NATIONAL QUALITY FORUM

+ + + + +

ALL CAUSE ADMISSIONS AND READMISSIONS
STEERING COMMITTEE

+ + + + +

TUESDAY
May 6, 2014

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:00 a.m., Bruce Hall and Sherrie Kaplan, Co-Chairs, presiding.

PRESENT:
BRUCE HALL, MD, PhD, MBA, Co-Chair
SHERRIE KAPLAN, PhD, Co-Chair
KATHERINE AUGER, MD, Msc, Cincinnati Children's Hospital
FRANK BRIGGS, PharmD, MPH, West Virginia University Healthcare
JO ANN BROOKS, PhD, RN, Indiana University System
JOHN BULGER, DO, MBA, Geisinger Health System
MAE CENTENO, DNP, RN, CCRN, CCNS, ACNS-BC, Baylor Health Care System
HELEN CHEN, MD, Hebrew Senior Life
ROSS EDMUNDSON, MD, Adventist Health System
W. WESLEY FIELDS, MD, FACEP, CEP America
STEVEN FISHBANE, MD, North Shore University Hospital and LIJ Medical Center
LAURENT GLANCE, MD, University of Rochester
ANTHONY GRIGONIS, PhD, Select Medical
LESLIE KELLY HALL, Healthwise
PAUL HEIDENREICH, MD, MS, FACC, FAHA, Stanford University School of Medicine
KAREN JOYNT, MD, MPH, Brigham and Women's Hospital

PAULA MINTON-FOLTZ, RN, MSN, Harborview Medical Center; UW Medicine
CAROL RAPHAEL, MPA, Subject Matter Expert
PAMELA ROBERTS, PhD, MSHA, ORT/L, SCFES, FAOTA, CPHQ, Cedars-Sinai Medical Center

ALISON SHIPPY, MPH, Consumer-Purchaser Alliance, National Partnership for Women & Families
THOMAS SMITH, MD, FAPA, American Psychiatric Association
RONALD STETTLER, United Health Group
CRISTIE TRAVIS, MHA, Memphis Business Group on Health

NQF STAFF:
CHRISTINE CASSEL, MD, NQF CEO
TAROON AMIN, Special Assistant to the President and CEO
HELEN BURSTIN, Senior Vice President, Performance Measurement
ANNE HAMMERSMITH, General Counsel
ANDREW LYZENGA, Senior Project Manager, Performance Measurement
ADEELA KHAN, Project Manager, Performance Measurement
KAREN PACE, PhD, RN, Senior Director, Performance Measurement
ZEHRA SHAHAB, Project Analyst

ALSO PRESENT:
SUSANNAH BERNHEIM, MD, MHS, Yale University
JEPTHA CURTIS, MD, Yale University
NIHAR DESAI, MD, MPH, Yale University
ELIZABETH DRYE, MD, MS, Yale University
JANE HAN, Society of Thoracic Surgeons*
LEIN HAN, PhD, CMS
LORI GEARY, MPH, Yale University*
JEFF JACOBS, MD, FACS, FACC, FCCP, All Children's Hospital
MARI NAKAMURA, MD, MPH, Boston's Children Hospital

JOHN MULDOON, MHA, 3M Health Information Systems*
SEAN O'BRIEN, PhD, Duke University
ROBERT McNAMARA, MD, MHS, Yale University
ISURU RANASINGHE, MD, PhD, Yale University
DAVE SHAHIAN,, MD, Society of Thoracic Surgeons*
MARK SCHUSTER, MD, MPH, Center of Excellence for Pediatric Quality Measurement*
LARA SLATTERY, MHS, American College of Cardiology
LISA SUTER, MD, Yale Center for Outcomes, Research and Evaluation
ALAN ZASLAVSKY, PhD, Harvard Medical School
LEORA HORWITZ, MD, MHS, Yale University

* present by teleconference

A-G-E-N-D-A

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

7:57 a.m.

MR. AMIN: Good morning, everyone. Welcome back to day 2 of the Readmissions In-Person Steering Committee meeting.

Thank you again for all of your work yesterday. It was quite a day reviewing all the measures that we got through but we were successful.

I wanted to welcome Christine Cassel, our CEO at NQF for a quick welcome to the committee.

DR. CASSEL: Thanks, Taroon. And thanks, Sherrie and to Bruce too for chairing this important work.

I was able to sit in and eavesdrop on some of your conversation yesterday. And so really especially appreciate the thoughtfulness and care and openness, the spirit really of open debate and discussion that goes on here at NQF.

And it was what makes the

deliberations but also the conclusions of this work carry the weight that they do. Not only in government programs but increasingly in private sector programs as well.

So, we just always need to stop just for a moment and recognize all of the volunteers and the expertise around this table that has made the commitment to actually make this multi-stakeholder model of healthcare quality measurement actually work for the nation. So we really appreciate that.

I wanted to just -- I take every opportunity I can to meet with our committees and talk with you about your work but also to let you know what's in a nutshell happening at NQF.

We are as you probably have seen recently right in the center of lots of really important discussions and debates about what our board is calling measurement science, really at the forefront of measurement science. Just in the last few weeks all the

attention to whether or not and if so how to risk-adjust measurement for sociodemographic status I think has ignited a needed national debate.

Carol Raphael and I were talking about this on the elevator, that these issues, many of these issues just are never done. We're just marching towards a better state, a better measurement science and a better way of implementing quality measurement. And I think that's a really good example.

We also are internally at NQF moving to a way of recognizing that endorsed measures -- there's more to say about whether a measure should be endorsed or not. Sometimes it's not just, you know, an up or down decision. But it's a kind of it depends. And often it depends on what it's being used for and whether it's for trial use.

We haven't really had the ability to make those kind of distinctions because we haven't really had the ability to do rapid

cycle feedback and follow-up.

And we now are actually moving to a place where we can do that with an open pipeline for measure submission, a model of standing committees so that it doesn't take us six months to assemble a committee every time we have a group of new measures.

We're not there yet, but with the full support of CMS we are piloting some of these approaches with some of the contracts we have this year. And I'm very hopeful and optimistic that it's going to give us a way of reducing the cost and the length of time that it takes to get this process through and to be able to be more adaptive then to what's happening in the scientific world as well as to what the clinical world really needs from us.

Another thing that is happening is that you know we have a process to evaluate the measures and endorse them or not. And then we have a process called MAP, the Measure

Applications Partnership, which selects which measures should be used for which federal programs.

And a lot of the discussion that goes on at the MAP is very similar to this discussion that goes on here. And in many ways we think it could be much more efficient if the two were sort of streamlined more to be able to identify what purpose -- maybe at this stage what purpose measures should be used for and feed that into the MAP process so that it makes that whole process more streamlined.

And in order to do that within the staff we're taking advantage of staff expertise in Lean and Six Sigma reengineering.

I don't know if you realize that Taroon is a black belt among us here. So, in his spare time he's actually helping us with the staff process of finding ways to cut the waste out of our process and become more lean and more efficient and effective in what we do. So we're very fortunate to have that kind

of expertise on our staff as well.

So you'll be hearing more about many of these things that are developing at NQF. But in the meantime the policies are the policies and it's really important that we adhere to these publicly available and carefully established policies which you're helping us to implement here today.

So let me just stop there, Sherrie, and see if anybody has any quick questions. I know you have a lot of work to do so I don't want to hold you up too long.

CO-CHAIR KAPLAN: Questions or comments for Chris? Dr. Cassel.

DR. CASSEL: Okay, thank you.

CO-CHAIR KAPLAN: Well, thank you. That was very, very helpful, especially the -- knowing that on the horizon there may be some of the synergy between the MAP process and this process I think helps with some of the frustrations that many feel for the purpose of measurement and a little concern about how

these things are going to be applied. So that was very, very helpful. Thanks, Chris.

So, I'm going to ask Taroon to kick off in a second, but yesterday it was noted that we laid out a bunch of ground rules and then assiduously ignored all of them.

And so today we are going to adhere to some ground rules hopefully more closely. We did a lot of hard work and you were extremely efficient but I think it will help if we can make the questions and the comments very concise. And then hopefully from the measure developers also the responses concise. And so that will keep us on track to make sure -- because we're going to start losing people as their plane flights and so on get going.

So if we can crisp up our comments and questions, and then also the responses from the measures developers, if those can be a little bit more concise will get us through today's agenda in a timely and hopefully full-

throated discussion way.

So, Taroon?

MR. AMIN: So, if we can -- as we get started here I would welcome the developers from STS to join us at the table as we get started.

But I'll actually turn it over to Adeela if you can just walk through a quick summary of day 1 here and then just quickly walk through the agenda for day 2.

MS. KHAN: Sure. So we actually were able to get through the agenda for day 1 and evaluate all the measures that were supposed to be.

Just a quick recap of where each measure is. 2502, the all-cause unplanned readmission measure for 30 days post discharge from inpatient rehab facilities passed.

2512, all-cause unplanned readmission measure for 30 days post discharge from long-term care hospitals was a measure where we weren't able to reach consensus. And

so we'll be following up with those developers but that measure will still be going out for public and member comment.

2375, PointRight OnPoint 30 SNF rehospitalizations passed.

2510, skilled nursing facility 30-day all-cause readmission measure passed.

2496, standardized readmission ratio for dialysis facilities was another measure where consensus was not reached.

2503, hospitalizations per 1,000 Medicare fee-for-service beneficiaries was consensus not reached.

2504, 30-day rehospitalizations per 1,000 Medicare fee-for-service beneficiaries was consensus not reached.

2505, emergency department use without hospital readmission during the first 30 days of home health passed.

2380, rehospitalization during the first 30 days of home health passed.

And 0327, risk-adjusted average

length of inpatient hospital stay was another measure where consensus was not reached.

Quickly we're going to be going over the following hospital measures today. We have two CABG measures, one risk-adjusted vascular procedure, two pediatric measures, one all-cause, one lower respiratory infection. And we have a PCI, AMI and outpatient colonoscopy.

MR. AMIN: Okay, thanks. Thanks, Adeela. I just wanted to point out two other things for today.

Again, a sincere try all for yesterday. There was a lot of work that was done. Again, NQF would not be able to achieve its goals without volunteers like you spending time through the workgroup calls and time that we had yesterday and today.

