-consuming even to obtain 30 day data. We have, however, established now a linkage with Social Security Death Master File, that will permit us to obtain long-term mortality, and we in fact have developed, and will be publishing soon, a long-term risk prediction

model for CABG.

Because I agree, that as short-term mortality has diminished, it is increasingly relevant to all stakeholders, I think, to know what the longer-term outcomes are. In some sense, the early postoperative period has also lengthened, because of our ability to keep patients alive.

So I think it's a very relevant question. Right now, the answer is no. We don't have that capability, but I think the next time we come back to you, we will have that operationalized.

CHAIR MORRIS: Okay, and then --

DR. ROGERS: I have a question, Arden, that will have pertinence, I think, with the next section that we look into, which is all mortality-related. As I understand risk-adjusted mortality, it addresses the global characteristics of patients for a given institution, and that's kind of how that's done.

I'm concerned that that may be a relatively blunt instrument with respect to mortality, and I wonder if there's a possibility of discussing or at least raising the issue of what was the specific patient risk of each patient who died?

It's one thing to have an okay mortality rate. It's quite another to have the wrong people dying, if you look at their individual risk. I know that's possible to do, and I wonder if the sponsors of these criteria might comment about that.

DR. SHAHIAN: Well, we calculate -- obviously, in order to calculate the global risk adjusted mortality or O to E ratio for a particular hospital or participant, we have to calculate the estimated risk of death for each patient. We tally them up, and then we compare that with the actual number of patients that died.

Now you can also, and we have done this, looked at deciles of risk, so that you

can low risk deciles and high risk deciles, and you can do it across, you know, as broad a distribution of risk as you want. We've done this, and actually the model performs very well across the spectrum of risk.

Performance of the model tends to fall off a little bit at the extremes, and this is true of any risk prediction model. But over the broad range of typical expected mortality rates, model performance is very good.

CHAIR MORRIS: Are you satisfied with that answer?

DR. ROGERS: I wish I knew. I'll stop here for the moment. Thank you.

CHAIR MORRIS: I think that that's actually a very insightful question, in that it's -- I'm not aware of very many situations where that has actually been looked at. What is the actual risk for individual patients, and are the patients expected to survive surviving, versus those that are expected to

have a harder time?

DR. SHAHIAN: Well, if I can just expand a little bit further, no prediction model does a fantastic -- there's not a prediction model in existence that does a fantastic job for individual patients. It can tell you for a patient, for a general patient that has renal failure, severe chronic pulmonary disease and an emergency operation, if you take 100 of those patients, five will die.

Unfortunately, most models won't tell you which five patients will die. The best metric that gets at this in terms of model performance is discrimination. Discrimination of our models has been pretty good.

So you take all possible discordant pairs of patients, where a patient lives -- one patient lives and one patient dies, take all the combinations of those, and then ask in how many of those individual

experiments, if you will, did the patient that died have a higher probability of death than the patient who lived? That's discrimination or C index.

That's about the closest we get to answering the question that you just posed. But we are not real great at, and no model in existence is great at, predicting for an individual patient, John Smith.

We can do a pretty good job of describing what generally happens to patients like John Smith that have John's particular combination of risk factors. But actually predicting for John is very difficult.

DR. ROGERS: Well, I think it's less about predicting for John than it is knowing about what the patient -- who actually died. That can be done, and I happen to have a slide.

But we can wait. I mean I think it's an important point. It could be that the patient -- usually, this comes in the context

of, well, you can measure our stuff in our hospital, but you haven't the faintest idea how tough it is to do what we do -- you know, the old adage of my patients are sicker.

In fact, this addresses it very well. It could be in a hospital that has a very good mortality rate for very sick patients, for some reason it doesn't do so well with patients who aren't that sick, or the opposite may occur.

I'm just, I'd just comment that I think the measure that we're proposing for all of our mortality, I think, doesn't take that into account. If indeed that is -- if we're capable of doing that, are we putting ourselves in an advantageous spot by approving something that's going to last for the next three years, when in fact there's something that could be done now that could be used in a more advantageous way over that period of time. I don't know the answer to that question.

DR. SHAHIAN: Well, I think model testing that should be performed and that we perform, and I think all responsible model developers do this, they assess calibration.

I think what you're talking about, I think, is calibration. How does the model do for patients with low risk? How does it do for patients with medium risk? How does it do for patients with high risk? Am I misreading you?

DR. ROGERS: Again, I wish I was smarter about this. What I do know is that what I see is that what it does do is it looks retrospectively, and immediately at the risk, the preoperative risk of the patients who died, and gives you some reflection of you may have an absolutely acceptable OE ratio. You may have a star or you may have a smiley face, similar to other hospitals.

But the patients who died probably shouldn't have died, as opposed to the patients who died at another hospital whose

risk, individual risk, was really quite high and they did the best they can. That would be my response, whether that's useful or not.

DR. DUTTON: I think some of what you're asking, Terry, has to do with the uses you make of the data, rather than the data itself. Risk adjustment and comparison to national benchmarks for mortality are very common in trauma. I know we've been doing it for 20 years in trauma.

Part of looking at that report is not just how do you do statistically O to E overall, but who are your unexpected deaths and who are your unexpected survivors, and what lessons can you learn from them. So it's more of an application than a requirement of the data itself.

DR. SHAHIAN: All right. Now I see what you're saying, and that's absolutely right. I think the way to get at that -- it really has to be at the institutional -- or you have Dr. Prager there who really has led

nationally an effort to look at every single death in his region and divides their hospitalization up into various time periods, and tries to identify was this a preventable death? What could have been done better?

So I absolutely agree with that, but I think that's not a risk adjustment issue so much.

DR. SEARS: One other comment. We don't take into account here any of the volumes of the programs. So a small volume program that has five percent mortality, but within the confines of competence limits may be normal for them, you know, for year-in over year-out, versus a program that does 500, where you might get a more accurate assessment of the true mortality.

So it's just a comment, but it makes using this as a measure somewhat difficult to adjudicate.

DR. STAFFORD: I think volume, though, comes up in some of the other

measures, and I suspect that's going to be a very large discussion, because if you look at the literature on volume, depending on what procedures you're talking about, it runs the gamut.

For a lot of procedures, we don't know. The question becomes is it the volume? Is it the volume per surgeon? Or is it the system that you work within that's really the matter, and not necessarily the numbers?

So I think what, at least in this large database, you kind of take some of that out of the picture and have to look at your own institution. But I think it's probably going to generate a lot of discussion as we get into some of the other measures.

CHAIR MORRIS: I agree with that. Will we have the opportunity to request harmonization of this measure, of mortality with volume measures, or is that already in the works for later?

MS. MURPHY: There will be an

opportunity to look at areas for harmonization.

CHAIR MORRIS: Okay. So let's keep that in mind, and let's go ahead and proceed to the vote. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: 21 said yes, 1 said no. Next vote is does the measure meet NQF criteria for Scientific Acceptability of Measure Properties?

[COMMITTEE VOTING.]

CHAIR MORRIS: I'd like to ask you all to just hit your vote one more time, and then hit send, directing toward Jessica.

[COMMITTEE VOTING.]

CHAIR MORRIS: 17 said completely meets the criteria; 5 said partially meets the criteria. Next, does the measure meet NQF criteria for Usability?

[COMMITTEE VOTING.]

CHAIR MORRIS: And one more time, please press down firmly on your button, and then hit send.

[COMMITTEE VOTING.]

CHAIR MORRIS: 21 said completely meets the criteria, 1 said partially meets the criteria. And then does the measure meet NQF criteria for Feasibility?

[COMMITTEE VOTING.]

CHAIR MORRIS: Okay, and please press down firmly on your button again and then hit send?.

[COMMITTEE VOTING.]

CHAIR MORRIS: 20 said yes, it completely meets the criteria; 2 said partially. Then lastly, does the measure meet all of the NQF criteria for endorsement?

So just to recap briefly, we talked about risk-adjusted mortality rate, and could it be better performed to better identify whether the people who should do well actually are doing well within institutions.

We had quite a bit of discussion about that. Ultimately, Dr. Rogers, did you feel satisfied with that discussion?

DR. ROGERS: Partially. I'm okay, that's fine. I just wanted to raise the point. I appreciate the opportunity.

CHAIR MORRIS: Yes. I do think it's an important point to bring up. Then there was the point that volume is not taken into account with this measure, and perhaps we'll find an opportunity to harmonize with a volume measure in the future.

I think that the underlying message there is that in some ways, in high volume centers, in medium and high volume centers, volume actually may predict mortality with more regularity than mortality does from year to year.

In low volume centers, that's probably even more so the case. Any other issues that anybody wants to bring up with regard to this measure before we vote?

DR. SHAHIAN: This is Dave Shahian. Could I just ask that your last statement contain a qualifier from us. The evidence for a strong volume relationship for CABG, which is I think what we're discussing right now, is really fairly problematic. For other procedures that we'll talk about, I would concede.

But for CABG, risk-adjusted mortality is a much better predicter, and in fact even John Birkmeyer, who's an advocate of volume as a performance metric, has published a paper with Justin Dimick showing that for CABG surgery, risk-adjusted mortality is a much, much stronger predicter than volume.

CHAIR MORRIS: I agree. I think that's an important point, and I think that volume is a better predicter for some specific operations.

Then there are other operations which will not necessarily come up in this particular session, but that will definitely

come up in the future, within general surgery, for which volume is really not a good predicter at all. So that is something to keep in mind. I agree.

So does the measure meet all of the NQF criteria for endorsement?

[COMMITTEE VOTING.]

CHAIR MORRIS: We had 22 out of 22 saying yes, it does meet the criteria. Okay. So now we're moving on to Work Group B, and first we'll start with Dr. Dillon, 0120, Risk-Adjusted Operative Mortality for Aortic Valve Replacement.

I'd also like for you to try and provide information regarding whether the measure developer has anything to report since -- for maintenance measures -- since the measure was initially endorsed. So is there anything new that we know since the measure was initially endorsed, for those that are being maintained.

Then we need to probably pay a

little more attention to disparities and what is known or what is not known about disparities in these measures.

DR. DILLON: Right. Team B will take over or take the baton from Team A, but we'll probably continue with a familiar theme here.

So NQF 0120 involves risk-adjusted operative mortality for aortic valve replacement. It looks at the percent of patients undergoing isolated aortic valve replacement who die within 30 days or within the hospitalization.

It is publicly reported. It is certainly a measure that is of great importance to the public. In terms of scientific validity -- and I guess, Dr. Morris, addressing your question -- in terms of what is known about some of the disparities, certainly the risk-adjustment process is well-established.

They have multiple years of

analyzing and tracking trends across institutions with this. The issue does come up in terms of being able to determine some of the disparities using this measure alone, in terms of subgroup analysis, which was one of the minor points that we discussed as a work group, that perhaps with stratification or further evaluation, STS database would allow for that information.

But the report itself does not allow, does not allow one to determine that at face value. It certainly is quite -- the Usability of it, we found as a group no particular issues or problems, and certainly the Feasibility was adequate.

So in general, our work group had no particular issues with this measure, and indeed I think there was rather a unanimous of congruence in terms of evaluating it.

CHAIR MORRIS: Anybody have anything else to say about this measure? Issues or comments, questions?

DR. STAFFORD: I have just a quick question. So for the overall mortality after CABG, the time frame for the denominator was 12 months, and here, the time frame for the denominator is 60 months. Is there something that informs that difference in time frames?

DR. PRAGER: I'm not sure. It should be on, sorry. I'm not -- the 12 months is not in the specs that I read, unless I missed something in the 60 months.

DR. STAFFORD: So in --

DR. PRAGER: I may be missing something. I mean we report 30 day or in-hospital.

DR. STAFFORD: So it's in the application for mortality after CABG, in the denominator statement. Denominator time window is 12 months. In this measure, 0120, same place, the denominator time window is 60 months.

DR. SHAHIAN: No, I think the issue there is that we aggregate data for the

valve cases, because there are far fewer of them historically than CABG cases.

I don't know if Jane is still on the line. I don't -- I'm not absolutely sure that 60 months is correct. But if she has that down here, I'm sure it is. But because of the smaller number of cases, we do aggregate the valve cases.

DR. PRAGER: But for the individual site, you'll get a report that will give you what happened in three months or six months.

DR. DUTTON: In other words, it's always 30 day mortality, and it's just a question of whether you're reporting a five year window or a one year window?

DR. SHAHIAN: That's right, right.

DR. STAFFORD: Right, yes. The numerator's the same. It's the denominator that I was struggling with.

DR. SHAHIAN: Yes.

CHAIR MORRIS: Okay. Are we ready

to vote? Anybody else want to bring anything up with that one?

(No response.)

CHAIR MORRIS: Let's move on then. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: And we said -- 20 out of 20 said yes. The next vote, does the measure meet NQF criteria for Scientific Acceptability of Measure Properties?

[COMMITTEE VOTING.]

CHAIR MORRIS: 20 said completely and 1 said partially. Next vote, does the measure meet NQF criteria for Usability?

[COMMITTEE VOTING.]

CHAIR MORRIS: 20 said completely and 1 said partially. Then the last -- then the second to last vote, Feasibility. Does the measure meet NQF criteria for Feasibility?

[COMMITTEE VOTING.]

CHAIR MORRIS: 21 out of 21 said

completely. The last vote here, does the measure meet all the NQF criteria for endorsement. Then just to recap again, it sounded like Team B actually approved of this measure pretty uniformly. There was a question or discussion of stratification by race, ethnicity, gender or other socially vulnerable markers.

And the report here does not clarify the presence of disparities, but that could potentially be clarified further by the developer. In addition, something that we haven't talked about quite so much was the plan for public reporting. So this is, as I understand it, not a publicly reported measure. Is there a plan in place for publicly reporting?

DR. PRAGER: David and I will echo this. Yes, there's a plan for over the next several years to roll out multiple measures for public reporting, and aortic valve is the next in queue, and we would hope within 12

months to have that be able to be publicly reported.

CHAIR MORRIS: Can you describe the plan?

DR. PRAGER: Can I describe the plan? The plan is in its early stages of creating a composite for public -- a composite metric for publicly reporting the outcomes for aortic valve replacement. David Shahian is leading this with DCRI.

DR. SHAHIAN: Yes. I can just speak briefly to that. We are developing -- as you know, for CABG, we have a combination of outcomes and process measures that we publicly report. There is not the analog of internal mammary artery use in the case of valve surgery, and we are going to confine this particular composite to strictly to outcomes measures.

So this aortic valve composite measure that we'll be publicly reporting will consist of risk-adjusted mortality and the

five major risk-adjusted morbidities, stroke, renal failure, reoperation, prolonged ventilation and I've left one out. Five majors, yes.

So we're developing that right now, and the goal is to develop it this year and publicly report it next year. It's been a big undertaking to publicly report the isolated CABG. We're still working the kinks out of that, and we wanted to be pretty far along in that process before we roll out a second measure.

But I think we can commit to 2011 for public reporting, or 2012, excuse me.

CHAIR MORRIS: Okay, thank you. So we'll go ahead and vote. Does the measure meet all of the NQF criteria for endorsement?

[COMMITTEE VOTING.]

CHAIR MORRIS: We had 21 out of 21 say yes. The next measure to discuss is 0121, Dr. Sears, talking about Risk-Adjusted Operative Mortality for Mitral Valve

Replacement.

DR. SEARS: Thanks. This is very similar to the last discussion, where we're looking at the percent of patients undergoing strictly mitral valve replacement who die, either within 30 days of operation or within the time of the hospitalization.

Our team felt that this was an important measure. There was really no problems with the scientific validity, usability or feasibility for the measure.

CHAIR MORRIS: Any issues anybody wants to bring up with regard to this one, issues or problems?

(No response.)

CHAIR MORRIS: That's a very short discussion. Okay. Is there -- no? Okay. Well, let's go ahead and vote. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: I'll ask everybody

to once again press firmly on your vote and then press send.

[COMMITTEE VOTING.]

CHAIR MORRIS: We had 21 out of 21 say yes. The next vote, does the measure meet NQF criteria for Scientific Acceptability of Measure Properties?

[COMMITTEE VOTING.]

CHAIR MORRIS: I'll ask everybody once more to press firmly on their vote and hit send.

[COMMITTEE VOTING.]

CHAIR MORRIS: 20 said it completely meets the criteria and 1 said it partially meets the criteria. Does the measure meet NQF criteria for Usability? Please vote twice this time.

(Laughter.)

[COMMITTEE VOTING.]

CHAIR MORRIS: 21 out of 21 said completely meets the criteria, and it seems that voting twice works well. Next, does the

measure meet NQF criteria for Feasibility.

[COMMITTEE VOTING.]

CHAIR MORRIS: 21 out of 21 said completely meets the criteria for Feasibility. Next, does the measure meet all of the NQF criteria for endorsement? We really had no discussion about this to speak of. I have nothing to recap. Is there anything anybody wants to say about the measure before we vote on it?

DR. SEARS: For public reporting as well, what's the plan?

DR. PRAGER: The plan is yes, we'll develop the aortic model. As David said, there is no IMA for these, so it will be based on the other publicly reported NQF-endorsed aspects, and it would probably be sequenced, and hopefully it will be less than the next full year after, because we will have experience in the model.

DR. WILHOIT: One small question about that that I was thinking about with

respect to what you said about the aortic valve replacement. We're looking here at a stand-alone measure, but what you're talking about publicly reporting are composites. Those are very different. So I just wondered, just thought that was an issue worth putting out on the table.

DR. PRAGER: Yes. I'll let David handle most of that, but our public -- what is publicly reported currently now is a composite metric for coronary bypass, based on a feeling created over years that this allows the opportunity to put in multiple measures and have various discretionary factors in the model that's reported. David, can you add to that?

DR. SHAHIAN: Yes. Just like our CABG model, the AVR composite model we develop will have drill-down capability, so that you'll be able to look specifically at the risk-adjusted mortality component, which will be the equivalent of this measure, and we will

come back to NQF at a future time, to get the composite, to put the composite before you for endorsement.

CHAIR MORRIS: Okay. Do you feel that that answers your question?

DR. WILHOIT: Well, one of -- you know, looking at the documents -- one of the criteria for even considering a measure is that it be brought forward for both quality improvement and public accountability.

If the measure isn't being considered for public accountability, as a stand-alone measure, which is what we're looking at, is it one, you know, does it meet that criteria? That's my question, I guess.

DR. SHAHIAN: Well, if it -- the composite then has X and Y in it, it's publicly report. If X is endorsed and we have the ability in that composite to separately report the X component of it, it seems to me that satisfies the requirement for public reporting. I mean youþll have it. Youþll

have this publicly reported.