I just wanted to also follow up that we'll try to do an evaluation or just a discussion around the dry run results on the 1789 all-cause measure today.

And given the recommendation for endorsement on the two measures yesterday that were -- had the same measure focus and the same target population we will likely have a follow-up call to have a discussion around best-in-class/competing in addition to the potential two measures that are up for discussion today.

But we'll first evaluate them independently and then have that discussion later on.

So again, I'll turn it over to the chairs if there's any other reflections on yesterday. Otherwise, you guys can lead us directly into the conversation.

Is there anything else you wanted to add, Helen? Okay, thank you.

CO-CHAIR KAPLAN: And my earlier comments were not meant to truncate anything. I just want to make sure everybody gets a chance to say what they have to say about the measures under consideration and do so and get

their points made and considered.

With respect to the reviewers, if you could each time you make a comment state your name because the people recording this can't see you. And they know which microphones we're sitting at so they can hear who we are but they can't hear who you are.

So if you could each time say your name. We're going to ask yourself to introduce yourself and your background quickly and then give a two-minute summary of the measure.

But if you could, when you're talking, after your introductions repeat your name again that will be very helpful. So could you introduce yourselves and then the measure.

DR. JACOBS: My name is Jeff Jacobs and I'm a cardiac surgeon from Johns Hopkins and All Children's Hospital in St. Petersburg, Florida. And I'm joined today by Sean O'Brien.

DR. O'BRIEN: Hi, my name's Sean O'Brien. I'm a statistician at Duke University Medical Center. And we serve as an analytic center, the STS database.

DR. JACOBS: We also have on the phone with us two members from the Society of Thoracic Surgeons, Dave Shahian who's a cardiac surgeon from Harvard, and Jane Han who is staff at Society of Thoracic Surgeons.

CO-CHAIR KAPLAN: Cathy, can we just make sure that those people have open lines?

OPERATOR: Yes, ma'am, their lines are open.

DR. JACOBS: So, the measures that are before the group today, actually the next two were developed in parallel. There's the Risk-adjusted Coronary Artery Bypass Graft Readmission Rate from the Society of Thoracic Surgeons and there's the Hospital 30-day All-cause Unplanned Risk-standardized Readmission Following Coronary Artery Bypass Graft Surgery

from Yale.

These two measures were developed in parallel and in collaboration between the two groups as part of a process in collaboration with CMS.

And the two measures are somewhat different in that the Yale measure is based on administrative data and the Society of Thoracic Surgeons measure is based on clinical data from a clinical database, a clinical database that currently has a penetrance of between 90 and 95 percent of hospitals in the United States, but not 100 percent.

And I think both groups, the Yale group and STS, view these two measures as complementary with strengths and weaknesses that complement the other measures.

And I think the way I kind of look at this simplistically is that the Yale measure is based on administrative data. An advantage and a strength of the Yale measure is that 100 percent of hospitals in the United

States that do coronary artery bypass surgery participate in the administrative data sets which the Yale measure is developed.

Meanwhile, in the Society of Thoracic Surgeons only 90-95 percent of the hospitals participate. But we feel that the actual risk adjustment in the measure is somewhat enhanced because of increased ability to use clinical variables both for defining the model of patients, the cohort of patients, isolated CABG patients, and for the variables used to risk-adjust.

I think that's a brief summary of how these two measures fit together and what we've done. And I think we'd be happy to answer any questions.

CO-CHAIR KAPLAN: Thank you very much. Bruce and Paul were the discussants on this measure. With respect to the evidence, Bruce, do you want to go first?

CO-CHAIR HALL: I think overall this is an outstanding measure and in terms of

evidence the evidence is strong.

DR. HEIDENREICH: I would agree. While I think it should be classified as an intermediate clinical outcome I think there's a strong rationale.

CO-CHAIR KAPLAN: Other comments and questions from the group? Are we ready to vote evidence?

MS. SHAHAB: So, we're going to vote on 1(a) evidence, 1 for yes, 2 for no. And your time begins now. I think we have all 22 votes. 1(a) evidence, 22 yes, zero no.

CO-CHAIR KAPLAN: Thank you very much. Onto performance gap.

CO-CHAIR HALL: I'll jump in front, Paul, and then you can follow. Overall, about a 13.5 percent readmission rate depending on exactly which set and period and so on that you look at.

In terms of risk-standardized rate it comes out to about 17 percent with a range of 12.5 to 34.2. The interquartile being

about 15 to 18.

So, about a 3 percentage point spread in the interquartile range. So, at least, you know, comparable if not as good or better than many of the other measures that we're dealing with. In terms of the performance gap.

DR. HEIDENREICH: I would agree. I would say it's at least a moderate level of potential for improvement.

CO-CHAIR KAPLAN: Other comments from the committee? Hearing none are we ready to vote?

MS. SHAHAB: Voting for 1(b) performance gap, 1 high, 2 moderate, 3 low, 4 insufficient. Time begins now. Just one more.

We have all 22 votes for 1(b) performance gap: 6 high, 16 moderate, zero low, zero insufficient.

CO-CHAIR KAPLAN: Great. For scientific acceptability we're going to move

to reliability first. Sorry, priority. Priority.

CO-CHAIR HALL: In terms of priority the developers make the case that it's between two and three hundred million dollar target roughly as a subset of all readmission costs that are a burden on Medicare.

So, again, I think it's -- that's probably an understatement of the magnitude of the issue in terms of priority. So, my feeling on priority was at least moderate if not better.

CO-CHAIR KAPLAN: Paul?

DR. HEIDENREICH: I would agree on moderate. It's not as common as readmission -- as some of the other medical diagnoses, but still substantial.

CO-CHAIR KAPLAN: Other comments? Are we ready to vote on priority?

MS. SHAHAB: Voting for 1(c) high priority, 1 high, 2 moderate, 3 low, 4

insufficient and your time begins now.

We have all 22 votes, 1(c) high priority. Five voted high, seventeen moderate, zero low, zero insufficient.

CO-CHAIR KAPLAN: Thank you. Now onto scientific acceptability and first reliability.

CO-CHAIR HALL: So I think the developers have provided outstanding information. I think in terms of reliability actual implementation cutoffs for distinction and whatnot are actually not described. Those would be determined at a later date. So in terms of the most kind of classic rigorous signal-to-noise it's not really possible to comment on that until some of those later details would be specified.

In the background I think every other aspect of data field reliability, consistency, reproducibility, and so on are met by what has been submitted in the material.

CO-CHAIR KAPLAN: Paul?

DR. HEIDENREICH: I did see, maybe you can correct me, but it looks like you did compare, say, one year and three years worth of data. If I got that right. And that had reasonable correlations. It looks like -- it looked like you had a score-out of 0.78 if you had 300 hospitals included. I don't know if you have any further clarification.

DR. O'BRIEN: This is Sean O'Brien from Duke University.

One thing we did with respect to reliability was to estimate the proportion of variation that was explained by true signal variation as opposed to random statistical fluctuations.

And for that type of analysis we basically used the sample of all of the hospitals, there's approximately 1,000 hospitals in the development data set.

And when you included all the hospitals that have at least 30 cases which

might be a very common, inclusive threshold for reporting results of a hospital, the reliability, the percentage of explained variation by signal was 47 percent.

And then we looked at thresholds of what if we only reported results for subsets of hospitals that have at least 50 cases, at least 100, or at least 200 and those are respectively around 50 percent, 55 percent and 65 percent.

So I think the results demonstrated the potential for higher reliability with -- even with a very inclusive threshold as being kind of, you know, adequate or moderate, and the potential to have very high reliability in the subset with larger volumes.

In terms of the comparing one year versus three year, that may be -- I think basically we saw high agreement between different outcomes done across different time periods. But actually I don't have the

details up in front of me.

CO-CHAIR HALL: So I'd like to support but back up and clarify something that Sean said.

So indeed, Paul, as you point out there was good consistency and reproducibility of fields and data.

With respect to the reliability that you talk about, Sean, again, I respect your numbers. When an institution submits a 30-case estimate you had a number of 0.47. For 200 cases it would be 0.641. The development was on a 3-year data set, right?

So, again, if that's how the measure were going to be implemented then those would be good reflections of the reliability.

But those -- the actual sort of application of the measure is not yet determined. And so, again, those are good numbers. As Sean correctly phrased it that represents great potential for high

reliability. And so I support what he said.

But those numbers reflect a 3-year data set and an example of a development, an example of the potential that could be achieved.

CO-CHAIR KAPLAN: Other comments from the group? Go ahead, Larry.

DR. GLANCE: So, I have a comment about specification. I have great respect for the fact that there is a tremendous amount of clinical expertise as well as statistical expertise that went into the development of this outstanding measure.

The comment that I have is that this is a measure for isolated CABG surgery. And as part of the specification it also includes patients who underwent a combined CABG and ventricular assist device placement.

The rationale as I understand it for including the VAD patients is that there are occasions where because of a quality issue a patient is unable to separate from the

heart-lung machine and requires a ventricular assist device.

On the other hand at many heart failure centers there is an intention going into the surgical procedure that these are very, very high-risk patients and that there is a high likelihood that the patient will in fact, although the planned procedure is a CABG, that the patient will in fact probably need to undergo ventricular assist device.

The reason this is important is because the readmission rate for CABG patients is very, very different from the readmission rates for ventricular assist device patients. So my question for the developers is, knowing this, why would you have included VAD patients as part of the specification for this measure.

DR. JACOBS: Well, thank you. This is Jeff Jacobs. And first of all, that's an excellent question and some excellent observations.

A couple of clinical facts that I

think would help understand the rationale for the way this was developed.

First of all, substantially less than 1 percent of all coronary artery bypass grafts performed in the United States are associated with the use of a ventricular assist device. Ninety-nine percent of them are not.

Of those that are associated with ventricular assist device usage most of them are unplanned. And a patient is taken to the operating theater, undergoes coronary artery bypass grafting, cannot separate from the bypass machine, a variety of interventions are tried including a machine called an intra-aortic balloon pump.

After all of those things failed then really the only option is to put the patient on a ventricular assist device. And most of the time that's in an unplanned situation.

It is true that in some heart

failure centers patients go to the operating theater for planned ventricular assist device insertion after coronary artery bypass grafting but that's rare. It's extremely rare.

That being said, it is a fact that patients who get a ventricular assist device either planned or unplanned after a coronary artery bypass grafting, should they survive and go home have a higher rate of readmission than those without a ventricular assist device, no doubt.

When the measure was developed the measure was developed using a data set where we just knew that the CABG was associated with ventricular assist device insertion.

Since that time the STS database has been modified so that ventricular assist devices are now tracked as to whether or not the insertion is planned or unplanned. And that's a change in the database since the measure was developed.

Therefore, moving forward we certainly can implement this measure with the definition of isolated CABG that would include patients who had an unplanned ventricular assist device but excluded those with a planned ventricular assist device.

CO-CHAIR KAPLAN: Thank you. I think this discussion shades into validity because we're talking then about the accuracy of the application of the measure rather than the reproducibility of it. So, anymore questions on reproducibility? Go ahead, Paul.

DR. HEIDENREICH: Well, I have a question about that, but we can hold it.

CO-CHAIR KAPLAN: Can we hold it? Yes, for the validity question. Because right now I'd like to stick to reproducibility.

So, we've heard the reproducibility is in the zone. And actually for some of us who look at these kinds of signal-to-noise numbers for those hospitals that have a fairly large number that's a good

number. It's a reasonable number to have given the database.

Any other comments on reproducibility? Are we ready to vote?

MS. SHAHAB: Voting for 2(a) reliability - 1 is high, 2 moderate, 3 low, 4 insufficient and your time begins now.

We have all the votes for 2(a) reliability. Eight high, fourteen moderate, zero low, zero insufficient.

CO-CHAIR KAPLAN: Thank you. Now we're onto validity. So, Bruce, do you want to?

CO-CHAIR HALL: I have a couple of points, but Paul, you were about to go ahead.

DR. HEIDENREICH: Well, just a follow-up to your comment that you can now have a field for a planned ventricular assist device is how can you -- are you confident that that won't be gamed? That seems very hard to control.

DR. JACOBS: Absolutely. I

anticipated that being the next question. I've got the word "gaming" written right here.

(Laughter)

DR. JACOBS: Oh, this is Jeff Jacobs. I'm supposed to identify myself.

Well, that's a great question. And I think one potential tradeoff of including patients with unplanned ventricular assist device and excluding patients with planned ventricular assist devices is that the system then is subject to gaming.

I think the way that can be addressed is through a combination of proper definitions, proper documentation and then audit of that documentation.

The Society of Thoracic Surgeons database is one of the most rigorously audited clinical databases in the United States with multiple sites undergoing site visits with audit every single year.

And it would be a relatively simple process during that audit to audit this

field and to make sure that the documentation is in place to document that the patient truly went to the operating theater with a planned ventricular assist device insertion.

And that would simply require a note in the chart that says that the family was consented for a planned ventricular assist device and that the clinical team felt it was likely that that might be needed because of the patient's severe heart failure.

So I agree with you that gaming is a potential problem. The solutions to addressing that are good definitions, documentation and good audit.

CO-CHAIR KAPLAN: Thank you. I think that may get to -- so Paul, you might want to bring that up again, or at least remind people of it when it comes to the use. Because unintended consequences is one of the use parameters. Bruce?

CO-CHAIR HALL: So in terms of reliability I thought I would present my 20-

second summary for the group. This is a risk-standardized readmission ratio. Ninety-five percent intervals are provided. Medicare fee-for-service greater than 65 then obviously but matched to the STS data. So matching between the programs is one of the prominent features.

This is isolated CABG. We've already heard some comments about that. The definition of isolated CABG is provided and well specified with note to that issue that we've already heard about.

Patients have to be discharged alive and then readmitted within 30 days from discharge.

The exclusions include under 65, patients that they were not able to match between the data programs, cases that are not deemed standalone by their definition which we've touched on, patients who died in the hospital or were discharged to AMA.

Now, patients who died in the hospital, there's some small controversy about

how to handle a death in the hospital with respect to readmissions because obviously that was not a good outcome and yet that patient is not eligible for readmissions.

Suffice to say that because that is controversial and there's probably a lack of 100 percent consensus within healthcare about how to do that at least as many people are excluding deaths as taking any other approach. So this is consistent with that.

If patients were not fee-for-service, excluded index more than 365, not first admission, all exclusions, all specified well.

Race and sociodemographics were not included but the developers provide good information about those variables. Again, this -- the information we're seeing is a 3-year development set. If it were a 3-year measure that might be questioned but it's not clear that in practice it would be a 3-year measure.

And the readmission is attributed to the first institution in the case where patients are transferred between acute institutions. And only one readmission would be counted.

So I think overall excellent specifications and high validity on all the aspects that I mentioned.

CO-CHAIR KAPLAN: Do the developers want to respond to that?

DR. JACOBS: Is there a particular question?

CO-CHAIR KAPLAN: Never mind. Paul?

DR. HEIDENREICH: I'd say the only concern for validity I would have, it's not a big concern, is the matching to CMS. I think given right now you're not allowed to match on Social Security number, is that correct? But obviously CMS could do that in the future. So, I assume there's not a 100 percent matching to the readmission to CMS.

DR. JACOBS: Right, so the issues we've had with Social Security number relate to our matching of the STS database to the Social Security Death Master File.

And that matching process worked quite well. We've published several papers and all was well until the Social Security Death Master File was modified. And with changes in the Social Security Death Master File a substantial portion of the deaths in the Social Security Death Master File are no longer re-disclosed. So that's not very useful for outcomes research.

As far as our matching with CMS, I think that it's -- I'll let Sean address the overall numbers but my understanding is that the overwhelming majority of patients over the age of 65 in the STS database are matched successfully to the CMS registry.

DR. O'BRIEN: This is Sean O'Brien. Yes, Jeff, that's correct. Using an indirect record linkage around 85 percent are

linked. But then if you look within the subset of sites that are actively participating it depends which direction you're linking from among the subset in Medicare what percent linked to the STS database within sites that are actively participating in the database, high nineties, 97 percent, 98 percent. So it's fairly complete.

And then going the other direction of course you don't pick up the Medicare Advantage plans, the HMOs. Of course those wouldn't show up in the claims-based measure either.

DR. O'BRIEN: Can I just -- this is Sean O'Brien again. Just to respond to the question about the time frame.

All aspects of the measure were really developed with consistency with other CMS readmission measures in mind. So a lot of the other readmission measures for AMI and pneumonia, et cetera, they were originally

developed with a 1-year time frame and they were subsequently in subsequent iterations converted to a 3-year time frame. So I think the measure developers had a 3-year time frame in mind just for consistency with CMS.

CO-CHAIR KAPLAN: Thanks for that clarification. Any other comments on validity? Go ahead.

MS. SHIPPY: I had a quick question about the specification. So, you had noted in your introduction that you had worked in collaboration with CMS. Can you discuss the choice for 65 and older for the patient denominator and CMS has 18 and older.

DR. O'BRIEN: CMS has 18 and older?

MS. SHIPPY: I understand that this is probably harmonization but it felt like it was an opportunity to have them discuss it.

DR. JACOBS: My understanding is that to be eligible for Medicare one has to be

either over the age of 65 or in renal failure. So, to be in the Medicare database and having undergone a CABG the two ways to get in there is either being younger and being in renal failure, or being over the age of 65. So that's why the age of 65 is part of this measure.

I'm not sure why a claims-based measure would be developed for over the age of 18 that's based on the Medicare data because I don't think a patient would be eligible for that unless they're on dialysis.

CO-CHAIR KAPLAN: Okay, when you put up your cards if you can turn them sideways so I can see them. They disappear when they're this way. Paul?

DR. HEIDENREICH: I know the Yale group has occasionally used the California data to test their model that was developed for 65 and older in Medicare but then to have it on a claims base for 18 and above. So I didn't deal with that measure but I know

they've done that with other measures.

CO-CHAIR KAPLAN: Thank you. Other comments? Okay, we're ready to vote validity.

CO-CHAIR HALL: So I have a question though. Is -- are we asking that the developers clarify that in the future the VADs would be so specified as planned or unplanned. Do we have that ability to request that? If they agree?

DR. JACOBS: We're very comfortable with that. We have the skills to do it and we're comfortable doing that.

CO-CHAIR KAPLAN: So how are we approving the measure though? As specified, as is. We can recommend that they make this change but we are approving or not approving the measure as it is currently presented. Ready to vote? Okay.

MS. SHAHAB: Voting for 2(b) validity, 1 high, 2 moderate, 3 low, 4 insufficient and your time begins now. Just

one more vote.

We have all the votes for 2(b) validity. Four voted high, seventeen voted moderate, one low and zero insufficient.

CO-CHAIR KAPLAN: Okay, we're onto feasibility.

CO-CHAIR HALL: I thought the feasibility was very reasonable. The main issues I guess that popped into mind would again be the penetration of the STS program into all CABG procedures across the country which I think is very, very high.

And then the issues around matching that have already been raised which might create some limitation around perfect matching. But again, the matching seems to be done at a very high level.

So otherwise, I did not see any major obstacles to feasibility. I defer to Paul.

DR. HEIDENREICH: I agree with that.

CO-CHAIR KAPLAN: Any other comments? Go ahead.

MS. HALL: I had a question about the proprietary nature and the potential fees and could that cause barriers for use by other organizations or particularly those who might be reporting to the public for public good or consumer organizations who might have an interest. Could you comment on that, please?

DR. JACOBS: So, the Society of Thoracic Surgeons is the largest professional organization of cardiac surgeons in the world. And almost all cardiac surgeons in the United States are members.

STS is a strong advocate of public reporting, a huge advocate. And currently our outcome data is publicly reported on two platforms, one through Consumers Report. And we partner with Consumers Report because that allowed public reporting from a respected organization at an arm's length from STS.

So, Consumers Report publicly

reports our outcomes measures with a web platform designed by Consumers Report and using STS data.

STS also reports our outcome data on our own website, www.sts.org. That information is available to anyone through the internet for free. So there's methods to access the results of our NQF-endorsed measures through our website for free and also through Consumers Report at an arm's length from us.

So I think the issue of public reporting and the proprietary nature of the database becomes essentially a non-issue because there's two ways to get that information from the website of STS or Consumers Report. And we're certainly a big advocate of transparency in public reporting.

CO-CHAIR KAPLAN: Thank you. Other questions or comments? Okay, we're ready to vote.