You'll just have more. You'll have this plus the ability to have this incorporated into a larger composite measure. But you'll have this. It will be publicly reported. We'll also have the capability of providing you with additional information on other outcomes as well.

CHAIR MORRIS: So let me just try and clarify that a little bit. With your public reporting of a composite measure, then you would also publicly report the components that went into the composite?

DR. SHAHIAN: We do that now. If you look at the public reporting that we have of the STS CABG composite, it gives the -- you have, it's four domains. One of those domains is risk-adjusted mortality for CABG. You can go to our website right now and get that information, or go to Consumer Reports.

CHAIR MORRIS: So your plan for the composite measure would include -- it

would be an individual composite measure for each of these operations then?

DR. PRAGER: Yes.

CHAIR MORRIS: Okay. So Dr. Prager is saying yes to that. Let's go ahead and vote, unless there's anything else anybody wants to add. Thanks for your comments, guys. Does the measure meet all of the NQF criteria for endorsement? And please vote twice.

[COMMITTEE VOTING.]

CHAIR MORRIS: 21 out of 21 said yes, it does meet the criteria for endorsement. Just to clarify for the transcript, as we started this, we learned when we started this meeting today that voting twice is only recorded as once. Early and often.

All right. The next measure is 0122, Risk-Adjusted Operative Mortality for Mitral Valve Replacement and CABG Surgery, and this Dr. Rogers.

DR. ROGERS: Yes, thank you. Just

a point of clarification. Is it okay press once, twice and then send twice, or do you have to do it in sequence?

CHAIR MORRIS: I recommend doing it in sequence; otherwise, you'll be voting 11.

DR. ROGERS: Thank you. We are dealing with risk-adjusted operative mortality, mitral valve replacement plus CABG surgery, and I think the same kind of issues that have previously been reported on the two similar measures apply here.

Except for my residual and probably singular concern about the risk adjustment, there was no issue with respect to the importance to report. Similarly, the same would apply to both Scientific, Usability and Feasibility. So it's pretty straightforward and I think similar to what we've already heard.

DR. MORTON: Just one question. Any exclusion criteria about this?

DR. ROGERS: I'm not aware of any. Dr. Prager, do you know?

DR. PRAGER: No. As long as it's a repair and a coronary -- I'm sorry. As long as it's a replacement and a coronary bypass, no.

DR. ROGERS: Did you have something in mind?

DR. MORTON: Re-dos.

DR. ROGERS: I'm sorry.

DR. MORTON: Re-dos. If it's a --

DR. ROGERS: Oh, I see. Yes.

DR. PRAGER: I think it is what it is. I'm not aware. David, do you know?

DR. SHAHIAN: I'm sorry, I didn't hear the question.

DR. PRAGER: David, the question is whether -- what exclusions? So the patient had a coronary bypass and then they're having a reop for a mitral and a coronary. Is that case excluded?

DR. SHAHIAN: No. The reoperative

status is just included in the risk adjustment. You have -- Iþd have to go to the specific measure, but I'm sure there are some --

DR. CIMA: There's a comment here that in one spot, it says "replacement," and other spots it says "replacement/repair." It's supposed to be replacement, correct?

DR. PRAGER: This is supposed to be replacement, yes.

DR. CIMA: There's a separate one for repair?

DR. PRAGER: Correct.

DR. SHAHIAN: Are we talking about the isolated MVR or the MVR CABG right now?

DR. PRAGER: MVR CABG.

DR. SHAHIAN: I'm having trouble finding that in my document. Are exclusions listed, Rich?

DR. PRAGER: I'm looking, too.

DR. HAN: There are no exclusions listed, Dr. Shahian. This is Jane.

DR. PRAGER: So no exclusions listed in the document.

DR. WILHOIT: One question I had about this measure, looking at the data that was provided, is that the volume of these cases seems to be low. In the other measure that had to do with volume, it reported a median of 27 and a mean of 38 cases.

And this is a five-year measure, which makes the denominator a little bigger, but -- because you're aggregating over five years. But even with five years of data, the number of centers for which results are reported is only 33.

That makes me wonder whether, how useful of a measure this is. It may be that it's useful for the higher volume places, and that's good enough. But it just seemed like it was -- that the results were different enough, in terms of the number of centers for which there's data. It seems like it was worth putting that on the table.

DR. SHAHIAN: Yes. Dr. Prager comes from one of those institutions nationally that does a very high volume of mitral valve surgery, and I think you're right. This measure will be much more relevant for certain centers where this procedure is practiced more commonly.

Aortic valve replacement is done at most hospitals. Mitral valve surgery, because of the special techniques done in repair, and the desire to try to get as many valves repaired as possible, rather than replaced -- these cases tend to gravitate to major centers. Rich, do you want to comment on that?

DR. PRAGER: No. I would just echo that, and I think your observation is absolutely correct. But I think the importance of it remains.

CHAIR MORRIS: Okay. Are we ready to vote? Any other comments, issues, questions?

(No response.)

CHAIR MORRIS: Does the measure NQF criteria for Importance to Measure and Report, and I'll ask you to hit whatever your vote is and then send, and then hit whatever your vote is and send.

[COMMITTEE VOTING.]

CHAIR MORRIS: I'm sorry to say this, but let's all vote one more time.

[COMMITTEE VOTING.]

CHAIR MORRIS: We had 19 out of 19 say yes, it does meet the criteria. The next vote, does the measure meet NQF criteria for Scientific Acceptability of Measure Properties?

[COMMITTEE VOTING.]

CHAIR MORRIS: 16 said yes completely; 3 said partially. Does the measure meet NQF criteria for Usability?

[COMMITTEE VOTING.]

CHAIR MORRIS: 16 said yes completely and 3 said partially. Does the

measure meet NQF criteria for Feasibility?

[COMMITTEE VOTING.]

CHAIR MORRIS: 18 said completely, 1 said partially. And then lastly, does the measure meet all the NQF criteria for endorsement, and we talked very briefly about the public reporting plan. We talked about the limited number of centers that does this sort of operation. Any other issues that anybody wants to bring up?

(No response.)

CHAIR MORRIS: Okay. Let's go ahead and vote.

[COMMITTEE VOTING.]

CHAIR MORRIS: 19 out of 19 said yes, this does meet the criteria. Okay. Let's see now. Next is Dr. Saigal, 0123, Risk-Adjusted Operative Mortality for Aortic Valve Replacement and CABG.

DR. SAIGAL: Okay. So this is in the same theme. They're looking at the clinical patients who have a combined aortic

valve replacement and CABG who die either in the hospital, after the surgery or within 30 days of their discharge.

In terms of the Importance of the measure, I think it has got great face validity, itþs very important in patients and it varies by center. In terms of the scientific validity of it, as modeled and tested and published, they didn't report anything about disparities. I'm sure that's actually available pretty easily.

I didn't see any information about how this measure has changed practice or had an impact, but I'm sure that's available as well. And Feasibility and Usability, I think, are also acceptable.

I didn't see a public reporting plan either, but I think it's probably in the same sort of plan that they've articulated already.

CHAIR MORRIS: Thank you. Anybody have any other issues, comments, questions

about this measure? Can the STS respond to the issues around no information regarding disparities on any data about the impact of this measure and can you confirm or add to the question of a public reporting plan?

DR. PRAGER: Disparities, again, is something that we have data that can be, that can be turned into information for all of us, frankly. Two, public reporting plan will be sequenced. I will be honest, that I am not sure we can give you a time frame, 24 to 36 months perhaps.

The aortics and then the mitrals, and then mitral coronaries, aortic coronaries, hopefully tied closely to each other. The third, other than intermittent publications of research on these types of patients, I am not sure we have, and using it in regional collaboratives and starting to look at what we may get into later about volume and outcomes in these cohorts of patients, I'm not sure we can tell you anything other than we expect

these numbers to continue to rise as technology has influenced this group of patients, having more hybrid operations, more coronaries done in the cath lab, and then more higher risk valves done in the operating room, including now percutaneous valve approaches.

So I think we are at a cusp, if you will, for this. David, do you have other thoughts?

DR. SHAHIAN: No.

CHAIR MORRIS: Okay. Any other issues before we go to a vote?

(No response.)

CHAIR MORRIS: All right. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty out of 20 say yes. Does the measure meet NQF criteria for Scientific Acceptability of Measure Properties?

[COMMITTEE VOTING.]

CHAIR MORRIS: Eighteen say completely; 2 say partially. Does the measure meet NQF criteria for Usability?

[COMMITTEE VOTING.]

CHAIR MORRIS: Nineteen say completely; 2 say partially. Does the measure meet NQF criteria for Feasibility?

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty-one out of 21 say completely. Does the measure meet all the NQF criteria for endorsement?

DR. SAIGAL: Could I ask one question before we vote?

CHAIR MORRIS: Yes.

DR. SAIGAL: The response about the mortality rates rising over time because the indications for the procedure in the OR have changed, and a higher-risk population is undergoing the procedure in the OR. Does that mean the risk adjustment model is failing to account for that, or should that be like changed, because of the nature of the

population undergoing the procedure?

CHAIR MORRIS: Dr. Prager, do you want --

DR. PRAGER: No, I apologize. I may have been misinterpreted. I didn't say that mortality rates are rising over time. We are starting to see a higher-risk population, because of percutaneous approaches and other hybrid approaches, so that more people are becoming, if you will, candidates for a higher-risk operation.

Those data, those analyses of those data -- we don't have all those data yet. This is just over the last year or two or three or four, and we expect to see more.

DR. SHAHIAN: So we're seeing more patients with specific high risk predicters, and the expected risk of the patient population is increasing. But in fact observed mortality has not been increasing as rapidly or remains stable or in some cases declining from any of these procedures. So,

and the data, these models are recalibrated every year, as well.

But it's mostly the expected risk that's been increasing.

DR. SEARS: One other thought here. Do we need to alter this a little bit with the advent of the percutaneous techniques, as they go through their evolution, and should this measure be more geared to open repair or open replacement?

DR. HALPERN: Actually, you'll see there's two coming up that are repairs, rather than replacements.

DR. SEARS: Okay. Well no, but I'm just talking about percutaneous versus open aortic valve replacement. Not repair, just replacement, either through a catheter technique, which is not an open-chest procedure.

DR. PRAGER: David, help us a little bit, but catheter technique is not in this model at this point; correct?

DR. SHAHIAN: Well, that's right. I mean we are collecting data on those patients, but those patients are all being done under protocol right now.

CHAIR MORRIS: So it sounds like that is -- in a way, it's sort of an exclusion, because they're being done under protocols, so they're not really captured here; is that correct? Okay, and potentially we'll be looking at this in the future. Okay. Any other issues, comments?

(No response.)

CHAIR MORRIS: Does the measure meet all of the NQF criteria for endorsement?

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty-one out of 21 say yes, it does meet all of the criteria. The next measure is Dr. Searsþ, 1501, Risk-Adjusted Operative Mortality for Mitral Valve Repair, and this is split from Measure 0121, and I'm hoping that you will clarify the reason for splitting from Measure 0121.

DR. SEARS: Yes. This is very similar to 0121, in the fact that you're dealing with a mitral valve. But this specifically looks at repairing of the valve and not replacing it.

So that there are various techniques that are used to repair the mitral valve today, and so it's incumbent upon us to recognize that the repairs are different than the mitral valve replacement.

When we talked about this within our group, we felt this was an important value for the society, for the Quality Forum to endorse. The Scientific Acceptability was pretty much uniformly accepted, as was the Feasibility and Usability of the measure.

DR. HALPERN: This is actually where I had more of a question of what repairs are being included, because the other ones were specifically replacement. These next two are repairs.

DR. SEARS: You're talking about

1502 as well? Well, 1502 is mitral valve repair plus CABG. This is strictly a mitral valve repair, no CABG involved with it. So theyþre different techniques.

DR. HALPERN: They're all open chest.

DR. SEARS: Right, they're all open. The mitral valve is visualized and then whatever technique the surgeon opts to use. That doesn't break down technique, obviously.

CHAIR MORRIS: Can you go a little bit further in your question? I'm not sure that I'm understanding specifically what you're asking.

DR. HALPERN: As somebody was indicating, there's percutaneous ways of repairing valves now. But this is specifically referring just to open-chest cases.

DR. SEARS: Yes. Again, I think, and David probably can clear this up more than I can, I think most of the mitral valve

percutaneous techniques are still being done under some kind of a protocol. Wouldn't you say that's right, David?

DR. SHAHIAN: Absolutely. That is -- I think, Rich, most of those are still being done in Europe, aren't they?

DR. PRAGER: Yes. There are some in this country. They're mostly protocol- or company-driven at this point in time and they are not in this grouping of patients. These are all operative patients. Could be robotically, it could be right thoracotomy, could be mediastenotomy, could be a left thoracotomy. They're operative patients, not percutaneous.

CHAIR MORRIS: And then split from the other measure because it's a repair and not a replacement; is that correct?

DR. PRAGER: Yes, that's correct.

CHAIR MORRIS: Okay, and is there more that you'd like to say about it? Okay. Any other issues with regard to this?

(No response.)

CHAIR MORRIS: Are we ready to vote? Great, okay. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty-one out of 21 say yes. Does the measure meet NQF criteria for Scientific Acceptability of Measure Properties?

[COMMITTEE VOTING.]

CHAIR MORRIS: Nineteen said completely, 2 said partially. Does the measure meet NQF criteria for Usability?

[COMMITTEE VOTING.]

CHAIR MORRIS: Nineteen said completely, 2 said partially. Does the measure meet NQF criteria for Feasibility?

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty-one out of 21 said completely. I'm afraid that there's not much for me to recap on this one, unless anyone wants to bring anything up. I actually

don't have any further comments. Please feel free to bring anything up that you think is important to note before the final vote.

(No response.)

CHAIR MORRIS: Okay. Does the measure meet all of the NQF criteria for endorsement?

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty-one out of 21 said yes. The next measure is 1502, Dr. Rogers, Risk-Adjusted Mortality for Mitral Valve Repair and CABG Surgery, again split from Measure 0122.

DR. ROGERS: Right. This is the evil twin of the prior measure evaluation. Risk-adjusted operative mortality for MV repair and CABG surgery. I have really nothing new to add that hasn't been said already. Meets the criteria for Importance to Report, Scientific Acceptability, Usability, Feasibility, and there are no listed exclusions. So we would recommend its

approval.

CHAIR MORRIS: And I assume the same issues with lack of information about disparities, lack of information about changes in practice, and lack of -- and future plan for public reporting of the composite measure.

DR. ROGERS: I have nothing to add to that.

CHAIR MORRIS: I hope that the lack of discussion is not because everybody's been beaten into submission by the carbs at lunch. But we'll probably pick up a little bit as we move on. So let's go ahead and vote. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty-one out of 21 said yes. Does the measure meet NQF criteria for Scientific Acceptability of Measure Properties?

[COMMITTEE VOTING.]

CHAIR MORRIS: Sixteen said

completely, 4 said partially. Does the measure meet NQF criteria for Usability?

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty said completely, 1 said partially. Does the measure NQF criteria for Feasibility?

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty-one out of 21 said completely. Does the measure meet all the NQF criteria for endorsement? This is the last chance to say something about this measure before we take a vote on it.

(No response.)

CHAIR MORRIS: All right. Let's go ahead and vote.

[COMMITTEE VOTING.]

CHAIR MORRIS: Twenty-one out of 21 said yes, it does meet all the criteria. The next measure is Dr. Halpern, 0124, Surgical Volume. A, Isolated Coronary Artery Bypass Graft Surgery, B, Valve Surgery, C, CABG and Valve Surgery.

DR. HALPERN: I think everybody was saving up their discussion for this one. So this only looks at volume. There is no other quality measure here, and therefore it's like only one number. It's not a numerator, a denominator.

We had a very intense discussion about this in our group, and all of us felt that volume alone cannot be -- is not an adequate quality marker, for reasons that have already been brought up here today earlier, including the fact that low-volume places may still have good quality.

There was actually an editorial in the Journal of Thoracic and Cardiothoracic Surgery, because there was a paper published out of Japan, where their high volume was actually our low volume in this country, and they had very excellent outcomes, pointing to the fact that processes may be more important than volume.

So with that, that meant to us

that it didn't even pass the first criterion.

CHAIR MORRIS: Any other things that came upon your discussion that you'd like to bring up now?

DR. HALPERN: No. Really, those were the main ones, and the main question that got discussed quite vigorously was whether volume alone could be a quality marker.

CHAIR MORRIS: Okay. Anybody else want to make a comment before we ask STS to respond? Dr. Morton?

DR. MORTON: Was there mostly concern about just volume in general, or was it the number that was given?

DR. HALPERN: It was volume in general as a sole marker for quality, because this is for quality improvement. It's important to know volume, but so it is important to know -- measure and report, but even the person who was on the phone who was of the public, i.e., not a physician but the consumer, felt that she was concerned also by

our discussion, that volume alone would be a quality marker, having listened to our discussion about it.

MS. KENNEDY: And I guess I can comment, because I was that person. I think just by itself, it could just be kind of misconstrued, if there's not something to accompany it, as far as the quality performance.

DR. STAFFORD: I would add to that. So not only, and you could have all of these quality measures on somebody's publicly available dashboard. So you could have volume and you could have all the other measures we talked about. But I can tell you, as probably most lay people, the first thing they're going to look at is going to be volume.

I mean that's a number. I mean people conceptually get numbers, and they may or may not look at all the other quality measures that go with it. So you could have a very high volume center, but then you go and

look at their mortality, you look at their renal failure, you look at their incidence of stroke, and it may actually be way worse than some of the smaller volumes.

So you're right. In context, you know, if somebody just looks at the number, which I suspect is what a lot of people would do, that could be a problem.

DR. HALPERN: Another thing that was brought up is that if volume alone becomes your quality marker, will people be doing more procedures than they actually need to, in order to make that volume criteria?

MR. FINDLAY: Well, I guess STS is going to respond here, but was this intended ever as a measure to stand alone, apart from --

DR. HALPERN: It actually was -- from my understanding from the discussion that we had, is that CMS actually had this marker and STS picked it up.

MR. FINDLAY: Okay. But it's

meant as a dashboard really. I mean in reality you wouldn't use this alone. The physician community wouldn't use it alone, and the worry is that consumers might and would, and that's misleading and we all know that. So I concur.

DR. BURSTIN: I'll just point out that NQF has at times paired measures, indicating one should always be reported with the other, next to each other, and I believe the AHRQ quality indicators that have volume attached to them, in fact, actually Melinda was the lead on this, in fact are paired measures.

So volume is only reported with the mortality measures, not as a stand-alone. Is that right?