MS. SHAHAB: Voting for number 3,

feasibility. One is high, two moderate, three low, four insufficient and your time begins now.

We have all the votes for feasibility. Eleven voted high, eleven moderate, zero low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Now we're on the issue of usability and use.

CO-CHAIR HALL: I felt the usability was high myself. I did not see any major obstacles.

CO-CHAIR KAPLAN: Paul?

DR. HEIDENREICH: There's always the potential that with a procedure that's elective you could have surgeons not doing the cases. Although, as you say, you've been reporting data for a long time so I don't think this would have any significant incremental impact on selecting cases based on risk of readmission.

And then it just has the limitation I think that all the CMS 3-year

based measures have in that it just takes -- you can't see a rapid change in your program through those data.

CO-CHAIR KAPLAN: So the issue of gaming was raised. And I want to make sure we're all having a discussion about unintended consequences. Karen?

DR. JOYNT: Just two quick comments. I think one, just to get back to the ability of these models to sort of reduce the information that you can get from low-volume hospitals. And for two reasons, in case you're looking for a safety signal and also for consumers as we talked about understanding what the difference is between an average hospital that's small and a hospital about which we just don't know their performance.

Again, this all gets back to usability as opposed to the validity and I'm sorry to be a broken record on this.

And I forget what my other comment

was.

CO-CHAIR KAPLAN: Thank you. Do the developers want to comment on that? I mean, this is -- excuse me for interrupting.

This is going to be -- I mean, low-volume hospitals are going to have the same plaguing problem we've all talked about and will probably continue to debate for the majority of our remaining careers.

So it is one of these perplexing problems. Do you just not evaluate the low-volume hospitals? Do you evaluate them and give them the mean? Do you do all the things that we've talked about that they're going to always have problems associated with the error of estimation.

Developer?

DR. JACOBS: This is Jeff Jacobs again. And I think this recent dialogue raised three issues. One, risk aversion as an unintended consequence, two, gaming, and three, how to manage low-volume hospitals.

And I'll say a couple of sentences about each one and then I can answer more questions.

Certainly with any form of outcome reporting risk aversion is a possibility. The solution to that is a good risk adjustment methodology. And I think this measure as well as all of our isolated CABG measures implemented by STS have a very vigorous risk adjustment methodology that's designed to prevent risk aversion.

I think it's extremely unlikely that this particular measure would lead to any new risk aversion because isolated CABG is a group of patients that are already subject to multiple other NQF-endorsed measures including a mortality/morbidity/multi-domain composite.

So I think that although risk aversion is possible with any measure it's miigated by proper risk adjustment.

Regarding gaming of the system which came into context in this discussion with the planned versus unplanned VAD but

certainly could also become an issue with other components of any risk adjustment methodology I think the solution to gaming is having, again, good definitions for all the fields, having proper documentation of those definitions, and the application of those definitions and having a solid audit program.

And as I said before, I think our audit program of the STS database is as good as any and better than most clinical registries in the country.

Finally, related to the issue of low-volume hospitals, this is a challenging problem for almost any measure, especially measures that deal with relatively rare procedures and relatively rare operations.

And I think within STS we've done a lot to make sure that appropriate confidence intervals are utilized so that the low-volume hospitals and their unique situations are respected and accounted for.

And I think I'll turn this over to

Sean to maybe make an additional comment about what we do to deal with low-volume hospitals. It's a topic we discussed on frequent phone conferences.

DR. O'BRIEN: This is Sean O'Brien. I don't think there's a magic bullet for dealing with the problem of small sample sizes.

As Jeff mentioned we report measures with measures of uncertainty so I think that's about the best you can do is to say what the evidence is and report that there's a range of possible performance that's consistent with the observed data.

CO-CHAIR KAPLAN: Thank you. Bruce?

CO-CHAIR HALL: So I would like to add that with respect to this particular measurement access isolated CABG the developers do shed light on this for us in their reliability information.

For instance, portraying that the

signal-to-noise version of the reliability assessment down to 30 cases gives a metric about 0.47.

So with 30 cases assessed that level of reliability is as good or better than probably anything you see in healthcare which I don't know if that's a reflection that isolated CABGs end up being a pretty homogenous reproducible query into quality I guess is one way to put it.

For whatever the explanations are I think we can be at least somewhat comforted by the notion that the reliability, the signal-to-noise assessment of this specified measure remains as good or better than anything else we see down to levels of 30 cases assessed and perhaps below.

So, I personally take that as some comfort and reassurance around the specification of the measure.

CO-CHAIR KAPLAN: Thank you.

DR. JACOBS: This is Jeff Jacobs.

I agree with everything you just said and just to provide some clinical context.

A hospital doing 30 cases a year of coronary artery bypass grafting is a really low-volume hospital. I mean that's --

CO-CHAIR HALL: Jeff, this is 30 cases over three years.

DR. JACOBS: Yes, I'm getting to that. So 30 cases a year since you're doing about 2 a month, a little over 2 a month. Thirty cases over three years means that it's one of the most low-volume hospitals on the planet. It's not where I would go for my coronary artery bypass graft.

CO-CHAIR KAPLAN: I think we're getting the drift. Larry, can you make a concise comment?

DR. GLANCE: Always.

(Laughter)

DR. GLANCE: So, this is very concise. I think the point that Karen makes is a really important one. And it is

crosscutting.

And the point is, and it will need to be looked at at some point because it's applicable to all the measures.

When you use shrinkage estimators what you end up doing is classifying virtually all of the low-volume providers as if they were average and grouping them together with other higher-volume centers that may in fact be average. So that is a problem and I think it will need to be addressed at some point.

CO-CHAIR KAPLAN: Absolutely. There is absolutely no disagreement that low-volume hospitals remain a perplexing and problematic issues for all of these outcome measures. And we probably aren't going to resolve that here now.

On the other hand, for this measure it looks like the low-volume issue is probably as not problematic as we're going to get.

So, having said that, any other

comments or questions? Okay, are we ready to vote usability? Go.

MS. SHAHAB: Voting for usability and use, 1 high, 2 moderate, 3 low, 4 insufficient information. And your time begins now. One more vote, please.

We have all the votes for usability and use. Thirteen high, nine moderate, zero low, zero insufficient information.

CO-CHAIR KAPLAN: Thank you. So now we're onto endorsement.

CO-CHAIR HALL: I have no additional concerns.

CO-CHAIR KAPLAN: Other comments or questions? Larry, did you have your gizmo up? Okay. Other comments? Are we ready to go? Voting.

MS. SHAHAB: Voting for overall suitability for endorsement, 1 yes, 2 no. Time begins now.

All votes are in for overall

suitability for endorsement. For measure 2514 Risk-adjusted Coronary Artery Bypass Graft Readmission Rate, 22 yes, zero no.

CO-CHAIR HALL: We thank our developers for their input. I guess we neglected to ask whether Dr. Shahian or Jane Han had anything to add, but too late for those of you on the phone. We thank you for your input.

And we'd like to ask the next developers, Yale, CMS to come to the table. Thank you. We have our developers for the next measure 2515 to the table. Lein, Lisa and Elizabeth are with us.

So if you wouldn't mind briefly introducing yourselves and then your measure.

DR. SUTER: My name is Lisa Suter. I'm from the Yale Center for Outcomes, Research and Evaluation. And we're introducing the measure 2514.

CO-CHAIR KAPLAN: Lisa, can you move just a tad closer to the mike? Thanks.

DR. SUTER: Sure. Can you hear me now? Great.

So I think the overall --

CO-CHAIR HALL: Lisa, I'm sorry. Did you say 2514? 2515?

DR. SUTER: 2515. My apologies.

CO-CHAIR HALL: Fifteen, thank you.

DR. SUTER: I think the discussion with the STS measure very nicely summarized the collaborative process which CMS allowed us to engage in with STS to develop these two measures.

They are as harmonized as two measures I think could possibly be. And the success with which the claims-based measure was able to achieve cohort and risk adjustment validation is certainly due to the close collaboration that we were able to participate with STS's surgeons and their workgroup.

And I think we've talked about how important CABG is as a readmission measure so

I'm not going to speak to that individually.

This measure differs slightly from the registry-based measure in that it measures all-cause unplanned readmissions after isolated CABG procedures. And similarly to use as a vetted guideline concordant approach to measure development that's been supported by the MAP.

As I mentioned the measure was developed in close collaboration with STS and every step of the measure development was performed in parallel with the STS measure developers and clinical experts.

I think as Dr. Jacobs mentioned both measure developers recognized that there are pros and cons to each measure, and that they each have a place in the measurement process.

The registry measure noting, as he said, a clinical-based risk adjustment model that I think has a greater face validity among clinicians but a lower penetrance in terms of

the number of hospitals that can be captured and assessed without burden upon the hospitals.

And as noted before we have reached over a 97 percent agreement in the cohort definition with the only discrepancies either being distinct measure decisions such as MAZE procedures which were made in concert with the measure developers, or with incongruities that can be ascribed neither to the claims nor to the registry data specifically and could represent errors in either data source.

I think the other clarification I'd just like to offer, and I'm sure we'll have other discussions as we move on, is that while this measure was developed in the over-65 population I think it was noted that we have assessed it in a California all-payer data source.

This is to allow flexibility for other users of this measure to use it in an

all-payer data source. But it was in fact developed in a full national Medicare over-65 population data set.

And finally, just to remind people that for other readmission measures that CMS has implemented regarding the low-volume hospital discussion there has been -- always been an opportunity for hospitals to be noted either that they are no different from average, or that they are too small-volume to be ascribed to a particular category. So that that inability to categorize due to small sample size is transparent to users. Thank you very much.

CO-CHAIR HALL: Thank you, Lisa. Lein, any comments or anything else to add?

DR. HAN: Hi. I am Lein Han from CMS. And I just want to say that we appreciate very much the collaboration with STS. The working relationship was very good and I really appreciate that. So thank you.

CO-CHAIR HALL: Thank you. So

we're in the category of evidence. I would like to turn to our primary discussants Ross and John and ask them to open the discussion.

DR. EDMUNDSON: Yes, this is ground that we covered here. But the evidence is pretty compelling that there is a readmission problem in this population and that there's opportunity for improvement in that. So I think that's well established.

DR. BULGER: I don't have anything to add. From an evidence standpoint it's very similar to the last measure which we looked at.

CO-CHAIR HALL: I'm not seeing any other cards so we'll vote evidence.

MS. SHAHAB: Voting for 1(a) evidence, 1 yes, 2 no. Time begins now. Just one more vote.

We have all the votes for 1(a) evidence. Twenty-two yes, zero no.

CO-CHAIR HALL: Performance gap. Any additional commentary above and beyond

what we've reviewed? John?

DR. EDMUNDSON: Just on the information provided that they have a mean of 16.8 percent readmission in the range of 12 to 22.1 percent. But again it's information that's redundant as to what we discussed before.

CO-CHAIR HALL: Wes?

DR. FIELDS: Yes, a question you can either answer now as developers or later in the process if you think it's more appropriate.

But I'm just curious if analysis of claims data reveals whether there is a greater degree of variation or less than optimal outcomes among the sites that don't participate in STS or not.

So I'm just curious about whether or not that's been part of your analysis and whether you think there may be greater variation in those few remaining sites that aren't part of the STS registry.

DR. SUTER: This is Lisa Suter from Yale. It's an excellent question. Because of the proprietary nature of the STS data we do not have the ability -- we at Yale do not have the ability to identify individual hospitals that either matched or did not match.

I can't actually speak to whether or not STS has investigated that among the hospitals that they were unaware that did not match to the Medicare data.

I know that many of the hospitals that were not included in the validation process because they did not link or match did not have active participation in the STS registry which represents about 10 percent of hospitals in the nation.

DR. FIELDS: But do you know any ballpark on the number of total CABGs in Medicare that did not appear to have an STS case match?

DR. SUTER: So, speaking to the

isolated CABG cohort that we identified, approximately 30,000 patients or one-fifth of the national isolated CABG patients identified in claims approximately did not link or match.

DR. FIELDS: To an STS case.

DR. SUTER: To the STS. I don't know the outcome rate among -- we did not investigate the outcome rate among those patients.

CO-CHAIR HALL: Wes, does that answer your question or is it as close as we're going to get for now?

DR. FIELDS: Well, we'll probably talk more about it. I just find it fascinating. I mean if ultimately what we're trying to do is to reduce the remaining variation.

You know, I have a lot of respect, regard for the STS process and registry and measure, but it raises a question about which measure is most likely to actually reduce variation going forward.

CO-CHAIR HALL: Great. Any other additional comments? I'm not seeing any cards raised for performance gap.

MS. SHAHAB: Voting for 1(b) performance gap, 1 high, 2 moderate, 3 low, 4 insufficient. Time begins now.

We have all the votes for 1(b) performance gap. Nine voted high, thirteen voted moderate, zero low and zero insufficient.

CO-CHAIR HALL: Priority, John?

DR. BULGER: So this has a similar priority to the last measure. And as noted before this is one of MedPAC's targeted diagnoses for readmissions priority.

CO-CHAIR HALL: Okay, any additional comments? Not seeing any.

MS. SHAHAB: Voting for 1(c) high priority. One is high, two moderate, three low, four insufficient. Your time begins now.

We have all the votes for 1(c) high priority. Eighteen voted high, four

voted moderate, zero low and zero insufficient.

CO-CHAIR HALL: Moving into the scientific realm, reliability and validity. John, Ross, you want to open the discussion?

DR. EDMUNDSON: Okay, reliability. This is a test/retest split sample. And with intraclass correlation coefficient here. Large numbers.

What I found interesting -- and they used Medicare claims for the years 2008, 09 and 10 as well as the Society of Thoracic Surgeons. But I believe this is on the claims data that you're doing the split samples, is that correct?

DR. SUTER: That's correct.

DR. EDMUNDSON: And then on that the random split samples for each hospital was the intraclass correlation coefficient was 0.331 which was judged as fair. Could developers comment on that relationship?

DR. SUTER: Yes, we heard during

the workgroup the concerns about the low intraclass correlation coefficient during the split sample retest.

So, trying to understand the stability of the measure result, we think that the most robust and conservative assessment is to fully separate the sample of patients so that in an individual hospital there is no overlap between the two samples.

So we randomly split each hospital's patients into equal portions and then we calculate the risk-standardized readmission rate at the hospital level in each of those samples.

And using a 3-year data set which was what was available to us we received -- we yielded an ICC of 0.33 which you would agree is outside of the range of 0.4 to 0.7 which is usually interpreted as fair to good or moderate for intraclass correlation coefficients.

Although in many, most of the

readmission measures other than the hospital-wide readmission measure that CMS has implemented use three years of data in order to achieve a sample size. And this is particularly true in a procedural-based measure such as CABG.

In response to the concerns about the ICC and based on some recommendation from prior NQF discussions for other measures using a method similar to John Adam's paper which was referenced in regards to RAND's method using the Spearman-Brown prophecy formula -- and we have this information available to the committee if you'd like to see it -- we estimated what would the intraclass correlation coefficient be if we were actually able to create a 3-year sample that could be split into two equal 3-year samples. So that you had the volume of the 3-year sample but you still had two completely independent and non-overlapping samples.

And when we perform that analysis

for CABG the ICC rises to 0.5. It rises for all of the readmission measures that our measure developer has in front of the committee today and we'd be happy to share this information. We have copies that we can provide to you.

CO-CHAIR HALL: Additional comments on this area, reliability? Reproducibility. Not seeing any -- go ahead.

CO-CHAIR KAPLAN: I would say that for these intraclass correlation coefficients this is roughly in the range of the things you see.

And in part -- and we were just having a dialogue about it. In part it's because within hospitals when you're looking across patients within hospitals you're dealing with a dichotomous variable of readmitted/not readmitted. So it's a little bit of a compressed variance problem.

But having said that there are, you know, this is in the zone where you --

exactly what you would see when you look at other kinds of measures we've already considered for the intraclass correlation coefficient.

CO-CHAIR HALL: Not seeing any cards we'll vote on reliability.

MS. SHAHAB: Voting for 2(a) reliability, 1 is high, 2 moderate, 3 low, 4 insufficient and your time begins now.

We have all the votes for 2(a) reliability. One voted high, twenty-one moderate, zero low and zero insufficient.

CO-CHAIR HALL: And validity? Ross, John, opening comments.

DR. BULGER: So a couple of questions on validity. In general, the C statistic was 0.63 which is similar to the last measure.

This is administrative data, not clinical data so I wondered if you at some point could speak to that.

The other question that came up

with the last data was with the LVAD patients. And they are in this group. And we had just talked about the ability to exclude one subset of those.

But my assumption, and you can speak to this, that because of the administrative -- yours is administrative data that you could only exclude them totally or not exclude them, but not subset them into elective and non-elective LVAD patients.

The face validity from the panel I think was strong as well. In looking at your exclusions they were similar to the last one we looked at.

There was a question of excluding patients from the panel in our discussions from the workgroup of excluding patients who died in the 30 days.

And I think you had mentioned, you made some comments back on that already that they were in because that was what was similar to the there Yale measures, to keep those in.

But I was wondering if you could speak to that as well.

DR. SUTER: Great, thank you. And I also just wanted to correct an answer. I had responded to the question of capture. My off-the-cuff math was off. So instead of being 20 percent it's 10 percent non-capture rate. I apologize for that error.

In regards to post-discharge mortality. So patients who die within the hospital obviously are not at risk for readmission. They are in fact excluded from the measure.

There are a small proportion of patients who die after discharge from the hospital. That is about 1.4 percent of patients. They do as expected have a higher readmission rate.

This measure is paired with a mortality measure which is in front of the surgery committee and will be reviewed by the NQF later this year.

We think that that allows the capture of a full spectrum of quality outcomes. So to prevent any unintended consequences of measuring readmission in this cohort of patients.

In regards to the VAD issue, as has been previously noted we do include VAD procedures based on the recommendation of a host of cardiothoracic surgeons who are involved in this measure development. And it is harmonized with the STS measure.

As you noted we do not have the ability to finesse the -- the ability to identify unplanned versus planned VAD procedures. In a cohort of about 150,000 patients with isolated CABG only 90 have VAD procedures.

About 50 percent of them have percutaneous VAD procedures and they have a readmission rate very close to the mean 17.5 where the average hospital readmission rate is 16.5 in non-VAD patients.

Those that have more invasive VAD procedures do have a higher readmission rate around 30 percent.

Because they represent 0.03 percent of the entire cohort of isolated CABGs and they are not clustered in any one particular hospital or type of hospital we as measure developers are open to bringing back to our workgroup the recommendation from NQF to remove these procedures if there's a strong feeling that they misrepresent the quality of hospital performance using this measure.

We went ahead with the best recommendations from the largest group of cardiothoracic surgeons so we felt confident in that recommendation and we are eager for the NQF's advice.

And I will also talk about risk adjustment. I don't know if people wanted to comment on VADs before I move onto risk adjustment.

So, in the materials that we

submitted we also submitted a technical report. And on page 81 and 82 there are two graphs that I think are particularly helpful to understand the risk adjustment validation that was performed with the Society of Thoracic Surgeons.

And this is in a matched cohort of patients as was previously discussed.

And I know that it's not easy to see this, but these are all of the hospitals that were identified as outliers among all of the thousand or so hospitals included in the validation process.

And each pair of lines and dots represent the risk-standardized readmission rate achieved with the claims-based measure and that achieved with the registry-based measure.

And while it's I know impossible for you to see across a room, the gray area represents -- the top of the gray area represents the national rate.

And what I hope is visually apparent is that those paired lines are extremely overlapping. So in addition to achieving an ICC of 0.9 something depending on what ICC method you use so you can see that the risk-standardized rate is highly correlated, the interval estimates, the uncertainty around that interval estimate, excuse me, around that risk-standardized readmission rate is also highly correlated.

And any time you draw a line to move patients or hospitals into a performance category which is what the nationally reported readmission measures do. They report them as better than average, worse than average, or no different than average, or too small to quantify. You have to draw a line.

And I'm happy to share this. It's also in your materials. But when you draw that line some people fall on one or the other side.

But I think what's very reassuring

about this is that the two measures produce performance estimates for each hospital that are qualitatively highly similar. Quantitatively they do miscategorize a few percentages of patients, but the specificity in the claims-based measures is close to 100 percent. It's over 99 percent which I think for a high-stakes measure which readmission may be we think is the proper emphasis in terms of being conservative.