CHAIR MORRIS: There's also a question about whether the STS supports public reporting of this particular measure. So could the STS please respond to the concerns about whether this is a valid marker of

quality standing alone, about whether this is meant as a dashboard actually, about whether they do or do not support public reporting of this measure, and to the concerns about manipulating the system potentially?

DR. SHAHIAN: I'm going to take that one on. Actually, the editorial that you referred to was my editorial, and I've dealt a lot with this volume issue. I think the discussion you just had has been a very high level and appropriate discussion.

We in general have greatly favored the reporting of risk-adjusted outcomes or composite outcomes, as opposed to volume, because some entities, including CMS, Leapfrog, others, have expressed an interest in volume, and because for some procedures that are less frequently performed than CABG, there is a volume outcome association that's much stronger, we included this measure.

But it was not without some trepidation, because we don't want to seem to

be endorsing this as a stand-alone quality metric, for all the reasons that you've discussed. You know, we have a measure here that is well specified in terms of the particular types of procedures that should be included if you're going to talk about volume.

So if a volume measure were going to be used, we're comfortable with the one that we have proposed here. But in general, you know, volume is simply a surrogate or proxy for outcomes measurement, which is the preferred measure. If in fact outcomes measures are available, then that's always preferred.

CHAIR MORRIS: Any other discussion on that?

DR. DUTTON: I'm sorry. I couldn't -- David, could you be a little more specific? I'm not sure what you just said. It sounds to me like we actually don't believe in this one, but want to recommend it anyway. Could you help us lay people figure that out?

DR. SHAHIAN: Well, there is a lot of -- there's a lot of interest in volume out there in the measurement world, and you know, I think if volume is going to be used, we would just as soon be a player and define it correctly and have an appropriate source for that volume.

However, we think that risk-adjusted outcomes are preferred. I would not lose any sleep if this measure, if you decided not to endorse this measure. But I would hate to have CMS or some other entity come back with another volume measure that you did endorse. I mean I think that -- so that's basically where we stand.

I think there are some procedures where volume is a pretty darn good metric or proxy. One is esophagectomy. Another one is pancreatectomy, where there is a very strong volume-outcome association. For the major operations in cardiac surgery, not so strong. Certainly for CABG. Very weak for CABG.

Moderate for valve procedures, I would say.

DR. KLEINPELL: Well then really, based on what you've just said, I would question why is this being proposed then, you know, in terms of there has been volume to outcomes for others, but not necessarily for this. But it seems that the impetus for submitting this is because CMS may do that. Is that -- that's sort of how I'm interpreting this discussion. If you could help me out a bit?

DR. SHAHIAN: Yes. I mean if there is a desire from other stakeholders to have a volume measure, we think it should be one that's well specified, that gets the right combination of, you know, procedures, and we just as soon have it be an STS measure. But we are not wildly enthusiastic about using volume as a metric, when there are good risk-adjusted results available.

CHAIR MORRIS: Melinda.

MS. MURPHY: There are two

approaches that NQF has taken in the past with respect to measures, particularly of volume, and Helen mentioned one, and that is to recommend that it always be reported with mortality. That's one thing.

Another, whenever measures have been considered for use in a composite, is that a measure be evaluated in terms of its use within a composite only, not as a stand-alone measure.

CHAIR MORRIS: So it seems to me that the appropriate thing to do here would be to proceed to a vote, to know that if we vote this down, that we could make the recommendation that it always be paired with a more meaningful outcome measure in the future.

DR. STAFFORD: Well not only that; I'd actually rather see all of the other measures presented stratified by volume. That might give me a better idea of an effect of volume. But I think even pairing it isn't

what I would want to see. I just want that out there.

CHAIR MORRIS: Okay. So let me just sort of reiterate, to make sure I've got it here. So your preference would be to not include it, but if it is included, to stratify other measures by this. So not necessarily to pair it with another outcome measure, but instead to not consider it to be a measure at all, but to stratify other measures.

DR. STAFFORD: Exactly, and report it that way.

DR. SEARS: If we report -- if we vote this down, can somebody else vote it back in through another committee, if you're going to use a CMS measure, for example?

DR. BURSTIN: There is no other measure.

DR. SEARS: Well, this --

DR. BURSTIN: It's theoretical, I think, a theoretical concern that if there's going to be another measure, you prefer to use

the one built off the registry.

DR. SEARS: Right.

(Simultaneous speaking.)

DR. HALPERN: I think then what we're more concerned about is outside agencies putting that forth as a quality marker, outside of the NQF. So like somebody mentioned Leapfrog was very hot for a long time on volumes as a marker of your ability to do -- you know, the more you did, the better you were, which is not necessarily the case.

DR. PRAGER: I would just add the historical aspect of this is this is an adopted measures by STS, one. Two, we really do, as David eloquently stated, you know, this is not how we believe you measure quality. However, there are others that utilize this to measure quality. If that's the case, then either pairing it or having it available, we believe, keeps the field level, if you will.

But we're not endorsing it as an independent quality measure, and that is part

of one of your responsibilities.

MS. STEED: But that doesn't mean that the NQF needs to endorse it.

DR. ROGERS: We actually could make a statement by not endorsing it.

CHAIR MORRIS: All right. So then in that case, let's go ahead and vote on this first criteria. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: I'm sorry. Let's go ahead and start over with that. We're going to vote on the Importance to Measure and Report.

[COMMITTEE VOTING.]

CHAIR MORRIS: So 4 said yes, 17 said no. This does not meet Importance. Beg your pardon?

Okay. So we don't have to go through the rest of the votes for this particular criteria.

I think based on the discussion

among the group, that we really want to add that the concerns, both that Dr. Stafford brought up and that others brought up, that either this should be paired with another measure or it should be used to stratify, but should not be used as a stand-alone measure.

DR. SHAHIAN: That's perfectly acceptable to STS.

CHAIR MORRIS: All right, next measure. Oh, next is Work Group C, and we have a break. We actually made up for lost time here. Should we go ahead and take our break right now and then come back for Work Group C? Let's do that. So a 15-minute break, and let's reconvene at five to 3:00.

(Whereupon, the above-entitled matter went off the record at 2:40 p.m. and resumed at 2:59 p.m.)

CHAIR MORRIS: All right. We're going to go ahead and get started. We have somebody from AHRQ on the phone. I'll ask all of you guys to take your seats. Okay.

This is, I believe we have John Botts on the line from AHRQ who we'd like to give the opportunity to say a few words about the measure, since you weren't be to be present this morning earlier. Can you hear me okay?

DR. BOTT: Well, we seem to be rather minimalistic in our comments. We're mainly here to be able to respond to questions.

But really, really quickly in regard to the two AHRQ QIs. One is of course a volume measure, a count measure for counting the number of the procedures. The other measure is a mortality measure which uses risk adjustment in the measure.

These are measures that use electronic claims data set to complete the measures, the electronic claims data defined by, inserted by the user into the AHRQ QI software. So that's as much as I have for an introduction. I believe a couple of other

folks are on the call, if they want to add anything.

DR. GEPPERT: And this is Jeffrey Geppert from the QI support team, and I have nothing particular to add, but I'd be happy to answer any questions.

CHAIR MORRIS: Then, is Dr. Romano on as well? Maybe not yet. We're ahead of schedule a little bit here, which is an unusual situation. My question for you is, we were just talking about mortality in volume measures, and talking about the importance of pairing them.

Can you discuss any plans for pairing these measures in the future, or is there anything that we should know about what you have in mind for these measures together?

DR. BOTT: I'll let Jeff comment on it further, but at the time, we do not have concrete plans to implement something different. We're exploring some things, but that's the phase we're in, that exploratory

phase.

So at this time, we do not have a particular plan for it with a due date and what version a change might go into. Jeff can feel free to elaborate further in any direction, if you'd like.

DR. GEPPERT: Yes. So the indicators, you know, could potentially be paired in two respects. So they were developed as a pair. They were intended to be analyzed together jointly, because the rationale for the mortality measure was the volume-outcome relationship, so they were not ever intended to be analyzed jointly, and have been since they were, you know, initially released in 2002.

The more sort of recent methodological pairing has to do with incorporating both measures into a single composite measure, sort of a volume-outcome composite measure. The construction of the composite measure uses basically the same

methodology that's incorporated in the QIs and what we cal the smooth rates.

Basically, it's the univariate shrinkage estimator, and in some of the Leapfrog work it's basically the same methodology, but the key difference being that sort of your prior, the thing that you shrink to, is a volume-specific mortality measure.

So a couple of things in respect to that. One, as John mentioned, there's an indicator development process that AHRQ employs, and so the earliest something like that could be implemented after it went through that AHRQ process would be at least a release from now, at least. Our most, our current release is scheduled for some time this spring. So it would be some time after that, after it went through a review process.

The second thing I'll just say about that is that all of the pieces that you would need to construct such a composite are currently available as output from the current

QI software. So you know, you need the volume measure to shrink back to. You need the mortality measure, and then you need the weight, the reliability weight that's used in the shrinkage, and that's an output from the software as a signal ratio, which is a parameter to the software, and a noise, a signal variance, which is a parameter, and a noise variance, which is computed and reported.

So all that sort of information is currently sort of incorporated into the software and available, you know, for researchers.

CHAIR MORRIS: Okay, thank you. The measure's going to be introduced by Dr. Siperstein.

DR. SIPERSTEIN: Great, thank you. I think some of the points have already been made. This measure, 0360, Esophageal Resection Morbidity Rate is described as being paired with 0361 that does look at the

hospital volumes.

The measure has been around for a number of years, and just in way of background, why look at esophageal resection? It's a relatively unusual procedure. However, it's a particularly high-risk procedure for a number of different reasons.

The patients themselves are fairly high risk due to comorbidity, older age, smoking, ethanol and also malnutrition secondary to the obstruction from their cancer. As they typically present, they've often received preoperative chemo and radiation therapy.

In addition, the operation itself is particularly high risk, with entry into multiple body cavities. This has really in the literature been one of the prototype procedures for correlating operative mortality with hospital volume. So unlike the prior measure, there's a long and established literature track record making this

correlation.

The measure uses mortality rate, where the numerator, simple number of deaths in patients undergoing esophageal gastric resection for cancer, and the denominator has to do with the number of hospital discharges. In our phone conversation, one of the issues, not an issue, it's addressed. But just in terms of there's a risk adjustment model that's built in. One of the questions for the folks on the phone has to do with a little bit of the vetting or the detail of that risk adjustment model. It also uses hospital discharges, as opposed to 30-day mortality, and one of the potential issues, although practically probably not a big concern, is for example hospital transfers are specifically excluded. So that may skew the data slightly, and there is a stratification of results that can be presented.

In terms of usability, the only potential issue is that because it's a low

volume procedure, those hospitals that are doing this in relatively low volume may have fairly wide confidence intervals in terms of reporting their mortality data on a center-by-center basis, but in terms of an aggregate, would hold up to statistical scrutiny. In terms of the feasibility, relatively straightforward, because it uses billing/administrative data.

CHAIR MORRIS: Thank you. Anybody who was involved in that work group want to comment on some of the things that we discussed in this measure?

(No response.)

CHAIR MORRIS: Dr. Dutton, is your mic on for a reason?

DR. DUTTON: Yes. I wasn't in that work group. I was waiting. But you mean, you did say that transfers were excluded in this measure?

DR. SIPERSTEIN: Yes. It's based on hospital discharges.

DR. DUTTON: It seems like a pretty substantial flaw, in that I mean it's death after esophagectomy is not a clean kill. It's a go to the ICU and get multiple organ failure and dwindle, and wouldn't a lot of those patients tend to get transferred to tertiary care centers, thus biasing this result pretty substantially?

DR. SIPERSTEIN: I mean as described as a potential flaw, I do not know what the track record's been in that department. It wasn't in the materials.

DR. GEPPERT: I've done some -- excuse me. This is Jeff Geppert again, some analysis of -- there's two ways to sort of think about that. So a lot of our QIs are inpatient measures, and so, you know, some of the steering committees that we've worked with have viewed that as sort of an inherent characteristic of the measure. Not that inpatient is supposed to proxy for something else, but it's reflective of the patient's

experience in that hospitalization.

The other perspective is the relationship with an out of hospital measure, and we did do some work with some linked mortality measures, linked vital records data. The general finding was that the vast majority of the mortalities occurred in hospital.

There were very few patients that were transferred post-procedure for this particular measure. We were capturing, you know, 90-some percent of the deaths. The reason, let me just explain the reason for the transfer is that over time, to avoid this double-counting at discharge.

So we're using state hospital discharge data, so we don't presume that one can link it from one hospitalization to the next. So if you don't link from one hospitalization to the next, and you only want to count a patient once in the denominator, you count them in sort of a receiving, into the -- you count them in the receiving

hospital.

So that's the rationale. But for this particular measure, it doesn't have a big impact, because there's not that many patients that fall into that category.

DR. WILHOIT: One thing that concerned me about this measure is that, as best as I can see, there's not a minimum number of cases required to report it. The median mortality is about six percent, which means that on average, I'd need to do 16 cases to have one die.

If I haven't done 16 cases, then my rate's going to be zero and I'm going to look really good. But we know from the next measure, 361, at least as best we can tell from the data that's provided there, a lot of places or a lot of surgeons wouldn't do 16 cases.

So you end up with a lot of zero results, and folks look really good, but not based on the fact that care is good

necessarily but that the numbers are small. So I'm wondering if the small numbers, at least at a facility basis, make this not particularly useful, and whether it's, you know, more useful on a larger scale.

But I'm not sure how helpful of a quality measure it would be for assessing hospitals or doctors.

DR. SIPERSTEIN: I think that's exactly what I tried to point out in the summary, and what came out on our conference call, was exactly what you're, you know, what you're saying, is that the purpose of this is not for an individual hospital to market their results, but for statistical purposes, you really have to look at aggregated results.

Also being interested from the sponsors, in terms of what the track record has been in terms of reporting, because a number of centers are using this and reporting it.

DR. GEPPERT: Can I make one

statistical point about the volume, the low volume? So the way that's addressed methodologically is in two ways. One is that the software actually does not report any rates for a denominator less than three. So there is a threshold in that respect.

Then the second way it addresses volume is it uses this shrinkage approach. So you're right. For a small volume hospital, there's a probability that the rate would be zero, even if the true rate were in fact not zero. So the shrinkage addresses that by pulling hospitals closer to the overall mean.

So a hospital that had a zero observed rate, even a zero risk-adjusted rate, would not have a zero smooth rate, which is what we call it, or a shrunken rate. The rationale for that is being that the shrinkage rate, the shrunk rate is a better predicter of future performance, a better predicter than zero would be essentially. That's the rationale.

But I'll let John just mention what we know about how this indicator is being publicly reported.

DR. BOTT: Yes, which I don't have a lot of information on. AHRQ doesn't really systematically go out and canvass the community as to who's using which measures how. But just informally because we're fairly close to a number of states, we do compile this largely for NQFþs purposes.

But we're really not going back to states to ask them what their experiences have been with the measure. If people have technical issues or questions or concerns or suggestions about any given measure, they use the AHRQ QI support line to deliver those questions and to resolve any issues they're having for consideration for future enhancements to the measures, which happens quite a bit because of the widespread use of the software and the measures.

DR. GEPPERT: The biggest topic

that we've received comments on through user support is not so much about low volume and reliability, but just whether we're capturing the right set of procedures. So we went through our clinical panel review a few years ago and made some refinements to the denominator with that in mind. Those have been the bulk of the comments.

DR. ROGERS: It's not clear to me, is this administrative data or is it -- it all is administrative data. So I have a question relative to the conversation we've had all day. Looking at what are relatively, particularly CABG, relatively high-volume procedures kind of globally, and the kind of useful information that can be fed back to those people who are doing it, no matter where they happen to be, with esophagectomies, I think we're talking about considerably lower volumes.

So can I ask the question, is the intent of this kind of study, to lead to a

kind of different conclusion that we've been pondering earlier, and the conclusion here would be, you know, if you're not going to -- if you can't anticipate you're going to do more than eight or 10 or 12 or 15 procedures, you shouldn't be doing them at all? Is that the intent of this measurement?

CHAIR MORRIS: Terry, are you asking the developers that question?

DR. ROGERS: Yes. Well, whoever might have the answer. I mean because it's a different animal, I think, than we've been talking about and has been pointed out. I see hidden in here this notion, and I'm not opposed to it. I think if I were going to have my esophagus out, I'd be sad to begin with. But then I'd kind of look at some place who actually did have some experience. So help me, sponsors, with that.

DR. GEPPERT: That rationale has certainly been put forth, that there's a safety accountability component for this

particular measure, and there's kind of a related measure for pancreatic resection, which it is to be discussed at a later time, but a similar low volume kind of procedure.

So certainly that's the suggestion, that if you're performing one or two of these a year, and we know on average that hospitals that perform that few procedures have significantly higher mortality rates than hospitals that perform, you know, 15 or 20 of these procedures, and that's a safety accountability issue.

DR. CIMA: Just to follow up on that, can you sort of clarify in Section 1(b)(2), where one of the issues we're looking at performance gap, and most people, you know, actually the STS has been presenting it differently.

They sort of give you a total percentage. You gave this distribution of medians and values. Can you just explain what that means, because it's got the 5th, 25th,

median, 75th, 95th, and then there's a series of numbers underneath.

Is that just for percentage or what exactly is that telling us about performance gap? Because that's going to be a real issue if you have a very low volume system. How is it going to impact this, versus higher? If you look at the UHC data, which is major, about 230 major academic centers, there's only like five that do more than 75 esophagectomies a year, you know.

So that's a real, you know, depending on where you sit on this spectrum of hospitals, that's a huge difference. So can you clarify what that performance gap data shows us?

DR. BOTT: Are you referring to the volume measure or the mortality measure?

CHAIR MORRIS: We're talking about the mortality measure, and Dr. Cima was talking about 1(b)(2).

DR. BOTT: 1(b)(2) in the

mortality measure application.

CHAIR MORRIS: We're talking about the mortality measure.

DR. GEPPERT: Yes. So that distribution information comes from basically the -- it's the distribution of the hospital performance, estimated from this Bayesian kind of approach. But it's basically the distribution in the hospital rates after you've done this shrinkage process.

So it takes a lot of our risk-adjusted rates, which can be high, but there's a lot of noise in that. It shrinks them down. Sort of, the more noise in the measure, the more it shrinks it down, and those distributions reflect the hospital distribution after that shrinkage has occurred.

CHAIR MORRIS: I have a related question. The signals-to-noise ratio that's described in (2)(c), the validity testing section, it looks like there's about three and

a half times more noise in there as signal in this measure, which is troubling, and probably related to everything else that's being said.

DR. GEPPERT: Well, not necessarily. It's troubling in the sense that I mean that's the reason you do the shrinkage. If an indicator has a high signal-to-noise, you know, .8 and above, then shrinkage isn't really necessary. There's a lot of signal in the risk-adjusted rate.

If the signal ratio is lower, as it is to be used in a lot of low volume, infrequent types of measures, then that means you want to do the shrinkage. So once you've done the shrinkage, then you've sort of accounted for that, and to the extent that there's variation that remains after the shrinkage has been done, that's true signal.

DR. SIPERSTEIN: Well, maybe I can partially answer that, because in looking through some of the references that were provided, when they aggregated data for lower

volume and higher volume centers, there was a very large gap between -- in mortality between the lower and the higher volume centers.

So you know, although it may average a five or six percent mortality rate, it ranged from like 2 to 12-plus percent in the various groups.

DR. GEPPERT: Right.

DR. SIPERSTEIN: So there's a very wide range, and I think the issue or the strength of this measure is in its aggregated view of the world, rather than trying to make any statement about an individual, particular low volume center.

DR. WILHOIT: And the thing, you know, the whole thing, the more we talk, the more uncomfortable I get, I guess. I know that, you know, if it's out there as an AHRQ indicator, you push the automated software button from the AHRQ website, and it produces and you do whatever you jolly well please with it, you know, without the benefit of being

aware of what the small numbers mean.

AHRQ has done a great job of automating things and making the data readily available, but then it assumes you know what you're dealing with. I don't have tremendous statistical expertise, but the more I hear about, you know, the noise being higher than the signal and using statistics to smooth that out, again, the less comfortable I feel with it.

DR. BURSTIN: I just want to point out that, and I know the folks from AHRQ know about this, this is Helen, that we actually had a discussion about the competing measure which you'll come back to at the end of this, that was recently endorsed from Leapfrog, where the whole basis of that measure was actually focusing almost predominantly more on volume rather than risk adjustment, clinical risk adjustment.

We commissioned an evidence report for this, and in fact found volume explained

about a third of the variation mortality for esophagectomies. We went ahead and put that measure through, as we did for pancreatectomy, a very similar RQI, a competing RQI.

I think one of the ideas was that AHRQ was interested in, and as and I think that Jeff was indicating this earlier, that as they move to the next version of the RQIs, they would potentially also add volume smoothing, in addition to the clinical risk factors smoothing, because we know these are areas that are really very, very highly volume-sensitive.

So I just want to at least put that on the table. You'll have a chance to revisit this after you evaluate it and have a chance to look at the competing measures issue. But you are absolutely right. The volume issue, as Dr. Stafford raised this morning, is huge, particularly in this procedure.

DR. HALPERN: I think it goes back

also to Terry's point about so what does each death mean. Like why do low volume hospitals have poor outcomes? Is it because of their process or just the numbers? Is it because they're not used to taking care of these patients and what might happen to them postoperatively?

CHAIR MORRIS: On the other side, if you have a low volume center that does one esophagectomy and the patient lives, then they have 100 percent survival. So it's not just that low volume places look bad; it's just, you just don't know what they look like.

DR. CIMA: I hate to just point that out, but you know, I agree with Carol. I mean I have less confidence for very low volume things, all the statistical manipulation that has to go into them to give us a number, then to publicly report that number.

It may be misleading. I know the data that says there's a tight relationship to

it, but in order to really put it out as a public reporting thing.

I mean so when we say Importance to Measure and Report, I think people would say yes, mortality rate for this procedure is important to know. But when you have to do all these manipulations to do it, does it really meet the criteria we're looking at?

DR. BURSTIN: I think some folks would say, particularly in an area where it is low volume and volume is such an important indicator, this is especially a place where public reporting is extra important for people to vote with their feet.

So that's why putting these data out here, and again, they've got -- I mean if you look at the statistical analyses AHRQ has done, they're able to explain. I mean happen to have the C statistic in front of me, John, but it's, you know, it's .851. It's a very good, robust model they're able to explain.

So I don't want us to, without

really diving into the statistics, just say it's complicated and therefore don't play. We should really take a deeper dive. And again, you can only look at the measure in front of you, knowing that they will potentially be looking towards adding more volume adjustment going forward.

But it has been used, I think especially, you know, God forbid any of us had to make a decision. It's kind of one of those things you'd probably go to the website really fast for, and try to find some information, because volume and mortality are so linked.

CHAIR MORRIS: We went through that, but I'm concerned that people in general won't really be able to do that at all, because I think even very sophisticated people don't really understand shrinkage necessarily.

One of the other issues that came up in our work group meeting was accountability at the hospital level. So we really don't know quite what this means for a

low volume hospital's accountability at that hospital level, and a medium volume hospital is too potentially, since this is such a relatively rare operation.

So that came up during our discussion as well. So, Bob, in answer to your question, what are we really voting on, I guess, you know, we have a few options here. We can vote on the measure as it stands. If we vote the measure down, we could make recommendations that it be paired with volume, or other recommendations. Okay. That it be linked. So Helen's saying that this is linked to volume. It's not really a stand-alone measure; is that correct?

MS. MURPHY: It's endorsed as an individual stand-alone measure with the recommendation that it be reported only with the pair, volume and mortality. That's the way they've submitted it.

CHAIR MORRIS: So that's what we're deciding on, whether that should be

continued, that this should be endorsed as a stand-alone measure that always be -- that we recommend always be reported with the volume measure. Allan, do you have more to add to that?

DR. SIPERSTEIN: No. I think again, it has to do with kind of philosophically what the purpose is, you know. I don't think the purpose of this is for an individual patient to go to the website to figure out whether their corner hospital has good results, because the statistics are too difficult to apply to an individual site.

It really has much broader applicability in terms of, you know, health care policy and how insurers want to direct their patients to given centers. So I think it has a higher, kind of higher level quality purpose than, as I said, looking at your corner hospital.

CHAIR MORRIS: I would like to bring up, in that context, a disparity issue.

Suppose you can't afford to go to the higher volume hospital?

DR. WILHOIT: The other thing is that while the intent may not be for you, the consumer who needs an esophagectomy, to go look on the website, I can tell you we have it on our website. You know, so do lots of other folks. So it is there, and marketed for people to look at.

DR. MORTON: You know, if we're looking at quality improvement, I guess in my mind what else is out there? Do we have the equivalent of an IMA process here for esophagectomy? There's not a lot that I know of short of volume, you know, that's been published out there to demonstrate differences, and I'm sensitive to the small numbers. But this seems to be the best thing out there at the moment.

CHAIR MORRIS: Does the STS have any stake in the discussion?

DR. PRAGER: I am sitting here as

a visitor. However, the reality of the discussion, the question is you've hit on all the salient features, and maybe you can tell, NQF can tell us, because I think the general thoracic is coming with measures in a few months.

But are they all pulmonary, or is there an esophageal measure that has more than volume in it? No here, to NQF.

DR. BURSTIN: I believe there's already, and I was going to check online. I believe there already is an endorsed general thoracic surgery measure from STS, which is a combination of, it scares me I remember these, mortality and morbidity following esophagectomy, specifically for cancer. I can pull up the details and share them with you, but it does not --

DR. PRAGER: So there is, yes. I didn't know if that was endorsed or not.

DR. BURSTIN: It was endorsed a couple of years back, two or three.

DR. PRAGER: Okay.

DR. BURSTIN: Again, this is a different data source. These are administrative-based measures. Public reporting is out there already.

DR. ROMANO: This is Dr. Romano. Could I address the linkage issue?

CHAIR MORRIS: Patrick, go ahead.

DR. ROMANO: Yes. This is Patrick Romano. I'm a physician member of the AHRQ QI support team based at UC Davis. I think this concept of linkage of these two indicators is an important concept, and it certainly is AHRQ's intent.

The way that I would describe this is basically that there's a certain volume threshold that hospitals ought to have, as a previous speaker said.

So for low volume hospitals, you would focus on the volume, and say well, this hospital is in such a low volume range that it's very unlikely that they would be able to

achieve high quality outcomes. Not impossible, but unlikely.

On other hand, when the volume gets up to a certain level, then it makes sense to look at the hospital's own experience, to look at the risk-adjusted mortality for its own patients.

So in that case, you want to know well, this is a high volume hospital. They do have the experience necessary, but are they able to achieve good outcomes given that experience?

So that, I think, describes why it's so important to look at these two indicators together, and certainly anybody who sponsors a report card is encouraged to present the indicators in that way. Many sponsors of report cards impose additional limits, such as a minimum number of cases to report mortality indicator, although that's not inherent to the design of the indicator.

CHAIR MORRIS: Okay. Is there

anybody who feels that their questions haven't been addressed adequately, or who would like to request more clarification?

DR. WILHOIT: I guess in light, in the light of Patrick's comment, which made a lot of sense, is it -- you know, is AHRQ open to limiting reporting the result to a hospital with a denominator of 15 or 20 or 25 or 30, or some number that's bigger than five or seven or ten?

I mean is that a way to address this, because I think we all understand that the mortality rate is important. It just seems that with a very small denominator, it's just hard to assume that it's meaningful.

DR. ROMANO: Well, that is the aim of the shrinkage, so that effectively if the volume is very low, then a hospital's mortality rate shows up simply as the mean mortality rate. So then the hospitals become indistinguishable from each other.

That has been the preferred

approach for all of the AHRQ quality indicators, and in our previous discussions with other NQF panels. But I'll defer to John about further details.

DR. ROMANO: I don't think, we don't have any particular plan at the time to come up with such minimum thresholds, as suggested. I defer to what Jeff had noted before, as that we need three to perform the calculation.

DR. GEPPERT: You know, the implications of what you're suggesting are very broad. I mean you look at, like, you know, the CMS mortality measures. They have a threshold, but the reason for the threshold is not because of statistics or the validity of inferences, but has more to do with concerns about confidentiality.

Statistically, you know, the whole rationale for the shrinkage is because it results in a better prediction than the non-shrunk estimate.

So the whole rationale, that if you're a consumer, and you're making a decision, you're going to make a better decision on average using these estimates than you would based on, certainly based on no information. That hardly seems like a formula for a good decision, but a better decision than you would make on just a simple risk-adjusted mortality measure.

So there is a direct connection between the methods and the usefulness for decision-making by consumers that provides the whole rationale for the -- you know, we're not doing this just because itþs good statistics. It's because it results in a better decision by a consumer at a particular hospital.

That's the whole purpose of this method. That's the whole purpose, so that the hierarchical models that CMS uses for AMI and CHF mortality becomes standard practice.

DR. CIMA: Just as a clarification, why does it have to be two

measures, two separate measures? Why can't a measure be designed that has integrated the two to give you a value?

I mean I'm just trying to understand why does it have to be two measures, that we have to get a faith and recommendation that it gets linked, versus why not there just be one measure?

DR. STAFFORD: Assume that with the Leapfrog tomorrow, when we look at the comparison of the three linked measures. The Leapfrog one looks like it integrates the two.

DR. BURSTIN: They're not exactly overlapping measures, though. The Leapfrog measure doesn't have any clinical risk adjustment. This one doesn't have any volume smoothing, per se. Correct me if I'm wrong here on the wording, Patrick. So in fact they're probably elements of both that are important here.

I guess one question might be, you know, is this something for the next iteration

of the QIs that you would recommend, that this just get, the volume get built into the measure. But for right now, I think the measure before us is what they have. They do pretty rigorous testing of their measures. Patrick, do you want to talk about future plans at all?

DR. ROMANO: I'll defer to Jeff and John.

DR. BOTT: Yes. Well, I thought we touched on that at the top of the call. Jeff noted the way in which volume is currently integrated into the measure, and there's some exploration of basically creating a composite, as Jeff characterized it before.

But we're at the front end of that conversation and that consideration, and it's certainly a consideration for a forthcoming version. I just don't want to right now promise that we're going to make that.

It needs to go through other steps in the evolution, and that some checks and

balances need to occur before making a decision to inform that decision if we're going to go there.

DR. GEPPERT: And just, you know, I mean I'm not sure that's -- if the concern is one of lack of transparency and statistical complexity, I mean, to make the recommendation, have it more statistically complex and less transparent is a little at odds. But you know, from a methodological perspective, the composite has a lot of things to its advantage, which is why it's under consideration.

DR. SIPERSTEIN: So if I can just kind of summarize what I think I'm hearing. I mean, obviously these two measures are being presented as a quote "paired measure." However, the hospital volume is not used as a risk adjustment factor. Am I understanding that correctly?

DR. GEPPERT: Well, more accurately, the hospital volume is not being

used to inform the prior distribution. That's the distinction with the Leapfrog measures. Hospital volume is used to inform the prior distribution in a Bayesian analytic context.

CHAIR MORRIS: Are you satisfied with that answer?

DR. SIPERSTEIN: Yes. I know what they're saying, yes. But I mean, obviously if you're not looking at both numbers together, it's very difficult to interpret. I fully understand the issue of the individual patient not understanding the statistical details, and potentially making, misinterpreting the information in terms of the quality of a particular center that they're looking for.

So the question is, and for a lower volume center, should there simply be an n/a next to it or saying that, you know, due to low volumes, we cannot report a statistically reliable number, as opposed to reporting, you know, zero percent versus 100 percent if you've done one case.

DR. BOTT: That's not what we're doing, so --

DR. SIPERSTEIN: No, I understand. You're throwing in a fudge factor in there, based on volume, to try to kind of regress it towards the mean a little bit. But still with a low volume center, your results are going to be very skewed by a very limited number of mortalities.

CHAIR MORRIS: I have a question actually for Carol. You mentioned that you are, in your organization and organizations like yours, you are looking at these numbers. Are you looking at them in the intended paired way, or are you looking at them individually?

DR. WILHOIT: Well, we do a couple of different things with the AHRQ indicators, and what we do in terms of our quality efforts, we actually don't report things that are really low volume, because of all the issues.

However, I know that on our

website, totally unrelated to quality directly, our marketing people post all kinds of things. If they can find numbers, they post them.

I honestly don't know if this specific indicator is there, but most of the AHRQ indicators, you can go through the marketing part of our website and pull things up, and anything that there's a methodology to run and there's data to run gets run, gets posted, and does not have clinical input necessarily to that. If that happens with us, I assume it happens elsewhere as well.

CHAIR MORRIS: I would say this sounds like it sort of speaks to your concern, is that right?

DR. SIPERSTEIN: Yes. I mean, you know, the question is you vet a measure. I mean, this measure's been out there for almost a decade, and you know, the question is there is, you know, I think as we had on our phone conversation, there is a lot of validity to

this measure, because I think as John pointed out, there is no other way to statistically deal with low volume. So you do the best you can, even though it's not ideal.

CHAIR MORRIS: So maybe a major question for the Committee is, do we want to go ahead and move ahead on voting to endorse this measure as a stand-alone, but knowing that it's paired with the next measure, or do -- and this is something that we all individually have to make a decision on -- or do we want to say that it really has to be a composite measure? That's sort of what we're wrestling with right now.

DR. SIPERSTEIN: And I guess the semantic point is, you know, what's the difference between these two measures being quote, "paired" and being quote, "composite"? I mean, is just it semantic or is it really a major functional difference? Just asking.

MS. MURPHY: And AHRQ would need to speak to that, but it very well could be a

significantly different result if you reported each of the measures, but report them together as a pair, versus them being integrated into a composite measure, where the way in which the data was handled might be different. So Jeff or John?

DR. GEPPERT: Well, the composite is basically a weighted average of the risk-adjusted rate, the volume-specific weight. So the difference between the weight is now where they're reported separately, and the way it would be reported as a composite is that, the way you would be reporting, instead of having a mortality rate, you would be reporting a weighted average of the mortality rate and the volume-specific mortality rate, where the weight is this reliability ratio.

So as Patrick was saying, for small hospitals, that ratio would be close to zero. So the rate that you would be reporting would be very close to the volume-specific mortality. So you can say it's very similar

to just looking at the volume itself.

For larger hospitals, the weight would be closer to one, although for these measures, never that close to one because the volumes never get that high. But they might be .5 or .6 for the highest volume hospital. Then the composite would be a weighted average of the observed risk-adjusted mortality rate, and the volume-specific mortality rate, with a weight of .6 and .4. That's what you'd be reporting.

CHAIR MORRIS: Okay. Does anybody, would anybody like to ask for any further clarification or additional questions with regard to this measure?

DR. GEPPERT: Just one last comment. So you can get that same result. As I was mentioning at the very beginning, you could get that exact same result by simply, you know, reporting the existing data sort of stratified by volume.

CHAIR MORRIS: Anybody else?

(No response.)

CHAIR MORRIS: All right, and I have one more question, and that is, if this came forward as a composite measure in the future, then could it potentially be examined as a competing measure to this, to both of these paired measures?

MS. MURPHY: To 360 and 361? I would suspect that it could, but I wouldn't know why AHRQ would retain the two if they built a composite.

CHAIR MORRIS: All right, thank you. Unless there are any other comments anybody wants to make, let's go ahead and take a vote, all right. So first vote, does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: 18 said yes, 4 said no. Next vote, does the measure meet NQF criteria for Scientific Acceptability of Measure Properties?

[COMMITTEE VOTING.]

CHAIR MORRIS: 3 said completely meets the criteria, 16 said partially meets the criteria, 2 said minimally and 1 said not at all. Third, does the measure meet NQF criteria for Usability?

[COMMITTEE VOTING.]

CHAIR MORRIS: 6 say it completely meets the criteria for Usability, 13 said partially, 1 minimally and 2 not at all. Next, does the measure meet NQF criteria for Feasibility?

[COMMITTEE VOTING.]

CHAIR MORRIS: 17 said that it meets the criteria completely, 4 said partially, 1 minimally. Then lastly, does the measure meet all the NQF criteria for endorsement? We had a little bit of a longer discussion with this. It's challenging to recap that.

I think bottom line, there was concern about low volume hospitals in

particular, and what their mortality rates mean in terms of whether they predict future mortality. We heard a little bit about the methods that AHRQ used to try to account for that and to correct for it to an extent.

People continued to express concerns about it, and particularly concerns about misinterpretation if this measure is maybe reported with the paired measure, but maybe extracted by anyone separately from its paired measure. So there were concerns about that.

Any other issues that anybody else wants to either underscore or bring up anew before we take our vote, our last vote?

(No response.)

CHAIR MORRIS: Okay. So does the measure meet all the NQF criteria for endorsement?

[COMMITTEE VOTING.]

CHAIR MORRIS: 14 said yes, 7 said no, 1 abstained. So it looks like the measure

-- looks like we as a group in general agree to endorse the measure. So the next measure is also Dr. Siperstein.

DR. SIPERSTEIN: Hopefully, this discussion will be slightly shorter than the last one. This really, as we've already discussed, is an identical measure, identical patient population in the metrics, but simply looks at hospital volume. And that ends my formal discussion.