CO-CHAIR HALL: Okay. I'm sorry, I myself am trying to follow what you said while looking at the diagram. So I'm slightly lost on that diagram. But I know Larry has a question.

DR. GLANCE: So, I just wanted to comment on the validation part. And I want to preface my comments by saying I understand the need for a measure based on administrative data because some of the hospitals in the U.S. are not part of STS.

Having said that, you're

absolutely correct in that there's a very high level of agreement between your measure and the STS measure. In fact, according to your reporting there's about 97 percent agreement between the administrative data and the STS measure in terms of classification as high quality, average quality and low quality.

Having said that, when you look at the sensitivity of the CMS measure for identifying high-quality and low-quality hospitals it's about 40 percent and 60 percent respectively. So, I'd like you to comment on that.

DR. SUTER: So, I think the challenge with this validation process is in this case we were validating the risk adjustment. So we made the assumption that the registry data represents the gold standard.

We don't actually know what the gold standard for performance categorization is in the United States for isolated CABG

procedure readmission rates. So, I can't really comment on whether or not sensitivity of 40 or 60 percent is appropriate or not.

I think it is a challenge when you are creating measures that you want to be responsive to change and useful to hospitals, and yet they're being publicly reported and you want the estimates to be stable and reliable.

And in that situation we often need longer measurement periods, larger data sample sizes and we favor in the claims-based measure and the registry measure uses the same hierarchical modeling that does pull people towards a less outlier position. But I think we felt in the situation of how these measures may be used that that was a reasonable tradeoff.

And certainly the policy of implementation is not our decision, but we work in close concert with CMS to make sure that we're responsive to their needs.

CO-CHAIR KAPLAN: I'd like to just sort of clarify some issues because from a measurement person's perspective criterion validity says there is a gold standard and you're comparing a new measure to that gold standard. There is no gold standard.

Convergent validity says I have two sources of information and they're telling me roughly the same thing. That's confidence-inspiring from the graph I'm staring at right now. Two data sources tell you roughly the same thing.

Discriminate validity, however, says I can tell hospitals apart. And from that perspective not so much. So, can you help us understand sort of in those terms?

So, criterion validity is off the shelf. Convergent validity we're seeing evidence of. What happens to discriminate validity and can you tell hospitals apart?

DR. SUTER: So, we were unable to de-identify the individual hospitals that are

discordant. So, it's challenging to dig into the discordant hospitals to try and understand why a particular hospital might have been considered higher quality or lower quality from one data source versus another.

I think it is an important question. Unfortunately I can't speak to it.

CO-CHAIR KAPLAN: Thank you.

CO-CHAIR HALL: Any other comments, concerns on the topic category of validity? Wes?

DR. FIELDS: Yes, I just want to come at this from a different direction. So it's sort of reciprocal to my earlier comment.

I would assume that if you have a clinical data set as the registry does that you'd have more independent variables that help you get at the nature of quality if you will and to distinguish between facilities and programs.

So I just want to ask a first order question. Compared to the CMS claims

data set in terms of numbers of data points how much larger is the data set the STS registry uses compared to the number of elements in a claim stream that CMS would receive?

CO-CHAIR HALL: I want to push on you, Wes. I can see this might shed some light but we're only considering the measure in front of us, right? We're not really considering its comparison to an STS counterpart.

So, if your question helps us get to this measure I think we're okay. So, if the developers can comment on it in that light, in that context. Or did I misunderstand, Wes?

DR. FIELDS: I thought I was restating Sherrie's question from a different context. I think the issue of how you define quality is pretty interesting. And I'm assuming that having more data elements from a clinical registry that's larger in scope

than a claim stream gives you the possibility of doing that.

I'm just asking them to quantify how many more variables are in the registry stream. Because -- so it's really a way of restating Sherrie's question about how you get at the nature of quality and distinguishing between facilities and programs.

DR. SUTER: So I think -- thank you, Lisa Suter. The response is two that are C statistics so our discriminate ability is essentially identical.

And I think the other is that both measure -- I mean, certainly our -- we as the measure developer see room for both of these measures in the world of measurement. They offer unique perspectives. They were developed in an incredibly harmonized fashion and offer advantageous synergistic information about hospital performance, not necessarily replacement performance.

CO-CHAIR HALL: And they do

provide in their methodology report these insights such as you're seeing on this graph to try to help us understand whether one approach or the other approach helps discriminate the quality better.

The underlying notion that more variables will help you discriminate quality better may or may not be true. More variables could lead you to decide there is no difference in quality. So, the underlying construct is still ill-defined or controversial.

Other comments? Wes, do you have more comments? Any other comments in this category?

CO-CHAIR KAPLAN: I'd just add one thing. More variables usually help you improve your estimates of reliability, not necessarily your estimate of validity. So, reliability -- you ask more things, you get a tighter, more reliable response. But not necessarily a more valid response.

CO-CHAIR HALL: Okay, I'm not seeing any cards raised so we'll move on validity.

MS. SHAHAB: Voting for 2(b) validity. One is high, two moderate, three low, four insufficient and your time begins now.

We have all the votes for 2(b) validity. Two high, twenty moderate, zero low, zero insufficient.

CO-CHAIR HALL: Feasibility? Ross?

DR. EDMUNDSON: Feasibility, I think this is claims data. This is very feasible. We can do this.

CO-CHAIR HALL: Any other concerns? I don't see any raised.

MS. SHAHAB: Voting for criteria 3 feasibility. One is high, two moderate, three low, four insufficient and your time begins now.

We have all the votes for

feasibility. Twenty voted high, two moderate, zero low and zero insufficient.

CO-CHAIR HALL: Usability. John?

DR. BULGER: So I think the usability is similar at least to the last measure. I think the question came up in our pre-work is, you know, this discrimination issue of high performers, mid performers and low performers. And if this goes like some of the other measures that have been used which are set at a midpoint really from a payment standpoint I think there was concern amongst the group how that would perform.

And the other concern was if it really only was able to discriminate the tails what the use to the public would be from that standpoint.

Otherwise, I think we were fine with what we were looking at.

CO-CHAIR HALL: Paul?

DR. HEIDENREICH: Yes, just in terms of going forward from CMS' perspective.

I think there's going to be 20 -- there's about 20 hospitals rated as good or better than expected. Is that about right? Which some people have said is very low.

I'm not sure what the right percentage we should label as outliers, but is there any plan to change that implementation when this is reported on the website? About which one -- what fraction are outliers versus what fraction are not outliers?

DR. SUTER: So from a measure developer standpoint -- this is Lisa Suter -- that you're referring to the 2008-2010 data. I can't speak to more recent data. I don't have those estimates in front of me. Presumably more recent data would be used for reporting purposes.

And in terms of whether or not to use the same performance categorizations that are used in other measures I'll defer to Dr. Han.

DR. HAN: Hi, this is Lein Han

from CMS. At this moment we plan to continue the way we display the data on Hospital Compare.

And I believe that you all know but I just want to describe again the four categories. We have three compared to the national rate that's better, worse, or no different. And the other one is the category of small hospital that have less than 25, we put them aside. So, that would be the way this moment -- yes, that's the way we're going to plan to display the measure. Thank you.

DR. SUTER: And I'll just add that currently the interval estimates that were used for the graphic that was up and are used for the other measures reported on Hospital Compare uses a 95 percent interval estimate.

If you felt, if the nation felt that a larger number of outliers was a more revealing information you could certainly change the interval estimate for reporting purposes in order to identify more outliers.

DR. HAN: We welcome suggestions if you have any ideas.

CO-CHAIR HALL: And just to clarify it looks like in the materials submitted that there were about 1.2 percent high/good outliers and 1.5 percent low/bad outliers. So we're in the 1.2 to 1.5 percent of institutions being labeled according to the specifications you just heard.

DR. SUTER: And may I also add that each hospital receives a hospital-specific report for the currently publicly reported readmission measures. This is Lisa Suter.

And in that report they receive detailed information about all of their patients in the measures.

So, while there's a lot of focus on who's an outlier on Hospital Compare, there is still a tremendous amount of detailed information reported back to hospitals.

And I know STS I'm sure has

similar reporting back to their hospitals, detailed information about patients in the cohort, who was included, who was readmitted. This is information that's not available to hospitals because a large proportion of patients are readmitted to hospitals who did not perform the CABG. So this kind of information is incredibly valuable for quality improvement purposes even if there is not a distinct large number of outliers on a publicly reported website.

DR. HAN: Hi, this is Lein Han from CMS. I just want to add to that.

We also, CMS also offer the Q&A service. It's like hospital when they get the data they can call -- they can email CMS any question they have.

Yes, I don't know whose phone number I'll provide for who to call --

(Laughter)

DR. HAN: But we do have this service. So, hospital can contact us any time

they want.

CO-CHAIR HALL: Larry?

DR. GLANCE: So, I have a comment on usability. I think we both recognize that we have two very, very high-quality measures but they're essentially looking at the same thing.

And the issue that I see as a potential issue is that although all of us understand in this room that risk adjustment isn't perfect and that depending on which risk adjustment model you may end up coming to different conclusions, I'm not sure that all the consumers of this information will understand that.

And I think it goes to the heart of credibility of performance measurement when you potentially release in the public domain two different report cards which significantly disagree on which hospitals are identified as high quality and which ones are identified as low quality.

And I would urge caution to the developers, CMS and STS, that when you decide which quality measures to release to the public that maybe you agree on releasing one set of measures as opposed to both for the hospitals where you have overlap.

DR. HAN: Hi, this is Lein Han, CMS. We think these CABG is a very important area. We all recognize that. And we think that these two models are pretty good. Two very good measures.

I think the consideration for CMS is what is the most cost-effective and less burdensome way for hospitals to implement the measures. And that's really our consideration right now.

For us claims is the most cost-effective way to implement input in measure. So, that would be our priority to have a way to implement most feasible to both CMS and to hospitals.

CO-CHAIR HALL: Any additional

concerns? Thoughts?

We know that the claims-based measure has been written into IPPS for 2017 and the STS one has not. That's my understanding. I don't think that affects our decision about whether the measure in front of us is useful but we -- I think it does relate to Larry's comment and in fact relates to Wes's as well. Larry's and Wes's comments both related to comparison between two measurement programs that could give some different results.

In fact, we know the registry-based program seems to identify at least twice as many institutions on the tails as the claims-based program. So there's a danger for what Larry's concerned about.

But in this case the plan is to implement. We know this measure is written into IPPS for 2017 and again, that I don't think affects the overall assessment of this measure or should affect it.

Any other concerns or questions before we vote usability? I don't see any.

MS. SHAHAB: Voting for usability and use, 1 high, 2 moderate, 3 low, 4 insufficient information. Your time begins now.

We have all the votes for usability and use. Three high, eighteen moderate, one low and zero insufficient information.

CO-CHAIR HALL: Okay, before we vote overall any additional concerns or comments? Summary comments? Wes, you're just smiling.

DR. FIELDS: Bruce, I'm just so happy to be here to participate in the process. Thank you so much.

(Laughter)

CO-CHAIR HALL: Why do I feel like Wes is coming after me. Any final comments? Not seeing any.

MS. SHAHAB: Voting for overall

suitability for endorsement, 1 yes, 2 no. Time begins now.

We have all the votes for overall suitability for endorsement for measure 2515 Hospital 30-day All-cause Unplanned Risk-standardized Readmission Rate Following Coronary Artery Bypass Graft Surgery, 21 yes, 1 no.

MR. AMIN: So, before we move onto the next measure I just want to remind the committee and the developers that these recommendations for endorsement are still contingent on a conversation related to competing measures.

So this measure and the STS measure will be discussed in terms of how they're, you know, whether they're competing and whether it's justified to have both measures in the portfolio.

And that is in addition to this SNF measure, SNF readmission measures that we discussed yesterday. We likely won't have

time for that discussion during today's in-person meeting but we will have a follow-up call to discuss that.

CO-CHAIR HALL: I've been trying to resist going down that road but your comment makes it irresistible to ask this question though.

We just talked about two measures but we know one of them is written into IPPS. So does that not affect that discussion around competition between those measures?

DR. BURSTIN: You know, it's a great question, Bruce. I personally feel like for this committee's sake it's really about the comparability of the measures themselves and about the questions of whether you can in fact from purely a perspective of use broadly have both of those measures out there. Will it add to confusion? Can people understand the nuances? How much does the difference in data source affect the way people may use them? So I don't think it has a particular

issue.

I think it may very well come up at the MAP certainly where they are specifically charged with looking at which measures for which programs. But I don't think it should particularly have an impact on the discussion around competing -- and in this instance it's not really harmonization. They are fully harmonized except for data source.

I think it is still a competing issue and I think Larry's comments really raised that issue significantly in terms of understanding comparability, the comments related about the 15 percent of people who aren't in STS who are in this measure.

I mean there's just many issues I think you'll have a chance to chew on I assume in a separate conference call to follow.

CO-CHAIR HALL: We thank the developers from CMS, Yale.

Are the same folks going to stay at the table for the next? Or will it be a

different team? New crew?

CO-CHAIR KAPLAN: Okay, this is measure number 2513 Hospital 30-day All-cause Risk-standardized Readmission Rate Following Vascular Procedures. The developer is Yale.

Could you please briefly introduce yourselves and then the measure. Is anyone on the phone? No, everyone is here in the room. Excellent.

DR. MCNAMARA: Hi, I'm Bob McNamara. I'm a cardiologist at Yale. Jeptha Curtis is next to me, another cardiologist at Yale. Susannah Bernheim, also on the team is behind us here and we have multiple people on the phone including Lori Geary who's a part of the team.

I understand you have the whole measure in front of you. I just wanted to have a few -- to give a few highlights regarding this measure.

It's a very important measure. Vascular surgery and readmission was

identified in the MedPAC report as one of the seven conditions that were responsible for up to 30 percent of the preventable readmissions. The high cost, high readmission, high variation right along with the information that providers and hospitals and physicians need for quality development and patients need for choice. So that was the first one.

The second highlight is going into this measure we knew it was going to be very complex. We knew we would need a lot of clinical input both on our team and within our technical expert panel, the technical expert panel which was highly competent and engaged in the whole process involving multiple different specialties that are going to be affected by this measure, vascular surgeons, interventional radiologists, interventional cardiologists as well as experts in methodology, policy and patient advocate. And they were involved from the beginning for many if not all of the major decisions.

A third also regards the complexity. As opposed to some of the other measures this is going to have many different procedures. So, identification of the procedures and ultimately the patients was going to be very critical.

We developed a few guiding principles right from the beginning to identify which procedure should be included.

First, it was going to be a major procedure that was going to be involved. We didn't want to include venal punctures, arterial catheterizations and things like that.

It had to be clinically coherent. Initially MedPAC called it other vascular meaning didn't want cardiac, didn't want intracranial.

We also made the decision not to include hemodialysis catheter-related thrombectomies and the like.

And the third criteria was it had

to be central to the hospitalization. We didn't want the vascular surgery to be a suturing of an artery from another surgery. So those were some of the guiding principles.

Finally, another major highlight regards the risk adjustment that we used. The typical hierarchical model including both the patient characteristics as well as clustering of patients within a hospital.

In addition to the patient characteristics we wanted to include the different procedures. So we grouped the procedures in eight different categories. They included both anatomical location at neck, thoracic, abdominal and limb as well as an unspecified.

And we wanted to be inclusive as possible, include both endovascular procedures and open. So there's many other aspects of it but just wanted to give you those highlights. And open for any questions. Thank you.

CO-CHAIR HALL: So we apologize at

the table. We've been doing a little bit of whispering while Robert was talking. We apologize for that.

I was an expert on this measure for Yale and so I'm going to recuse myself from this discussion and that's what we've been whispering about. So I'll turn over to Sherrie.

CO-CHAIR KAPLAN: Okay. So, I did not review this measure and the other reviewer, Paulette, is not also with us today. So I am going to be looking at Bruce's --

MR. AMIN: There are a number of workgroup members --

CO-CHAIR KAPLAN: Who were on the workgroup.

MR. AMIN: Yes.

CO-CHAIR KAPLAN: So I'm going to look at Bruce's notes as best I can and the rely -- who was on the workgroup? Hands? Okay, so at least some people here have -- I did not review this measure so I will look at

Bruce's notes as best I can. And then we will count on the workgroup members to kind of pitch in here.

Okay, so with respect to the evidence, comments from the workgroup?

MS. KHAN: So the workgroup 1 members were John Bulger, Bruce Hall, Mae Centeno, Ross, Paul, Larry, Cristie and Paulette.

CO-CHAIR KAPLAN: Comments? Larry?

DR. GLANCE: The evidence is very strong for this measure.

CO-CHAIR KAPLAN: Other comments from anybody else on the workgroup? Okay, I guess we're ready to vote.

MS. SHAHAB: Voting for 1(a) evidence. One is yes, two is no and your time starts now. We still need two more votes.

We have all the votes for 1(a) evidence. Twenty yes, one no.

CO-CHAIR KAPLAN: Okay.

Performance gap. Larry, do you want to speak to that?

DR. GLANCE: So there's good evidence of a performance gap. Between the 10th percentile and the 90th percentile the risk-standardized readmission rates were 12.3 percent versus 14.9 percent respectively. So, about an over 2.5 percent absolute difference.

CO-CHAIR KAPLAN: Yes, the interquartile range was 12.9 to 14.3. So, but the range looks like 10 to 18 percent from the highest to the lowest. So there appears to me as well to be compared to some of the other measures we've seen a performance gap. Paul?

DR. HEIDENREICH: I agree there's a gap. It doesn't seem as large to me as some of the other gaps. But that still leaves moderate.

CO-CHAIR KAPLAN: So I think perspective is everything. I saw one where the interquartile range was 0.9 percent. So it kind of depends. But at least it's in the

range I think of the ones that we've seen. Any other comments? Ready to vote performance gap?

MS. SHAHAB: Voting or 1(b) performance gap. One high, two moderate, three low, four insufficient. Time begins now.

We have all the vote for 1(b) performance gap. Four high, seventeen moderate, zero low, zero insufficient.

CO-CHAIR KAPLAN: Priority. Larry, do you want to -- anybody from the workgroup want to say anything?

DR. FIELDS: I'd just say this falls again -- this is one of MedPAC's seven conditions which account for 30 percent of all readmissions in the Medicare program. So it's a high priority.

CO-CHAIR KAPLAN: Thank you. Anyone else? Voting priority.

MS. SHAHAB: Voting for 1(c) high priority. One is high, two moderate, three

low, four insufficient. Time begins now. One more vote.

We have all the votes for 1(c) high priority. Sixteen voted high, five voted moderate, zero low and zero insufficient.

CO-CHAIR KAPLAN: Scientific acceptability. First up is reliability. Larry, do you have comments?

DR. GLANCE: So this is 30-day all-cause unplanned readmissions. This was done using hierarchical modeling as per the standard approach.

They adjusted for age, sex, demographics, procedures and clinical covariates using the hierarchical condition categories, a fairly standard approach.

In terms of reliability testing the standard approach yielded an intraclass correlation coefficient of 0.4 which is very much in the zone, maybe in the upper level of that zone.

CO-CHAIR KAPLAN: Thank you.

Others? Are we ready to vote reliability? Any other comments? Okay.

MS. SHAHAB: Voting for 2(a) reliability. One is high, two moderate, three low, four insufficient and the time begins now.

We have all the votes for 2(a) reliability. Two high, nineteen moderate, zero low and zero insufficient.

CO-CHAIR KAPLAN: Now, validity. Larry?

DR. GLANCE: So, the measure developers convened a technical expert panel who expressed strong support for the face validity of this measure.

They validated this model in an independent data set. It had a C statistic of 0.67 which is at the upper end of the zone of acceptability for these readmission measures.

They looked at calibrations both graphically and also using a standard methodology and the model was well calibrated,

showed goodness of fit.

CO-CHAIR KAPLAN: So, the discriminate validity is sort of yet to be determined. Is that correct? So we don't have a sense of whether this discriminates well between hospitals.

DR. GLANCE: I don't recall exactly how many hospitals were labeled as high-quality and low-quality. Maybe the measure developers could address that?

DR. MCNAMARA: Yes, this is Bob McNamara. We did not do that analysis feeling that this is -- to develop the measure, the implementation can be any cut point that you want. Not to put it in more of a policy decision.