(Laughter.)

CHAIR MORRIS: All right. So previously what we were talking about were concerns that mortality was not adequately predictive, and several people raised the point that volume is a little bit more predictive when we're talking about esophagectomy, particularly for anything other than a high volume center. Anybody want to bring up any particular points around measuring volume here?

DR. WILHOIT: I had one question.

On 2.F-3, it lists thresholds, and there's Threshold 1, which is six or more, Threshold 2 is seven or more per year, and then it repeats Threshold 2 as seven or more. Is that just a typo? Are there really only two thresholds, or is there a third threshold that's meant to be there but isn't there? I couldn't tell.

DR. BOTT: Jeff, can you see where the person's referring to?

DR. GEPPERT: Give me one minute here.

CHAIR MORRIS: Carol, could you repeat the location of that?

DR. WILHOIT: 2.F-3.

DR. BOTT: About two-thirds of the way down the form.

DR. GEPPERT: Threshold 1 is supposed to be six or more, and Threshold 2 is seven or more. The distinction between Threshold 1 and Threshold 2 is when we did our literature review, often the studies used

slightly different thresholds if they were using some sort of cutoff or reported results based on different thresholds.

So we were, the intention was to kind of report the range of thresholds that have been observed in the literature, six or seven.

CHAIR MORRIS: And can you confirm there is no Threshold 3?

DR. GEPPERT: There's no Threshold 3, yes.

DR. BOTT: It looks like they just accidentally copied and pasted Threshold 2 again. Sorry about that.

CHAIR MORRIS: Okay. Anything else anybody wants to add to this, to be a part of the discussion?

(No response.)

CHAIR MORRIS: I can say that in our work group, in our telephone conference in the work group, this didn't really provoke much conversation at all, did it, Allan?

DR. SIPERSTEIN: Well, I think the two measures were really discussed together as a paired measure. So we really didn't have a totally separate discussion about hospital volumes, because it was really brought out in the first discussion. That's, I think, what happened in the room here today.

CHAIR MORRIS: Thank you.

DR. WILHOIT: But I think the one thing that's just really interesting here is on 1(b)(2), is it gives the volume by quartile. The first quartile is one procedure; the second quartile, 1.4; third quartile, 2.4; and fourth quartile is 8.4. So 75 percent of hospitals, it looks like, do less than two and a half a year. So it just, it really emphasizes, I think, how important this is.

DR. STAFFORD: I was going to say the opposite. If we have the mortality and the outcome measure, why have a structure measure like this, especially because that

data's being gathered at the same time?

CHAIR MORRIS: Do the folks from AHRQ want to respond to that?

DR. BOTT: I don't have a particular comment. I don't know if Jeff does. He was more involved at the inception of the measure than I was or Patrick.

DR. GEPPERT: I guess I'm not quite sure I understand. So why report the volume separately?

DR. DUTTON: Yes, exactly. Why report the volume separately?

DR. GEPPERT: For the reasons that we were talking about before, where it's sort of an accountability issue. We want to be able to identify those hospitals that are performing a very, very low annual volume. There's a slight methodological or slight definitional difference between the two indicators, which was the further rationale.

The volume measure is focused on a particular procedure. The mortality measure

is a slightly restricted subset of that, which requires a diagnosis of esophageal cancer. The volume outcome relationship has been primarily documented on the basis of the procedure. But for the mortality measure, we wanted a more homogeneous definition of the denominator.

CHAIR MORRIS: Are you satisfied by that? You look quizzical. Let me see if I can rephrase your question, and make sure that it seems clear. It sounds like you're saying if they're already correcting for volume with shrinkage in the first measure, then why are we measuring volume again?

DR. DUTTON: Yes. I think that covers it.

DR. GEPPERT: Well, I think it provides additional information, right? I mean, if you have two hospitals that have the same mortality rates that are both at the mean, and one hospital has a minimum level volume and another hospital has the higher

level of volume, then that could have -- with the rationale, you have a higher degree of confidence in the mortality rate of the higher volume hospital.

DR. DUTTON: Okay. That's getting to be a fairly subtle concept, especially if you look at this from the point of view of public reporting. If Hospital A has the same outcome as Hospital B in terms of mortality, but Hospital B does three times as many, you think I should prefer Hospital B, that it's a better quality hospital, simply because it does more, even though they have identical mortality outcomes?

DR. GEPPERT: No. It's a question of which measure is a better predicter of future performance. That's the rationale behind all of these methodologies.

DR. DUTTON: So you're saying that volume is a better predicter of risk-adjusted mortality than risk-adjusted mortality. I mean, that's essentially the argument you're

making. I could understand where statistically that would be true, but I'm not sure, for public reporting purposes, it's very easy to say.

DR. GEPPERT: I don't think I'm saying that. I'm just saying that if you separately ran a correlation from one year to the next, and you did it based on risk-adjusted mortality in Year 1, statistical mortality in Year 2, and you stratified that analysis by volume, you're going to get a higher correlation for the higher volume hospital than you are for the lower volume hospital.

That's the sort of independent piece of information that the volume provides, plus the fact that it's a slightly different metric, for the reason that I said before.

DR. ROMANO: And the volume information, I think, helps you put in context the risk-adjusted mortality information, so that if the volume is very low, then you know

that that risk-adjusted mortality rate is really measured with a lot of random error, and you shouldn't put a lot of weight on it.

Now if it's shrunken down toward the mean, but it's still -- it's not worth putting a lot of weight on that measure. On that other hand, if the volume is high, then you know that you'll really get additional information value from looking at the mortality measure as well.

From the contracting perspective, certainly a payor might choose not to contract with a low volume provider, simply based on an evidence-based volume performance threshold.

DR. HALPERN: I think, though, what he's trying to say is if you're a patient looking at this data, that may not be so readily evident to them.

CHAIR MORRIS: I think -- I really think that we've said about all we have to say about this. I do think that it's been a good discussion.

I'm glad that you brought those -- I'm glad that everybody brought their points up, because I think we needed to clarify this and air a lot of those concerns, and to get as full an explanation as possible regarding the value of these measures.

There is one other issue that I think came up in the work group conference call, and that is -- and Allan, I'd like for you to particularly remind me if I'm not remembering this correctly or if you had a different interpretation of this.

I think the group was talking about the time span during which this would be measured, and whether it would be a 30-day measure versus an index hospitalization measure. Didn't the group -- I believe it's put forth as an index hospitalization measure, isn't it, and we thought 30-day would be more useful?

DR. SIPERSTEIN: Correct.

CHAIR MORRIS: Potentially. Are

we at least going to ask about it?

DR. SIPERSTEIN: Yes. Well, I think I mentioned that briefly during the former measure. But yes, it's index hospitalization, because it's administrative data. But clearly this measure is the number of procedures done, and so it's not going to be confounded by that.

CHAIR MORRIS: We could pick up mortality in the administrative data. The CMS data is very good on mortality.

So nobody under the age of 65 would be measured that way. Okay. Well, that clarifies that. Is there anything else that anybody wants to discuss with regard to this measure before we go ahead and vote?

DR. SEARS: I just want to bring up one other point. The problem here is the technique used to do the esophagectomy, I think, as well. Some people use thoracotomy and abdominal exploration and some people use, you know, abdominal and neck incision. So

that there are different ways, and I think it all filters down to what the mortality could be, depending on the techniques you use.

So I think these are hard to measure, because they're not asking for the technique, whether it's a thoracoabdominal incision or an abdominal incision or a thoracotomy or a neck incision.

CHAIR MORRIS: Okay. So perhaps some technical issues as well in terms of mortality. Let's go ahead and vote, unless you would like for the measure developers to make a comment about that, Nick.

DR. SEARS: They can if they'd like. I think I threw it out there because I think, I mean, I did a lot when I was a resident and a few when I was an attending, and I mean, it's just like it varied from case to case. You know, it depends where their tumor is and what you're going to be doing for the patient.

DR. CIMA: Just to clarify, you

know, the first one we said was just cancer. This one's volume, and now we're saying that they're paired. That it's volume against volume of all type of esophagectomies.

Although the vast majority are done for cancer, at certain institutions, certainly at my institution, we do a lot for motility problems, patients that have had caustic injuries, things like that.

So what would the AHRQ say about the influence of that, you know, because there are certain other reasons to do esophagectomies, and then we're saying it's paired.

DR. ROMANO: Well, the rationale there, I think, is that in the course of performing esophagectomies for other indications, surgeons and surgical teams gain experience, which is likely to improve their outcomes for all esophagectomies, including the largest upset for cancer.

So it's basically saying that the

experience that you get on other indications is relevant to your treatment of patients with esophageal cancer. It's giving the benefit of the doubt, if you will, to hospitals and surgeons that do a significant number of esophagectomies for benign disease. So those get counted.

CHAIR MORRIS: Thank you. Let's go ahead and move ahead with the vote. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: 18 out of 22 said yes, 4 said no. The second vote, does the measure meet NQF criteria for Scientific Acceptability of Measure Properties?

[COMMITTEE VOTING.]

CHAIR MORRIS: 8 said it completely meets the criteria, 11 said partially, 3 said minimally. Next, does the measure meet NQF criteria for Usability?

[COMMITTEE VOTING.]

CHAIR MORRIS: Let's have everybody hit their buttons one more time and hit send again.

[COMMITTEE VOTING.]

CHAIR MORRIS: 7 said completely, 14 said partially, 1 said minimally. Does the measure meet NQF criteria for Feasibility?

[COMMITTEE VOTING.]

CHAIR MORRIS: 17 said completely, 5 said partially. Then the next vote is does the measure meet all the NQF criteria for endorsement? And I'd like to remind everybody that this is endorsement of just this measure, the volume measure.

It's not -- as several of you have pointed out, our conversation included both this measure and the previous measure, when we were talking about the previous measure, and then again when we were talking about this measure. But we're really just voting on this measure right now.

And before we do our final vote,

would anybody else like to say anything else about it?

(No response.)

CHAIR MORRIS: Okay.

DR. ROMANO: Hello?

CHAIR MORRIS: Yes.

DR. ROMANO: Oh, I just wanted to point out also that I'm not sure if this was within the scope of your review, but there are separate reporting tools that AHRQ has produced, to help users and report card sponsors in the process of reporting information on these measures to the public.

So there are templates for public report cards. There's also a system called Monarch, which provides an electronic interface for generating web-based report cards. So people may want to look at those to see examples of how AHRQ suggests that these indicators could be reported to the public.

CHAIR MORRIS: All right. Let's go ahead and vote.

[COMMITTEE VOTING.]

CHAIR MORRIS: Let's hit our votes one more time and then send again.

[COMMITTEE VOTING.]

CHAIR MORRIS: Okay, and then one last time, press hard. I think the batteries are wearing down.

[COMMITTEE VOTING.]

So 16 voted yes that it does meet all of the criteria. 5 voted no, 1 abstained. Okay. So with that in mind, it sounds like the measure passed for endorsement. I think that that was an important discussion, to pull out the different issues within reporting on mortality, for esophagectomy particularly.

Our next measure is going to be introduced by Mr. Rivenburgh, and this is 1526, Transfusion Consent.

MR. RIVENBURGH: Measure 1526, Transfusion Consent. The description is the percentage of patients with a signed consent for blood transfusion who received information

about the risks, benefits and alternatives of transfusion prior to the initial transfusion, or the initial transfusion was deemed as a medical emergency, applicable to inpatients of all ages.

The numerator was the patients who signed the consent, or those patients who got their initial transfusion which was deemed as a medical emergency. The denominator was all patients who received red blood cells, platelets or plasma.

The exclusions were listed as none, but there was a question in 2(f)(1) about -- well, let me pull that up real quick. Patients greater than four months of age that had been selected for measures was used from the eligible measure population of inpatient discharges. So there was a little bit of confusion in reference to that, that we had discussed a little bit at the time.

CHAIR MORRIS: Thank you. This is a JCAHO measure. Does anybody else want to

add issues, comments, questions, before we ask for JCAHO to respond?

DR. STAFFORD: I have a question. I wondered why in the numerator they just chose blood, and in the denominator, they said red cells, platelets or plasma?

So why not have blood or blood products or have the numerator and denominator be the same from a language standpoint, because I think different people, people might see blood and think of red cells, as opposed to blood products. So I would make those standard either way.

DR. WILHOIT: The other thing that puzzled me a little bit was why emergency was a numerator event rather than an exclusion. It seemed more like that belonged in the exclusion bucket rather than in the numerator.

MS. ZAMBRICKI: We're on 1526, right? Okay. I had two comments. The first has to do with a Feasibility issue, and that is related to burden. I was wanting

clarification as to whether a preoperative consent that lists blood products and explanation given would be considered meeting this measure.

Then a second broader issue is looking at the literature supporting this measure, I was unable to find a connection between patients receiving an explanation and signing a consent, and reduced use of blood products.

There was one study from Australia that asked the opinion of people as to whether if they knew there were options would they choose options, and they said yes. But considering that we have experience with patients signing consents and explanations given for blood administration, there is no evidence to show that doing that results in less use of blood products.

I think we would all support communicating with patients and families about their care, and that really should be the rule

for everything, whether it's getting a CT scan it should be explained that the radiation can have a cumulative effect. So I just don't see the science that this is going to improve care.

DR. STAFFORD: Well, and I wonder if part of the point was for decreased utilization, and part of it I was seeing more under the umbrella of patient-centered care, making sure that patients are informed, sort of the latter part of what you were talking about.

Unfortunately, I would echo your thoughts, which is that there wasn't necessarily, or there isn't necessarily as much literature out there on patient-centered care, and I think that's just something that needs to be developed, not that there's evidence against it.

MS. ZAMBRICKI: The patient-centered care is really a culture that is a thread running through the entire hospital

stay.

DR. STAFFORD: And I'm going to add onto that. In terms of the burden, so blood is considered a pharmaceutical. It's considered a drug by the FDA. You need a prescription for it, which is why a physician has to order it.

So why don't we just take this all away? Every time I want to give somebody, you know, a beta blocker, do I have to get consent for that? I mean, when you really -- if you really want to take this down that slippery slope, I can see that happening.

While I understand there are some specific things related to blood and blood products, in terms of morbidity, the mortality issues aren't as clear, actually, and so I think the burden for this is really huge. I absolutely agree that we need to have discussions with patients about whatever we do when we can.

But if I have to get a consent for

every single thing that I do in the hospital, I'm never going to be able to take care of patients.

DR. SIPERSTEIN: I just wanted to comment that one of the additional complexities is that, you know, the indications for giving blood and the circumstances are very different in different parts of the hospital. What goes on on a medical oncology ward is very different than what happens on an orthopedic service and what goes on on a cardiac service, or goes on on a liver transplant service.

You know, in some of these instances, it's very, very predictable in terms of what's going to happen, and in other situations it's very unpredictable with what's going to happen. That just adds another layer of complexity in terms of trying to have a uniform model of patient discussion.

DR. SAIGAL: I had a comment about -- I wasn't sure the way they described it.

The denominator is people who have a consent signed for a blood transfusion, and the numerator is the people that got more information about blood products. Is that what it is? Because the way it's written, it seems to imply that, to me at least.

DR. CARPENTER: Well, as I understand it, the denominator is everyone who got a blood product, and the numerator is everyone who has a documented consent for that blood product, be it platelets or red blood cells or anything else, and back to Christine's comment.

I don't know, it wasn't clear to me what qualifies as consent, because a typical operative consent may have a box that includes consent for blood transfusion. That's not generally the main conversation that goes around about that consent. You're usually consenting for the operative procedure, and not that it's an afterthought.

But does that qualify or is it a

separate consent document that was required for those? I think that needs clarification.

DR. HALPERN: I think most operation consents, having been to at least seven different hospitals in my career, most operation consents include blood consent. But I agree. It's not the main focus of your conversation. You do say to the patient now we might need to give you blood. These are the risks of transfusions, you know.

DR. SAIGAL: But it says percent of the patients with a signed consent for blood transfusion, who receive information about the risk. So the denominator is people with a signed consent for blood transfusion, and then the percent of those that receive information about the risks. That's the way it's written.

CHAIR MORRIS: Well, the denominator is people who received red blood cells, platelets or plasma. That's the denominator.

DR. SAIGAL: I mean the brief description of the measure. So maybe it's a different sense.

DR. STAFFORD: Yes, it's different. I think the assumption is is that a signed piece of paper is informed consent, and we all know that's not the case.

DR. MORTON: Well, I was going to make the point about, that was just made earlier, that when you sign the consent for the OR, that's generally part of it. Even some general hospital admission consent forms have it, too.

I would say that blood is a little bit different. In a lot of ways, blood is, you know, an organ transplantation. We know there's a lot of downstream complications that have been associated with blood transfusions.

There's some patient-centric issues around this, if you're a Jehovah's Witness. So I think this was one I'd generally like to see discussed and consented.

DR. ROGERS: If the intent here is to have -- to lead towards more appropriate use of blood, it seems odd that we'd approach it in this way, because this doesn't address the issue of whether the blood, the desire or the impetus to give blood was in any way appropriate.

This just measures whether the patient agreed with the doctor, who may be completely wrong about the suggestion that they actually get blood. I'm really uncomfortable about this as a valuable measure, because I think if we're talking about appropriate use of blood, we're looking at the wrong audience, or asking the wrong question.

DR. AFSAR-MANESH: And to address that, there are measures coming up that will address that. But you're right. This doesn't really do that.

MR. RIVENBURGH: And I think the question falls to is are we looking at the

risk of giving blood, or are we looking at the issue of making sure that the patient understands all of the risks and benefits of what they are going to be receiving in this particular case, and that it's being fully explained to them from a patient perspective, not from, yes, itþs on the surgical consent, you know, it's on the medical, you know, the hospital admission form.

But are we truly saying to them these are all the things that could possibly go wrong when we give you a unit of blood?

DR. DUTTON: I think the latter is what we're aiming for, but I'm not sure this says -- it's feasible to capture that, because you're trying to capture a conversation with a piece of paper retrospectively. I'm not sure that works. Incidentally, it may be the patents who refuse the blood transfusion that we're more interested in.

CHAIR MORRIS: There was one other issue that I had with this particular measure,

and that was that it was unclear to me -- Dennis, you may be able to answer this or you may go and ask the developers to answer it -- but it was unclear to me whether this included signed consent for every single unit that was transfused or is it separated by a 24 hour period? How is that determined?

(Off mic comments.)

CHAIR MORRIS: So just the first unit in some period, some time period -- because you wouldn't want just the first unit in their life. You wouldn't want -- just the first unit a month.

DR. CIMA: But the question is a surgical consent, okay. You sign it. But then let's say you've had a long hospital course or something, and then you're a week away from surgery or something like that, and all of the sudden an intern comes by and says we're going to give you some blood.

I mean is that same consent from the time of surgery applicable here? It's not

-- there's no time specificity. I think I agree with everything that's been said. I'm not sure it adds value at all. But that becomes an issue then.

You know, what consent are we looking at? You can go through a chart and someone's been in a hospital. There are probably 20, 30 consents for different things if they've been in there long enough, and then you're going to have to try and pull this out. So it's going to become a burden for very little added value, in my opinion.

DR. HALPERN: And I think a lot of times, you know consents, like where I practiced previously, the consent from surgery, the blood consent was good for 30 days.

DR. BURSTIN: Just a question for the developer. It specifically does say in the notes for abstraction that for hospitals that use a general consent for treatment that includes transfusions, select yes. Do we know

how commonplace that was? You did do reliability sampling of this. How much of an issue is that, especially for surgery consents, in addition to just general hospital consents? It's pretty broad.

MR. FINDLAY: That was part of my question, too, is how often is general surgical consent with various other components around that, how often is this included in that? I would assume that most of the time?

DR. WILHOIT: I looked at the document that was attached, which was really long and cumbersome, and I thought it was hard to find things. But on page 41 is a flow chart, and that I figure out, because I don't know the codes that are on here.

But it separates out transfusion consent, and whether education addressed risks, benefits and alternatives to transfusion. But I can't follow the coding, so I'm not sure what ends up in what bucket. So I'm not sure -- you know, from the

abstraction instructions, I'm not sure what's being measured either.

CHAIR MORRIS: Any other issues before I recap?

(No response.)

CHAIR MORRIS: Okay. I was going to recap before they respond, just to try and be thorough, and make sure that we've covered the things that concern everyone.

First of all, it sounded like there were some issues with standardizing the language throughout the measure. Is this red blood cells, platelets, plasma, everything? It sounds like the language changed a little bit during the measure.

Secondly, and this is not necessarily in order of priority. Secondly, including emergency in the numerator versus simply excluding it from the measure was an issue that was brought up.

Next, Feasibility and the burden on the hospital, and there were a lot of

different concerns around the burden on the hospital abstractor or whatever the unit of -- whoever's responsible for doing the measuring, particularly the burden around whether this opens the door to requiring a consent for many things that don't require consent right now, including other things that are considered drugs.

There was a concern about whether this is sort of a one-size-fits-all medicine, or one-size-fits-all measure, meaning that in various locations in the hospital the need for blood is very different. The opportunity to have a discussion with patients is very different.

For example, the trauma bay, which would hopefully be excluded or at least included in the numerator, is very different from say the orthopedic ward, which is very different from say the oncology ward.

Then there were questions, again going back to the burden, in what qualifies as

a consent. Is a simple checkbox adequate? Does there have to be documentation of a conversation? How many consents do we need? What's the time span that consent is good for? That may vary in different hospitals, particularly in the VA hospital versus the rest of -- versus many other systems.

And then, perhaps most importantly, is there any evidence of an impact on practice using this measure? I can't remember who brought that up, but I think it's probably the most important issue that came up.

So we'd like to give J-Co an opportunity to respond. I know that this is a lot of different issues to respond to, but I'm hoping that you can cover them.

DR. GAMMON: Well, on the numerator, we go with a standard with CMS and Joint Commission measures. What we usually do is the denominator is the larger general area.

Once we mention the red blood cells, plasma and platelets, then we don't usually mention it again in the numerator. So that's why there's a little bit of difference, perhaps, between the numerator and denominator, that maybe you're not used to.

We had looked at excluding the patients that had emergency transfusion, but no one could come to consensus about what an emergency transfusion was. So we were going to look more for documentation that the initial transfusion was deemed a medical emergency, because we thought that would be clear in the documentation. So those people wouldn't have to have a signed consent.

As far as the feasibility and the burden, we know each hospital does this a little bit differently, and so we looked at the initial transfusion. I know that you could have a transfusion in one area and then another.

But because of that burden, we

would just look to see if there was a consent for the first initial transfusion, and we also know that, you know, you're going to be talking about different things, risks, benefits and alternatives, depending on your hospital and perhaps on the product that you're going to be giving and given.

As far as the time span, again, we kind of go with whatever the hospital expects, but we're still going to look just for the initial transfusion consent.

As far as the evidence, this measure was not intended to show a difference and a decrease in blood products, but more to have a patient educated and maybe to have the process to make sure that it's being documented and the patient is really receiving the information.

Because it's so important because of the side effects and the morbidity and mortality that can occur. We want them to understand that. So it's more the process and

more patient-centered, getting the patient involved in their care. That's what we feel was the value of this measure.

DR. KLEINPELL: So just to clarify, it's having a signed consent. It's not having a signed consent plus documentation of additional information that was given to the patient about risks and benefits and such, yes?

DR. GAMMON: Well, it is a signed consent. But we were looking more for the information that was there about the risks, benefits and alternatives. It's more like the process that they would have to go through.

DR. KLEINPELL: So how would a hospital identify that that was done, aside from the fact that the consent was signed?

DR. GAMMON: Sometimes at the bottom of the consent, it says -- the doctor will sign that they've had a discussion with the patient, or in the consent itself it will say the patient signs that I've been given

this information.

DR. CIMA: So it has to be a specific blood consent? It can't be like a general surgical consent.

DR. GAMMON: No, it can be whatever, as long as they mention those three things, that the patient is aware of the risks, benefits and alternatives, and no matter -- whatever way the hospital wants to present that. We noticed that a lot of the hospitals have their own separate transfusion consent for that.

DR. WILHOIT: Are the abstraction instructions, do they make that clear anywhere, that it requires the, you know, specific risks and benefits? I couldn't find it in the abstraction tool.

DR. GAMMON: We had the data dictionary and we had a data element that says the information addressed the risks, benefits and alternatives. The abstractors didn't feel this was a burden.

CHAIR MORRIS: Okay. Any other questions around this?

DR. STAFFORD: I have a couple of comments. So again, you're assuming that a piece of paper in the chart. What you're really getting at is patient-centered care, and what you're assuming is that a signed piece of paper, which you're telling us can look like any number of different things, has actually provided the information to the patient, and actually did true informed consent.

So included in a true, informed consent you want a teachback, you want to make sure that the patient understands what you've said to them, that they can repeat back to all the appropriate information, and that they truly understand the risks and benefits of what you're talking about.

If you're really going to go -- I mean, I know it's on the Joint Commission site, and I know about informed consent, and

I think that's -- so I'm afraid that putting this burden on practitioners is not going to get what you're really looking for.

So it's not going to measure what you really want to measure. So to have this measure, and to have it not measure what you're looking for, I don't think, is the appropriate thing to do. And getting at the first unit of blood or the first transfusion; so I practice in a big academic center, I have patents come in and out of my ICU all the time, who have been transferred from the NICU, who come from another service.

Now I'm going to have to go look in the computer, try to find their transfusion record, to figure out if they've been transfused before, because then I don't have to -- or I'm going to have to take the time and consent everybody. So that just adds more burden to the everyday practice of the practitioners.

DR. GAMMON: Could I just say that

we didn't make it specific that the provider had to do it. We know sometimes they have a transfusion specialty officer or someone, an APN, or a physician's assistant could also give this information about the ordering.

Also, the Joint Commission allows hospitals to determine which treatments should have an informed consent. We don't say that blood has to have an informed consent. But we know that most hospitals do have to have informed consent. So we call this transfusion consent, and we looked at that, to make sure that the patient has been instructed and has the information about the consent.

DR. HALPERN: I have a question though then about the emergency. So are you having two separate numerators that you're comparing, or you're adding in the emergency products there?

DR. GAMMON: We either look at the initial transfusion, or we looked to see that the first one was deemed a medical emergency.

DR. HALPERN: As two separate measurements within the same thing?

DR. GAMMON: Well, you open up the chart and you find out whether it was one or the other, and they're treated the same. I mean you're still going to pass the measure if it was deemed a medical emergency, or if they have had this information.

DR. SIPERSTEIN: What's the definition of medical emergency and what documentation would be required to fulfill that?

DR. GAMMON: We would look for documentation, that the blood was given for a medical emergency. I think there's some forms that every hospital has that they have to sign, because they have to --

DR. HALPERN: Not if you're in the middle of a trauma that just came in.

DR. GAMMON: Well, we're not asking to sign it at the beginning. I mean just in the medical record. It could be just

a retrospective review.

DR. CIMA: Didn't you just say that you're not asking for a specific consent? But doesn't this say you have to have a specific consent?

DR. GAMMON: We just have to have documentation of a signed consent, and also that they were given the information, or the first transfusion was a medical emergency.

DR. WILHOIT: I found the abstraction instructions finally. They're on page 98 of the document that we got, and it doesn't say anything about information about risks and benefits.

It does say, as somebody pointed out earlier, it says notes for abstraction for hospitals that use a general consent for treatment that includes transfusion select yes. So a general consent for your admission plus transfusion or your surgery plus transfusion sounds like it counts. There's no requirement that there be risks and benefits,

which means that it's not necessarily getting at the patient-centeredness.

DR. SIPERSTEIN: Yes. After hearing some of the discussion, it doesn't sound like this adds anything. It really doesn't.

CHAIR MORRIS: Okay. So we've heard a lot of different things about this measure. We've said a lot of things about the measure. I think we -- I think we have a lot out there.

We gave J-Co a chance to respond, and then it's time for us to go ahead and vote. So does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: Okay. We had 22 out of 22 who said no, and so we're able to move on. We're going to skip a little bit here. One of our panelists has to leave a little bit early, so we're going to 1532, Dr. Afsar-Manesh is going to talk about Plasma

Transfusion Indication, and then 1539, Platelet Transfusion Indication.

DR. AFSAR-MANESH: Thank you. So the measure number again is 1532. The title is "Plasma Transfusion Indication." This is actually in a series of three transfusion indications that we're going to be reviewing this afternoon. This is the first in the series.

The description of the measure is the percentage of transfused plasma units, with pre-transfusion PTI and all resulting clinical indication documented applicable to inpatients of all ages. Of note, this does have an exclusion for trauma.

So in general, our work group reviewed this and we do recognize that plasma transfusions, the same as the other transfusions, are performed frequently in the inpatient setting, but there is considerable variation in the utilization of this rare resource, and that it is important to

acknowledge and improve our utilization of this resource.

However, there are a number of concerns that were brought up that I would like to share with you. When it came to the Importance to Measure and Report this, again, there hasn't been clear indication that putting the indication or doing the INR is I fact going to improve your quality outcomes or decrease your utilization.

So there was some concern about what this would translate to as far as improved quality outcomes.

We had some concerns about the Scientific Acceptability. The exclusions, we felt, needed to be broadened. So for example, when you have active bleeding, or in some cases again when you do ECMO in emergencies we talked about. Again, there are a number of different cases where you would need to do the transfusion. Again, we would need to have that be in the exclusion criteria.

The Usability, there weren't much concerns in that area. But then for Feasibility, again, we had a number of different concerns that I'll share with you. Again, one of the major weaknesses that we saw was that there are currently not any clear guidelines or indications for transfusion.

So therefore telling providers that you need to put the indication for the transfusion is going to again lead to even greater variability, on top of which because we don't know what we're looking for, we could have people just put whatever indication. It's not really going to change the outcome and get us the quality improvement that we want.

Another concern that we had was there could be some concerns with cost of implementing this if the institution doesn't have electronic health records. Again, you'd have to go in and abstract the PT INR and also the consent form.

There is a tool that the store at the Joint Commission has, which is similar to the tool or is the tool that we just spoke of, and there is definitely a web-based component to that.

But from our understanding, you still needed to have vendors and abstractors obtain that information. So again, as we're adding a couple of these measures, we had a little bit of a concern about the feasibility of getting that.

Another concern that was brought up is that the PT and INR, if they have to be drawn, there's not really a clear indication of how long before the transfusion they have to be drawn. So we wanted to have that clarified.

There was a pilot that was done for validity and their reliability testing. That's on page eight of the PDF document, in that it says that the measure-specific issues were revised to strengthen and provide

additional clarity for the data elements, but it doesn't speak to exactly what was clarified.

Again, we'd like to know what were some of the barriers and challenges in abstracting this in the pilot phase that we should be aware of. Even though that apparently has been clarified, it can shed some light into some of the obstacles that could be presented for institutions as they try and obtain this information.

Then lastly, from our understanding, the measure addresses the first three events, which I'm assuming is the first three transfusions, the first three times that you get transfusions, and each time it addresses the first three transfusions. So we were wondering kind of where the three and three came -- where that number was derived from, and what was kind of the reasoning behind that. And that's it as far as our concerns.

CHAIR MORRIS: Anything else anybody wants to add to this one?

DR. DUTTON: Yes. As somebody who does this a lot, transfuse people, the indications for plasma transfusions are changing very rapidly right now, and I would argue that it's almost to the point where the only clinical indication for transfusing plasma platelets is bleeding, and that any kind of prophylactic transfusion is almost off the books now.

So this is going to be a hard to define the indications part of the numerator very clearly, because the science is moving. The other thing I'll point out is this is referenced to PT and INR, but it would be perfectly reasonable to transfuse plasma on the basis of the TEG, for example, or other tests of coagulation.

DR. DILLON: In addition, is there any consideration in terms of point of care testing versus lab testing?

MS. ZAMBRICKI: I wanted to make a comment again about feasibility and burden. If I understand this measure correctly, it calls for both. It calls for reporting of the value, and then a statement about why you are treating it. It's easy that is a burden, that if the value is tremendously extended just writing value transfused is sufficient.

You do not have to have the provider then write some statement like to improve or treat coagulopathy or improve bleeding time or something like that. So if it stays, I would say "or," "one or the other" would be adequate.

DR. WILHOIT: One question that I had was about the accuracy of abstraction, and I wasn't quite sure what it meant. But under the testing results, which are on page 8(2)(b)(3), for example it said the originally, when the cases were abstracted, there was a rate of 78 percent. The reabstracted rate was 70 percent.

I'm not sure. I assume that was sort of tests, you know, two different sets of people doing the abstractions to measure accuracy. But I noticed on all of these measures, there was a big difference between the two, which also then raises the question of is it a measure that can be accurately abstracted.

DR. HALPERN: I also agreed with your statement that the exclusions have to be broadened, because you're not going to take time when you have somebody who's exsanguinating in the OR necessarily to either draw a lab or document why you're doing it.

DR. STAFFORD: I would agree with that. I mean in our massive transfusion protocol, it's almost one to one now, which is what's come out of most of Iraq and Afghanistan, in terms of for trauma or for massive bleeding. You don't wait for an INR, and by the time you get an INR, it potentially could be normal.

So you could get dinged for having given that, and so I think that's an issue. Actually, I think it is important to measure all of these. I would, I mean I sit on a P&T Committee at our institution, and I know we look very closely at blood usage and plasma usage, and have done a number of interventions with certain providers, because of how they were using it.

I know that most of the forms that when we sign the orders, we have to give an indication already. So I'd be curious, at least for those who are here, if they know at their institutions, do you already have to do this? Because my suspicion is is that in a lot of large institutions, at least, you do.

Now it doesn't get to the smaller ones who may not monitor it, but is this something that's best left at the individual institution level?

DR. DILLON: I think the other area we have to be careful of is the age, this

sort of all-inclusive age range, because I'm not sure that the indications for the management of neonatal sepsis in a coagulopathy will, you know, be the same as what's now going on with the adult system. So I think putting it to all ages is of serious concern on my part.

DR. HALPERN: I would say that, in addressing the prior comment, we do have a transfusion committee that carefully monitors what we do. But we do not have any statement on the transfusion form saying exactly why you're doing it.

CHAIR MORRIS: Okay. So just to recap, again, numerous issues or questions with this measure. Please let me know if you feel like I haven't adequately covered the particular issues that you're concerned about.

First of all, indications for transfusion are changing. It's a moving target, and so this makes it very hard to determine precisely what the indications

should be.

In addition, tests besides PT or INR may be equally appropriate for obtaining before doing a transfusion, or it may be appropriate not to do any tests at all, because of the patient need.

Secondly, there were issues around the feasibility and the lack of clear guidelines, which really refers back to the first one. So it's not clear that reporting this indication is related to the desired outcome.

There is a cost, concerns about the costs of implementing this measure, and unclear parameters, exactly how this should be reviewed by hospitals or measured by hospitals. Unclear where the data was derived regarding the fact that this is supposed to be the first of three transfusions.

A lot of concerns about the indications, and concerns about the accuracy of this measure. Also concerns that there

would be -- that we don't know what the barriers were to abstracting these data in the pilot phase, and it would be very helpful in terms of understanding of how it impacts hospitals, to understand what the barriers that were in the hospital phase, in the -- I'm sorry, pilot phase.

Then lastly, there were concerns about the exclusions, and in some ways that the exclusions should be broadened, but also that the exclusions in other ways were too broad. For example, the all-inclusive age range was a concern.

So it was felt that this was both not adequately sensitive and also not adequately specific. Would JCAHO like to respond to those, and does anybody have anything else that they want to bring up for JCAHO to respond to right now? Anything succinct that they would like to bring up? Okay.

DR. GAMMON: Okay. Well, the

blood bankers and the panel felt that as much as possible, you should do an INR before each time that you gave plasma. We do allow a TAG as well as an INR. It was changed after the pilot. The indications for this one was actively bleeding, and we do realize that if someone's having a massive transfusion, and sometimes when you're actively bleeding, you're not going to get a pre-transfusion lab for that.

So those patients would pass without that. We did the first three transfusions because of the abstraction burden. We know that patients receive more plasma than just three, but until we can get to the electronic phase, we just -- we're going to look at three.

The concern of the panel was also that a lot of patients are getting these for procedures when the INR is not very high. So you know, they weren't bleeding. So that was their concern, that they would be looking at

this, and then sometimes when you're looking at a surgical record, you can't even tell if plasma's being given and for what reason.

CHAIR MORRIS: I think that there was agreement among many of the group that this is actually conceptually very important. We're just not sure that this -- so I think as a group we really agree with JCAHO, that this is an important thing. Just not sure that this measure really captures what it is that we want to capture. Can you speak to that a little bit more?

DR. AFSAR-MANESH: Absolutely. In the small group discussion that looked at a series of transfusions, I think there was overall agreement that there are variations in utilization, and there is some data for that that we could, and some hospitals have committees.

But in general nationally, we could look at our utilization and decrease that or make sure that we're appropriately

using this rare resource. I don't think it's a matter of the importance of it. It's just exactly as you mentioned, which is this is not the best way at capturing that, because of those mentioned areas.

But I think once those are addressed, that this could potentially be something that could be reevaluated.

DR. WILHOIT: Other things in the abstraction details. It indicates that the lab value being looked for is the most recent one. But it could be a day before, a week before, a month before, a year before. There's no time constraints on what most recent is. That seems like a significant issue as well.

It also goes up to the third unit given, and doesn't say that you have to have rechecked after previous units, or that you should or that you shouldn't. Again, there's nothing about that.

DR. HALPERN: But along the lines

of a lab test, you could have had a completely normal INR preop, and then you get into unexpected bleeding, and you end up having to give a transfusion. You don't have time, anesthesia may not have time to write down why you did it.

DR. SIPERSTEIN: To that point, it looks like this measure was intended for the patient who has a high PT on coumadin, for example, who's scheduled to undergo a semi-elective procedure, and just to make sure that we're not willy-nilly giving excess units of fresh frozen to reverse that.

But it really does not apply at all to the intraoperative massive bleeding patient. It really is a very different clinical situation, where we use clinical parameters, not laboratory parameters to make that decision. So I think, you know, if this measure were really restricted to that initial group, then it makes some clinical sense in terms of documentation.

But the latter group is thrown into that really, in my opinion, kind of makes it clinically irrelevant.

CHAIR MORRIS: I think that summarizes it pretty nicely. So I would say that in our first vote around the Importance to Measure and Report, if this is voted down as inadequately important, it doesn't necessarily mean that we think that the concept is unimportant.

But just potentially the impact of this particular measure, or the outcome or evidence around this particular measure. Anybody else want to say anything before we move on to a vote?

(No response.)

CHAIR MORRIS: Okay. So the first vote, does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: Summary of responses, 2 for yes, 20 for no. We'll move

on to the next measure, which is also slightly out of order. It's 1539, Platelet Transfusion Indication.

DR. AFSAR-MANESH: Perfect. So the description of this measure is the percentage of transfused platelet doses with pre-transfusion platelet count results, and clinical indication document applicable to inpatients of all ages.

Again, not to sound redundant, I'll just summarize it very briefly. We realize that this is a rare resource that the utilization of it should be done thoughtfully. But again, reporting it as far as looking at your pre-transfusion platelet counts and indication outlines all the same problems that we highlighted in the previous measure, 1532.

I'd be happy to review all those again, but if no one has any questions about them, we can open it up to discussion, if anyone feels particularly different about platelets.

CHAIR MORRIS: Anything that anybody wants to add with regard to platelets, compared to plasma previously?

DR. CIMA: Just to clarify for the Joint Commission what my view on this is, that this is even more difficult than the other one, because if they're on a medication that inhibits platelet function, they may have a congenital abnormality that inhibits platelet function.

You can have too many platelets and still have platelet dysfunction. I mean azotemia, yes uremia, any of those things. This is really even more difficult than the plasma one. I think these are all very good, but perhaps a different avenue would be to say you have to have a blood utilization review committee in your hospital, as opposed to trying to do it piecemeal, you know, as opposed to this.

CHAIR MORRIS: That's a good suggestion. Would JCAHO like to respond to

this at all, or shall we move on to the vote?

DR. GAMMON: I think it's about the same as the last one. I mean it's based on the same.

CHAIR MORRIS: Okay. So does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: We had 22 out of 22 saying no, and again, I just want to stress that this does not mean that we think that the concept is not important, but just that the measure needs a little retooling. Now we're going to go back --

DR. ROGERS: Arden, can I just make one comment? I want to follow up on what Dr. Cima said, because I think it's really important, and Dr. Morton mentioned it earlier.

Receiving blood products is in fact a transplant of some type. It is not just a thing, like can you get this pill, I'm

going to get this, whatever. I think we in general as a profession have been very sluggish to recognize the seriousness and the importance of this whole environment.

I think that the NQF could do something pretty powerful actually, and make not only a suggestion, but lead the conversation to the recognition that a normal process in a hospital would have a blood utilization panel. They should be specialists who are ordering this, not just anybody who has hospital privileges.

So that's just a little bit of an editorial, but I think it's something that we -- that's a message I'd like to get to the NQF and perhaps we can help that along.

CHAIR MORRIS: Potentially similar to hospital restraints now. You need a lot of levels of approval before a patient is restrained for any particular period of time, and maybe something along those lines would be appropriate for blood transfusion as well. So

we will take note of that. Any other comments before we move on to 1527?

(No response.)

CHAIR MORRIS: Okay. 1527, Red Blood Cell Transfusion Indication, Dr. Carpenter.

DR. CARPENTER: So this is the third of the grouping. This should be more straightforward, because red blood cell transfusion is a little more straightforward, but shares some of the same problems.

It is also a newly-proposed measure. It has the two parts as the other ones do, a measurement part, which is hemoglobin hematocrit before the transfusion, and a documentation of an indication in the chart.

Both conditions need to be met to satisfy this, this criteria. The discussion that we had as a group and I think it's similar to what we've had here is that this is an opportunity for improvement. This is an

important area that we have. There is an opportunity to decrease the risk and the expense of unnecessary transfusions, so that this is an important area.

However, the sort of on the Scientific Acceptability part of it, shared some of the similar problems with the other measures. The timing of the laboratory measurements relative to the various transfusions, what a documented indication was.

Was it a documented lab value? Is it wording of documentation? Is that really what is standardly charted now, or is that going to be a burden. In addition, the same criteria.

Exclusions were really not well thought-through. Intraoperative use, use around trauma, use around dialysis, use with active bleeding. That all needs a lot of work to figure out what the exclusions are there.

My reading of their reliability

and abstraction of this was really quite poor, with a match rate of 60 percent, I think, when they went back to re-extract. I think it's around what's an indication and what's not an indication in the chart. So the use of this wasn't very reliable.

In addition, it does require quite a bit of work in chart abstractions, since many of these things are not captured electronically. They're not routinely charted. So the abstraction of this from the charts seem to be quite difficult.

The rest of the conversation we've had around these measures, to some degree.

CHAIR MORRIS: Anything else? Do you have something?

DR. DUTTON: Sure. At the University of Maryland, about half of all red cell units are given in the operating room, and about half of those are given in the trauma center. My practice gives 600 units a year of uncross-matched blood. Not having a

trauma exclusion here is insane. I mean we teach, it's in the textbooks, that you should not be waiting for hematocrit to give blood to somebody who needs it, somebody who's truly exsanguinating, and to not exclude emergency situations here, I think, is a big mistake.

CHAIR MORRIS: Anything else anybody has to add?

(No response.)

CHAIR MORRIS: Okay. Would JCAHO like to respond to the issues that have been brought up for this particular measure?

DR. GAMMON: Well, this is, you know, the closest pre-transfusion value was not a problem for any of the hospitals to collect. We originally had for the test it had to be within 24 hours.

And then for some chronic patients, you know, chronic blood use patients, they thought it could be expanded a little bit to 48 hours. But there was usually a pre-transfusion lab that the abstractors

didn't have a problem with.

The other thing on the match rate about 60 percent, that was only one of the data elements for red blood cells, and it was a little bit different, because every hospital does have a different way of if they do document, of what they use for the documentation criteria.

Even though the abstraction burden has been mentioned for this one and for others, we are moving toward electronic health records. The values are readily available there, as well as the trauma codes.

It was never the intent for someone who's having a massive transfusion to be getting a pre-transfusion lab value before each one, and that would be excluded as well as uncrossed blood units that were being transfused.

Also, some of the units that we've been finding that are used to prime pumps, ECMO machines and also some of the bypass

machines that are getting that as well, so --

CHAIR MORRIS: Let me just ask a question. As I read the numerator and denominator and exclusions, it looks a little bit different from the way that you're presenting it right now. Can you explain that?

DR. GAMMON: Yes. We had a panel meeting just in November, and we're revising the specifications, and we've been looking. We had that new data element called "red blood cell unit exclusions."

It was supposed to have been brought over for number two, as well as number five, because we've excluded patients from five for the massive transfusion uncross-matched blood, and those are the units that prime flow on pumps. It just didn't get brought over for that, red blood cells.

DR. CARPENTER: Can I say something? I think, you know, a lot of people spend a lot of time reviewing these, and then

we get here and then they've been changed already, because they weren't maybe fully vetted or really gone through before they were proposed, it seems.

So what we spend time on, what we evaluate, what we discussed is not what you say the current measure is. I think that's discouraging for the group here, to hear that maybe it wasn't fully vetted before it was proposed, and now we're going to vote on something that you say isn't even the current proposed measure.

So I don't know that that can be fixed at this point, but I think it should be understood by your group that a lot of work goes into evaluating what's proposed, and if it's proposed before it's really ready for proposal, then it just slows that whole process down.

DR. GAMMON: The data element is in your packet, the red blood cell unit exclusions. It's just that at the top it says

it's applicable for five and for two, and the two didn't get put over to the exclusions for the measure. That's all, which is unfortunate.

DR. DUTTON: I'll throw a little more data on the table about the burden of abstraction here. I don't, you know, I spend a lot of time for research purposes trying to look at transfusion from anesthesia records. So that's half the transfusions in the hospital potentially.

It's not easy to get those off of paper anesthesia records, which are right now today 85 percent of the universe.

CHAIR MORRIS: Okay. Melinda, you just clarified something for me that might be good for the group to hear as well. But we're voting on the measure as it's written; correct?

MS. MURPHY: That's correct. The only that we would consider something else is again the same thing we've talked about all

day. You'd vote it down and then you would suggest whatever changes, or we would hear what we've just heard, that changes have been made and we need to see all of those changes for a revote.

CHAIR MORRIS: Okay. I think in addition to that was it's not unusual for measure developers to continue to try to make their measures better over time. But I really have to agree with Dr. Carpenter, that if you think about the number of hours that we spend and our time is pretty valuable, that it is a little bit discouraging, even if that's normal. We'll be voting on the measure as it's written.

So let's go ahead and proceed to a vote, unless anybody has anything else to say about this measure. Does the measure meet NQF criteria for Importance to Measure and Report, as it's written?

[COMMITTEE VOTING.]

CHAIR MORRIS: Okay. 2 for yes,

19 for no. We'll move on to the next measure, which is also Dr. Carpenter, 1541, Blood Administration Documentation.

DR. CARPENTER: This is another Joint Commission-proposed new measure that is in the family of the blood management project. This is looking at when a blood product is administered, documenting three things that happen during that process, and each one of these needs to be met to pass this criteria.

First is an identification process matching the unit that's been prepared to the patient. The second is dating and timing of the transfusion, and the third is measuring vitals pre-transfusion, during transfusion and post-transfusion. So all those items need to be met to satisfy this criteria.

In discussion and reviewing this, I think all the group thought all three of those were really important things to do, critical things to do. What wasn't clear is how big a problem this is for the hospitals

right now. These are standard protocols that hospitals follow with essentially all transfusions, except in potentially life-threatening situations.

So what is the gap? What's the opportunity for improvement? It wasn't clear from the documentation. Although there certainly were references to this continuing to be an occasional problem, it wasn't put in perspective with how many transfusions are given and the relative risk of this.

Most of the risk seems to occur in the identification process. We felt, without doing a complete review, that in most hospitals this is standardly required process. So that was a first question, is how big is the gap? How big is the opportunity for this measure, even though we all thought it was, these were important things to do?

In terms of the acceptability, it's a lot of information to get out of a medical record. So the other part of concern

was even if this was all done, is this all documented? Is this all something that Can be achieved out of the medical record consistently?

We had a similar concerns about exclusions as we did in the previous blood product use measures that we discussed. The intraop measurements, I think, proved to be quite difficult for this, as you might imagine, especially if it's not electronic system.

So abstraction, as far as I could tell from their system, seemed to be challenging and difficult, because of the number of data elements required to satisfy this one criteria.

So that was basically the summary, is that is this really an opportunity for improvement, and is -- there's a significant burden to collect this data from the chart.

CHAIR MORRIS: So just to reiterate, there was a strong sense among the

group that this measure is valuable that topped out. Can you describe what happens at your institution?

DR. DUTTON: Well, I'll point out we also give uncross-matched plasma, universal donor plasma as well. So that's a minor change to this that would be needed.

It is very important, obviously, to make sure you identify the right patient to get the blood, and Dr. Carpenter is correct, in that a handful of patients every year are killed outright in the United States by ABO mismatch, transfusion mismatch, and probably the Joint Commission actually knows the number, since they're all reported as sentinel events.

All of the things that are asked for here, these three sets of requirements, are all sort of standard operating procedure for any hospital I've ever been in, and/or audited in those hospitals as well. I think that's because they're ADD standards to begin

with, and Joint Commission standards for practice anyway.

If this is something that the hospital's doing anyway and reporting anyway, and this is just a matter of rolling that up and do a national indicator, I guess the burden of doing so would be less, since you're already gathering all the data necessary.

On the other hand, I'm not sure what value it adds, putting this measure on the table, if it's something that everybody's already doing.

DR. HALPERN: I see that they have a statement here that the frequency is 1 in 1,000 events. That's in 1(b), number one. But my question would be how often -- of the sentinel events that happen, what has been the common root cause problem? Is it a misidentification of the patient, or a miscross-matched unit?

DR. WILHOIT: The data provided in 2(b)(3) says that the -- for the 274 units

that they reviewed, the rate was 89.4 percent. That's a ten percent deficit from what we would expect. So it sounds like it's a real issue. It sounds like it's an important issue to address if we're only scoring 90 percent, you know, and the hospitals that usually do testing are usually, you know, better --

DR. HALPERN: Are they including emergencies?

DR. WILHOIT: But unless it's an exclusion issue. However, the thing that's really problematic for me is that they reabstracted this and got a rate of 67 percent.

Well, if it's a 20 percentage-point difference from the first abstraction to the second, which I think is what this means, then it sounds like the methodology for abstracting hasn't been well enough defined to have clean data. So I'm not sure what we're left with.

DR. CARPENTER: You know, I

interpreted that rate of 80-some percent as something was missing in the documentation. Not necessarily that it wasn't done, but it was something was missing, a post-vital or something that wasn't specifically documented.

Because clearly this is being done in our hospitals at a higher rate than listed there. It's just not documented or abstractable consistently.

CHAIR MORRIS: Any other issues? Okay. Would JCAHO like to respond to the question about whether there's actually evidence of a gap, and questions regarding whether or not the abstraction process is adequately defined, given the large discrepancy in the first and second data retrievals?

DR. GAMMON: Well here in the United States, we are just now beginning to collect hemovigilance data on adverse events. But if you look at data from UK hemovigilance system, they've been looking at data for the

last 15 years, and they've had more deaths from bacterial contamination of the wrong blood and also from the administration of the wrong blood more than they've had for HIV infections.

So and patient identification is just -- I know it's a standard of care, and we hope it's being done. What we found out that a lot of times the identification is there, although our rate was, you know, not like 90 percent when we looked at everyone's. It was different infill that was missing, as Dr. Carpenter had mentioned.

You know, we have people that collect core measures, and they're very used to collecting data elements like these, and the abstraction burden was not that much for them. I think it was more for the people that were doing it for testing alone, and that's not usually their main job.

I think the difference, and like I said, the difference in the rate between the

original abstracted and the reabstraction was for the missing documentation. A lot of times, if someone's in surgery, you can't really tell when everything was given. I mean it's just very illegible.

But the main thing about this data element is that if our hospitals are going to participate with the hemovigilance, you do need this data. This is the exact data that's needed to participate with them, so they can figure out if there is going to be an adverse event, they have to have this information. So we've aligned with them, and they just hope hospitals will begin reporting on those things.

CHAIR MORRIS: Anybody have any other questions or comments?

DR. STAFFORD: I just have two comments. I would A, be somewhat reticent to use data from other countries or cultures and trying to extrapolate it to the U.S. health care system.

So the UK is very different than it is here. I guess the other point that maybe I would like to make is, as we've all said, I think we all think all of these measures actually are probably important somehow, and it's more how they come about.

My proposal would be that this might be something to turn into a patient safety goal, where you might say, you know, with the goal being to reduce X, Y and Z related to inappropriate or, you know, transfusion of blood products, and that then you require hospitals to put in place some method of monitoring that.

Just like you've done for rapid response teams. Because similarly, when there is a lot of talk about National Patient Safety goals and rapid response teams, the original proposal, proposed goal was to actually say you had to have a rapid response team.

When the measure finally came out, it was worded differently and was more broad,

which allowed institutions to get to the same end point, but using different methods that actually fit their institution and their culture and their resources, because not every institution has the same resources to do this.

CHAIR MORRIS: Any other comments? Do the developers want to say anything else about the measure? Okay. Let's go ahead and vote. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: We have 8 who said yes, 13 who said no. So the noes outweigh the yeses, although this is a little bit closer than some of our previous measures. So I think that we're not going to go ahead and continue with the vote on this particular measure.

Next is 1542, Mr. Rivenburgh, Preoperative Anemia Screening.

MR. RIVENBURGH: Measure 1542, Preoperative Anemia Screening, the description

of which is a percentage of selected orthopedic, cardiac and hysterectomy elective surgical patients with documentation of preoperative anemia screening 14 to 45 days prior to the anesthesia start time.

The numerator for this were patients with documentation of preoperative anemia screening 14 to 30 days before the anesthesia start time. The denominator were selected elective surgery patients that fell under these criteria.

The exclusions were patients not admitted from home, and this is one of the areas where we had some concerns as to what were the differences and levels of anemia that can be out there and the testing sources at the time of the testing, as much as 45 days in advance.

The time frame is particular, and it's very -- is it clinically relative, in the sense that if a patient has their anemia tested 45 days in advance before a

hysterectomy-type procedure, and they happen to be bleeding extensively and then that bleeding stops and their hematocrit and hemoglobin then normalized.

Some adjustments. There were no risk adjustments necessary, and again we were concerned about all of the exclusions that were listed.

CHAIR MORRIS: Anybody else have anything to add to that?

MS. ZAMBRICKI: Just a point of clarification. The measure description says "45 days before anesthesia start date," and the numerator inclusion said 30 days before. And then there was a difference in the age. The description of measure says "Equals 18 years," and the numerator said greater than 18 years. So just some standardization of the terms.

CHAIR MORRIS: So we're looking for clarity on standardization of the terms, 14 to 45 days before the anesthesia start

date, or 14 to 30 days before, and then particularly with regard to the age. That probably is a simple typo, but we're looking for clarification of it.

Then secondly, we're concerned about the timing of the anemia testing. I Can tell you my particular, one concern I have about this is seeing somebody in clinic for the first time say one week, having an opening in my operative schedule and it's clear that they need to have an operation.

So I get them into the operating room within a week. Well, they're not inside of that 14 to 45 day window. How does this play out? Not an emergency, but somebody who does need an operation.

That was one of the things that stood out for me immediately. Does anybody have any other issues with regard to this measure?

DR. CIMA: Does it have to be -- I mean is it specifically? So what if it's the

day before?

DR. STAFFORD: I understood that to be an exclusion, that they look at the record, and if the patient was scheduled for surgery less than 14 days, then that is an exclusion. That's what I read in here.

DR. HALPERN: They're trying to get -- it sounds to me like they're trying to get rid of the sort of more emergent cases, and this was the easiest way to do it on a global level, without increasing the burden of data extraction.

DR. STAFFORD: The baseline rate was in the 30's. How important is this to those of you who are surgeons and doing these kinds of cases? Is this something that's clinically important or something that's not clinically important? It sounds like it's not being consistently done.

DR. MORTON: I think age matters a lot. You know, that's why they have the exclusions, I guess, in there. The thing that

pops into my mind, we recently took a look at what we're doing preop with all of our patients, and there's a lot of labs that are being done that aren't always necessary.

So we looked at this because of the cost that's involved with it, and there's not a lot of data to support that the screening actually makes a big difference. I'd be happy to hear if there are some data about it.

DR. HALPERN: Actually, the New England Consortium for Vascular Surgery, anemia is one of the predictors of mortality for lower extremity bypass.

DR. MORTON: I know it's a predicter, but you know, the ability to act on and do something else.

DR. SIPERSTEIN: Different cases have very different risk factors for needing transfusion. So there are many types of surgery where transfusion is almost never done, and therefore you know, the preoperative

CDC, if it picks up something, you know, yes, you act on it.

But doing any formal anemia screening would be, you know, not that clinically effective in that group. Whereas you've got other groups of patients like undergoing hip replacements, that you know, the risk-benefit equation may change.

CHAIR MORRIS: The three types of operations in which this would be done would be orthopedic, cardiac and hysterectomy.

MS. ZAMBRICKI: So would that be like a carpal tunnel would be included in that?

DR. CARPENTER: I think they have, I didn't look at it right this minute, but they had an appropriate list, total knee, total hip mostly. Maybe there was some spine in that. But it was mostly appropriate, although many of the -- most groups have backed off on the number of cases that need preoperative hematocrit evaluation.

Certainly older people getting total hips and total knees need it done, but you know, a healthy person with a tibia fracture doesn't need it done. They're not going to lose a significant amount of blood and they're hematocrit's unlikely to be a problem if they're otherwise healthy.

DR. STAFFORD: Yes. I mean I think getting back to kind of what, I think, and certainly you can correct me if I'm wrong. You're really trying to get at those elective cases, because we do know that being anemic is a risk factor. It's a risk factor for actually getting more blood, and it's a risk factor for infections and also mortality and morbidity.

So I think what you're actually getting at is those truly elective cases where you could do something. So somebody comes in, they're anemic. You may have -- if you have six weeks to work up their anemia, to put them on erythropoietin, and perhaps even let them

then donate their own blood, so that they could have it for the operation, that's appropriate.

So I think there's probably a small subset of patients where this could be used, and along those lines, getting back to the exclusion for not coming in from home, well, the large number of those patients who would fit in this population are going to be in nursing homes or assisted living facilities. So you may be excluding a patient population that you actually want to benefit.

DR. DUTTON: There is a strong association between anemia and bad outcomes, no question.

But that may be because sicker patients do or some of the same things that's making the patient anemic, malnutrition, age, cancer, whatever, is increasing their surgical risk. Is there any evidence at all that addressing it makes any difference in outcomes?

DR. STAFFORD: And I think that's what we don't always know, and I guess the only way that I could see a definite outcome would be if it keeps you, if you can build up their blood stores, give them, you know, have blood available so that if they need to be transfused, they get their own blood.

That would for me seem to be the one place where you could really make a difference.

DR. CARPENTER: Not that this directly addresses that, but that used to be very common. We used to give a lot of auto-blood or direct to donors.

But that's become increasingly unpopular, and many people think it's less safe than giving anonymous blood. So I don't -- it used to be done commonly for elective orthopedic procedures, and itþs not done very often anymore.

DR. DUTTON: Yes, autologous blood is fine, your own. But directed donation

blood, yes, is more dangerous than random donor blood.

CHAIR MORRIS: Okay. So the issues that were brought up were timing of anemia testing, concerns regarding the exclusions, and help me out here. Other issues? The value, the value of this measure. The evidence. Would JCAHO like to respond to these?

DR. GAMMON: Sure. The time line is from 14 to 45 days. There's the -- NATA has developed some guidelines, and that's -- they have said they should have it by 30 days, and a lot of that is to do with the Medicare refunding. It has to be within 30 days in order to be able to qualify.

Also feel it takes that long to actually treat these people with some of the medications and to build them up to treat, to manage them, to detect it. We looked at just high blood use, like the elective surgeries that have that much time.

There's been quite a few studies that show that if you can bring up their hemoglobins, that they won't need the blood during the surgery. So this would be very important.

We're having a lot of patients that are older, that are going to be getting hips and knees done, and this could really decrease the blood use and decrease the outcomes, you know, improve their outcomes if they don't have to add blood.

A couple of hospitals that did this in the pilot had nothing but -- they're using less blood, they've had great results, and it's actually going over to other patient groups, because it's been so successful.

DR. SEARS: Have you documented -- I'm sorry, sorry.

DR. HALPERN: We might be asking the same question, but is there actually evidence that giving erythropoietin pre-op and building up their crit reduces mortality

afterwards? Because erythropoietin itself has risks.

DR. GAMMON: I know.

DR. SEARS: Giving hemoglobin and giving iron is the same thing. It takes a while to build it back up.

CHAIR MORRIS: Can you be a little bit more specific about the data that was collected in the pilot study?

DR. GAMMON: We looked at patients that had elective surgeries, hysterectomies and cardiac, and we looked at what the date was for the pre-anemia screening. And then we looked at what was their anesthesia start time and the date.

Then we looked to see how many days difference there was, and noticed a lot of people fell out because they weren't having it within that time frame. A lot of them were having it a lot closer to surgery, when there wasn't enough time to do much about it,.

We also needed to make sure that

we were collecting the data on people that weren't scheduled in less than 14 days. That's why we made these --

CHAIR MORRIS: My question is you said that in your pilot study that hospitals used this and wound up giving less blood. Can you say how many hospitals used this measure and how much less blood they gave, and how that related to the amount of blood that they might normally be expected to give?

DR. GAMMON: I'm not saying --

CHAIR MORRIS: Was there a statistically significant difference?

DR. GAMMON: I'm not saying they used our measure to do that. But when we went to their hospital, they had already been doing this, the pre-anemia screening, and they had been using less blood as a result of it. So they were very supportive of this measure, is what I was saying.

DR. HALPERN: Did they give you data to show that that happened and that it

actually affected the patient outcomes?

DR. GAMMON: I don't know if they published it or not.

DR. CIMA: The data that was put in here for demonstrating performance gaps, this is one of the things that bothered me about it, was high outliers for SSI tended to be hospitals that had, you know, patients that had more anemia.

But we don't talk about intervention; we don't talk about comorbidities, you know. If this is what the data is that's supporting it, it doesn't, it's not tied directly to what you're asking. You're just -- there's an association between people with anemia and bad outcomes, but checking that beforehand or -- are those modifiable risk factors that you necessarily want to do?

If someone has a bleeding tumor, a tumor, ovarian, an endometrial cancer that's bleeding, the treatment is to get it out. So

waiting, I'm not sure the data here, at least what you've used as performance gaps, don't correlate with what you're talking about.

DR. GAMMON: Well, we're looking at the elective surgery patients, though, that have that opportunity. I mean if someone's bleeding, you're not going to --

DR. CIMA: Well, those patients have an opportunity. The question is, is it right to wait, and will you have an improved outcome?

CHAIR MORRIS: Let me just clarify what I was asking about. I'm getting a sense from you that even in a pilot study, you don't have to have fabulous data and it doesn't have to be statistically significant necessarily. But I'm just trying to understand. I'm getting the sense from you that the hospitals that participated in the pilot study thought that they had a better outcome.

But I guess that is very, very qualitative, and I'm looking for something

that might be a little bit more quantitative. It sounds like maybe perhaps you don't have that data. That's okay. I just want to know what it is if you have it.

DR. CIMA: But I guess one of my questions too is I mean I think we've been talking about two related outcomes. Number one, trying to transfuse less units of blood, and number two, trying to reduce morbidity and mortality.

CHAIR MORRIS: And the first is probably a more short-term thing that would potentially be easier to measure. But if we don't have the numbers, then we don't have them.

DR. GAMMON: I don't have a specific number of a hospital that participated with us that I can have -- direct you to. I just have the aggregate number of all the hospitals that participated, and that is anecdotal. I don't know if they have a published study that they were doing this

process and they've been able to save on blood.

CHAIR MORRIS: Any other issues anybody wants to bring up?

(No response.)

CHAIR MORRIS: Okay. I think it's time to vote. Does the measure meet NQF criteria for Importance to Measure and Report, specifically impacts, evidence of a performance gap and outcome or evidence?

[COMMITTEE VOTING.]

CHAIR MORRIS: We have 3 that say yes, the measure meets criteria, and 18 that says no, the measure does not meet criteria. You received kind of a long list of issues regarding the measure.

I think once again that we all agree this is actually important. This is very important in concept, and that we're -- I believe that many of us are hopeful that a more defined measure may come forth in the future, that will help to address the concept.

So next is 1547, and this is our last measure for today, Preoperative Blood Type Testing and Antibody Screening. Again, Dr. Carpenter.

DR. CARPENTER: So this is a sister measure to the one we just discussed. It's the same group of patients, so it's patients who are at higher risk for getting blood transfusions during their hospitalization, who are admitted for elective surgical procedures, certain cardiac procedures, certain orthopedic procedures and hysterectomy.

It's simply a measure about whether their preoperative type and screen or type and cross-match was completed prior to the starting of surgery. So it's everybody that was presented with those conditions, and whether they had this completed at the time of surgery.

The justification for this is really that this is important to do, and that

doing this at the last minute might make it so that appropriate blood was not available in a timely fashion for someone that needed it. The proposal sort of realized that there were very few studies documented that this was a problem. The studies about this really were minimal.

There's one study that found seven percent that it was not completed before surgery. But overall, there's very few studies saying what the magnitude of this problem is, and then if it's not done, what are the implications for complications or mortality and morbidity following that?

So that was one thing that the group discussed, that this is mostly standard practice. It does not, it certainly does not happen every single time. Sometimes the ball gets dropped and it is not completed, and it's sometimes done on the day of surgery, which is a potential problem, especially for the first case of the day. Not quite as big a problem

later, because it usually can be done by the time surgery starts.

I thought that and we thought that the cases were acceptable. That list was, we've already discussed that as a reasonable list. The rate, I believe that they measured, was compliance rate with this measure was quite high, about 98 percent or so was, it was being completed. It was not 100 percent. So maybe the gap is not huge, but it is there.

It's probably important to have this done for cases that have a high risk for needing transfusions. Most of these elective cases don't need blood intraoperative. Maybe some of the cardiac ones, but certainly not the orthopedic ones.

So I think the biggest discussion was around how big is the gap, and is this really an opportunity for improving patient care or not, because what is the consequence of getting this done on the day of surgery rather than before surgery. But otherwise, it

was pretty -- more straightforward, I think, than the other ones.

CHAIR MORRIS: I'd like to make a comment about this, in terms of how big the gap is. Being done 98 percent of the time sounds like a lot to me, and that sounds actually commendable.

In my hospital, I don't think that it happens that commonly, because we frequently have first-case delays because there is no type and screen or type and cross-match in the computer system.

The way that I can see this being an issue of consequence is of course when we get into the situation in which we need blood right away, and the hospital may actually be low on blood. I don't know how many of your hospitals that comes up in, but Detroit's about an hour away from Ann Arbor, and it's not uncommon actually that we have to call to Detroit for blood.

So if we have somebody who needs

blood and we don't have appropriate blood in the hospital for them, that can be a real issue in our hospital. We do a lot of cases. It's a VA, so it's smaller volume, of course, than the university.

But I could definitely see this coming up on a somewhat regular basis in our hospital. I don't know if other people's hospitals are similar.

DR. HALPERN: I know some hospitals also require two types to complete the type, and that's where I think the problem comes in. Both in the hospital where I came from and in the VA where I currently work, they require two types if you've never received blood before, to confirm your type.

So that's where I see it as an issue, you know. What is completing the type? Because a lot of times completing the type is doing that second blood draw, which often does not -- not often, but not infrequently doesn't get done until the morning of surgery.

DR. WILHOIT: One of the things that I thought was problematic here were the exclusions. If a type and screen or type and cross-match wasn't ordered, that was an exclusion. But it seems like that might be the situation in which it was most of a problem.

Also, there was an exclusion for patients not admitted from home, and that was a proxy for a non-emergency admission. Again, the exclusions just weren't working for me.

DR. HALPERN: I don't have a problem so much for the patients not admitted from home, because I think it's hard, you know, when you try to do a large volume chart abstraction, to figure that out.

I think, you know, the patients who are not admitted from home tend to be the ones who are going to be more urgent and not so elective, and I think they really wanted to focus on the elective patients.

DR. WILHOIT: And I don't have a

problem excluding patients not admitted from home. But it seems like then there should also be an exclusion for emergency cases, or something -- you know, and that's the only -- I mean that's the whole proxy for non-emergent, and it seems like it's a fine exclusion, but not adequate.

DR. SAIGAL: I think the nursing home point is important as well. A lot of patients from nursing homes would need this done, and it's not an emergency. It's just they're in a different location than home.

MR. RIVENBURGH: But it clearly says these are elective cases. So if the patient's coming from a nursing home and it's an elective case, you know. I mean I agree with --

DR. SAIGAL: But they're excluded, right? They're --

MR. RIVENBURGH: Right, and they shouldn't be in that particular situation, if it's elective and they're coming from a

nursing home. But the emergent case, obviously, is a whole different ball game.

DR. DILLON: Are the groups defined strongly enough or properly enough that this isn't just going to just drive everyone ordering a type and cross on every single patient within these specialties?

CHAIR MORRIS: It was the same group, wasn't it, orthopedic?

DR. CARPENTER: Yes. I thought the groups were -- I thought they had a list of diagnoses. I thought that was reasonably straightforward.

DR. DILLON: And we know that from the cardiac patients as well. I didn't look at the cardiac list, and that's -- but again, my concern is just from a pure cost point of view, that all of the sudden now we're going to have 85 type and cross every morning, and it's just going to overload the blood bank, in terms of being able to handle these.

DR. STAFFORD: Yes, and I think in

a particularly in teaching institutions, the default would be I'm not really sure which ones I'm supposed to get it on, so I'm just going to get them on everybody, and that way we're covered and we don't get dinged.

So those are some of the unintended consequences of putting some of these things out there. People find work-arounds, and we know work-arounds cause their own problems.

DR. CIMA: But I mean it says type and screen or type and cross. So it's very clear that you Can do one or the other. So you just have to have very strict protocols about which ones would get it.

CHAIR MORRIS: Okay. So several issues. One, a question about the gap and how big of a problem this is. Another is, is this an issue of consequence. Would this matter? I think that it is an issue of consequence, to an extent. But then there are some points about the exclusions, patients without an

order to type and screen or type and cross are excluded. That doesn't necessarily make sense to me.

Patients not admitted from home. A very valid point was brought up that those maybe precisely the patients having elective operations that need to have this measure. And then the last issue, could you please restate that, Dr. Stafford?

DR. STAFFORD: It got to what happens in a lot of institutions, when you have very specific cases that you apply a rule to, or a measure to, particularly in large teaching institutions.

People aren't going to remember who's supposed to have what, so the default will be to order a type and screen or a type and cross on every patient who goes to the operating room, which then causes more anemia, because we're bleeding patients we don't need to, overworks the blood bank, and uses resources that could be used better elsewhere.

CHAIR MORRIS: Okay. So a concern about overuse. Would JCAHO like to respond to these issues?

DR. GAMMON: Yes. Our pilot rate was 92 percent, and really we had a self-selected group of hospitals that were interested, I believe, in blood management. So I think that that reflects it probably, it can't be used for the universe of hospitals.

We also used the exclusion of you had to have it ordered, because in order to be in the numerator, you have to be in the denominator. So we could have selected elective surgical patients, and if they didn't have a type and screen ordered, then they couldn't get to the numerator.

So we wanted to exclude anybody that didn't have an order for type and screen or type and cross, and only concentrate -- if you did have one ordered, then was it done. We were really hoping and, you know, there's a hospital-wide initiative for the safety

surgical sheets where you have it on there, and you're checking before you go into the surgery -- did you have your -- is your blood available, because you know, and some of it is anecdotal, because not everybody gets -- Can be captured in some kind of a rate of how many people didn't have the blood ready by the time that they went to surgery?

Because nobody's really capturing that right now, and then you know, sometimes they have to end up getting uncross-matched blood if they don't have their blood type available.

CHAIR MORRIS: Does anybody have any further questions about this measure?

(No response.)

CHAIR MORRIS: Okay. Let's go ahead and vote. Does the measure meet NQF criteria for Importance to Measure and Report?

[COMMITTEE VOTING.]

CHAIR MORRIS: Four say yes, 17 say no. So that concludes the discussion of

this particular measure.

We now have a few moments of member and public comments before we adjourn for today. So I'd like to invite the developers and the public to speak, if they'd like to, anybody on the phone or here in person.

(No response.)

CHAIR MORRIS: It's very quiet out there. So I guess what I'd like to do next is to thank our developers. I know a tremendous amount of work went into creating these measures, and a lot of sweat equity there. I'd also really like to thank our panel, our steering committee, for devoting quite a bit of time, precious time and for their presence as well, and for their stamina today.

Hopefully tomorrow will be a little bit less grueling. It will be shorter, and we only have one set rather than three sets of measures to go through. Anybody want to add anything?

(No response.)

CHAIR MORRIS: The room will be locked overnight. If you would like to leave anything in the room, that's your call. If there's anything you feel uncomfortable leaving in the room although it's locked, please do take it with you. We're starting at 8:30 tomorrow morning.

DR. ROGERS: Great. Just one quickie. You mentioned, and I want to make sure that JCH hears the message, that this is a very, very important issue, and the fact that we have not supported the measures as they've been presented in no way reflects the importance and the way we would -- we'd love to see something positive out of this, rather than the negative, and I think that's the message I'd like to propose.

CHAIR MORRIS: Thank you. All right, thanks, everybody. Good night.

(Whereupon, the above-entitled matter went off the record at 5:42 p.m.)

F

O

Q

R

W

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Surgery Endorsement Maintenance 2010 Steering Committee

Before: Arden Morris, Chair

Date: 02-28-11

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

Court Reporter