We had talked about that with the prior measure. We can address it now if the committee wants to. I don't know if there's much more to say on that.

CO-CHAIR KAPLAN: Any other comments from the working group or the

steering committee? Hearing none voting validity.

MS. SHAHAB: Voting for 2(b) validity. One is high, two moderate, three low, four insufficient. Time begins now.

We have all the votes for 2(b) validity. Zero voted high, twenty voted moderate, zero low and one insufficient.

CO-CHAIR KAPLAN: Thank you. Feasibility?

DR. GLANCE: So this is a highly feasible measure. It's based on widely available administrative data.

CO-CHAIR KAPLAN: Any other comments or questions? Nope? Ready to vote feasibility.

MS. SHAHAB: Voting for feasibility. One high, two moderate, three low, four insufficient. Time begins now.

We have all the votes for feasibility. Seventeen voted high, four voted moderate, zero low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Usability and use. Larry, do you have any comments?

DR. GLANCE: This one's a little bit more difficult to comment on. It's a new measure so we don't have too much information on the usability of this particular measure.

CO-CHAIR KAPLAN: Kathy?

DR. AUGER: I think it's a little challenging to assess usability and use if we don't know how many outlier hospitals there are. So we don't know whether it's able to really discriminate high performers from low performers. And so it just makes it challenging for me to assess.

CO-CHAIR KAPLAN: Taroon or Helen, you want to comment on when a measure is early on in the phase of development how that works?

MS. PACE: Yes. So, basically what we ask the developer to do is to do two things when it's a new measure. To write up how they think it can be used in improvement,

how it will be used for improvement and what are the plans for its use in accountability applications.

So, we can look at that section of their form or maybe the developers just want to remind people what they indicated as far as how this measure can accommodate improvement as well as plan for accountability applications.

CO-CHAIR KAPLAN: Developers?

DR. MCNAMARA: This is Bob McNamara. I think we're going to let Lein talk about that from CMS.

DR. HAN: This is Lein Han from CMS. So, I think I have to give Yale the developer, our contractor, credit. Because for this measure to keep the integrity of the measure they actually include cases from both inpatient and outpatient settings. Am I correct?

So, right now CMS is trying to figure out which program, either IQR, it means

inpatient quality reporting program, or outpatient quality reporting program, that this measure should be included for which program.

So, we are working on how to implement it. Depends on the -- our consultation with our leadership about which program it's supposed to be. But I can see that this measure could be for both programs.

Did I address your question about implementation?

CO-CHAIR KAPLAN: I remain confused. So if it's readmission to the hospital following a vascular procedure that could be done in either the outpatient or the inpatient setting what we're looking at is the readmission within 30 days of the procedure in whatever setting it's done, is that correct?

DR. MCNAMARA: Yes. This is Bob McNamara. I can address that. I appreciate Lein giving us credit. I think we gave her a headache with this decision.

We decided that really to be clinically coherent and to make sense for the clinical community to include both inpatients and outpatients many of the procedures are done as an outpatient procedure more based on the hospital characteristics or provider convenience or facilities rather than on patients.

And to try to say, okay, we're just going to do the inpatients then people could come out of the measure just by changing the setting, even if it's the same procedure. So that was the logic behind it all.

To talk about that is to come into the cohort you can have an outpatient procedure at a hospital facility. But the readmission has to be to the hospital. So it's not another outpatient procedure for the outcome.

CO-CHAIR KAPLAN: And for the lumpers and splitters among us, so vascular procedures seem like a big lumping category.

And to the extent that if you find as these things roll out that some of these vascular procedures look and behave different from other vascular procedures is there a plan when you're thinking about use in trying to categorize smaller clumps?

DR. MCNAMARA: Well, I think that's always a question of, as you said, lumping and splitting, of how do you want to do it.

The MedPAC had lumped them together and I think that we thought that many of the service lines within the hospitals, how it's set up is such that one entity could cover them all.

Certainly in the future some people could pull out different ones but I think the way practice is currently that it made the most sense for us to include them all.

CO-CHAIR KAPLAN: Thank you.

DR. ROBERTS: Is there a

difference in risk for those vascular procedures in inpatient versus outpatient?

DR. MCNAMARA: Yes, in a word. There's differences across which procedures you have. But the feeling was that it's not necessarily, it was, you know, the patient picking the patient's work, that lower risk that would be done as an outpatient, not necessarily that the facility as an outpatient was what was causing them lower risk. So for the patient level they were coming in as a lower risk to be done as an outpatient.

But if a hospital has to choose whether they do it as an outpatient or as an inpatient I wanted to include them all because it's really based upon the patient, not upon the facility or location.

CO-CHAIR KAPLAN: Larry?

DR. GLANCE: Does your risk adjustment model include an indicator for whether the procedure was performed as an inpatient versus outpatient?

DR. MCNAMARA: No. The way that the cohort is developed was based upon coding. And the inpatient codes are ICD-9 and the outpatient codes are CPT. So, you could develop it, you know, you could identify that from there.

But again, basically because of feeling that patients -- an individual patient will get their procedure, or could get a procedure as an inpatient and outpatient based upon a facility rather than based upon patient characteristics, it wouldn't be appropriate to adjust inpatient versus outpatient. You're trying to adjust it based upon the procedure being done.

CO-CHAIR KAPLAN: Larry, did that answer your question?

DR. GLANCE: Just a follow-up question. Did you look at whether or not there was a tendency for patients who were -- procedures that were performed as outpatient procedures, for the same procedures to be

readmitted more often if they were performed as an outpatient versus an inpatient?

In other words, the idea being that the inpatient procedures, they're already admitted to the hospital, whereas for the outpatient procedures you're sort of -- in a way you would think that those patients are slightly more likely to be readmitted because of being sent home more quickly.

DR. CURTIS: So this is Jeptha Curtis. I can comment a little bit on that specifically.

So, the question is whether or not there's a downside to outpatient procedures. And when we say outpatient procedure we're not really talking about necessarily patients who are going home the same day. Oftentimes we're talking about procedures that are being performed on an observation stay basis as opposed to an outpatient stay. And that's pretty much the major rationalization for including them.

So, most of the patients who are outpatient or observation stay have the exact same hospital utilization. They're in overnight and they go home the next day. And this is almost all endovascular procedures being performed on the neck or in the legs.

And that's, you know, everything is the same except for whether or not the hospital administrator characterizes it as an inpatient or an observation stay basis. So there's really no other information that comes with that.

The lower risk of readmission associated with that observation stay population really is driven by the fact that there are low-risk populations no matter what the designation is. And that's why we adjust for the procedure, not for the setting.

CO-CHAIR KAPLAN: Thank you. Any other comments?

DR. MCNAMARA: I'm sorry, it's Bob McNamara. Just to add on that that there's

not a 1 to 1 correlation between the CPT codes and the ICD-9 codes. You can't exactly say. So certainly within -- there's many more CPT codes, but they can be adjusted to different ICD-9s. Or correlated.

CO-CHAIR KAPLAN: Thank you. Paula?

MS. MINTON-FOLTZ: Can you tell me if this excludes same-day transfers? So if I came in as an outpatient procedure and was immediately admitted afterwards would that count as a readmission?

DR. MCNAMARA: I believe it would not. We included that it was a 24-hour difference in the level -- or the date of the procedure and the admission.

And I think most of those within Medicare rules would be listed as an inpatient procedure, even if -- as Jeptha said, many of these are done in the same area. It could be done in the same operating suite, whether it's an inpatient or an outpatient officially.

And then even if they were intake to the surgery as an outpatient, if they decided to change that as an inpatient from the back end Medicare it would look like an inpatient, not as a readmission.

CO-CHAIR KAPLAN: Thank you. Karen and then Karen.

DR. MCNAMARA: That was Bob McNamara again.

MS. PACE: Right, this is Karen Pace. I just wanted to make a comment that we really wouldn't want to include a risk factor related to where the procedure took place because that may be one of the things that might be a difference in the care provided.

So it may be something that's useful for drilldown and quality improvement when you're looking at your data and what patients are being readmitted. But it generally wouldn't be something that would be considered for a risk factor.

CO-CHAIR KAPLAN: Karen?

DR. JOYNT: I just want to say something kind of similar which is just to commend you for going to what I'm sure was a lot of trouble to put the outpatient things in. I think that is hugely important, especially as care sort of shifts place to really think about quality spanning across different settings.

I just have more of a technical question maybe for NQF people which is with this measure, compared to others who really don't have a clue for how it's going to work in terms of the outliers. Is that something that we are expected to know as we think about whether or not we feel that the measure is appropriate?

Or do we just sort of say it doesn't matter if it identifies 4 percent or even 25 percent as outliers, that's separate from the measure itself? Any guidance would be helpful.

MS. PACE: So, I guess we could

look at the -- the information that they provided was just kind of the distribution of the scores. And we could look at that.

But I don't think that has to be your defining decision. You may want to not say high on usability and use because that's a question in your mind, but the real question is whether this has the potential at this point to be useful for improvement and accountability.

And you know, one of the things when it comes back for endorsement maintenance will be to have some real data on how that has played out. So I don't think it's an ultimate defining decision for use and usability.

CO-CHAIR KAPLAN: Yes. So the way I understand this process. Correct me if I'm wrong, really quickly. If we endorse this for use then it goes out and they get the information, Karen, that you would be looking for within 3 years or they don't. And when it comes back for re-approval then we reconsider

whether or not the information is suitable for re-endorsing this measure? Does that help?

DR. JOYNT: One more question. With all these measures that are again sort of a longer time frame and the information is not fed back to the hospital quite as quickly as might be optimal for quality improvement is that again something we should consider in approving a metric, or something that goes to ways that we hope that all these measures get used better in the future?

Because I think that's a limitation that cuts across a lot of these. It's nothing to do with the statistical power of the model, or the way that it's set up, or the way that you've chosen procedures. It's just I think it's a real problem if we're supposed to also think about the usability for improvement. So how should we think about that?

CO-CHAIR KAPLAN: That's also come up before. I'll let you go ahead.

DR. BURSTIN: This has come up in multiple discussions for us, particularly around the readmission measures and the lag time to get -- the amount of time back.

Again, I think it's something you could factor into usability as you're voting. I also know it's something Lein and others from CMS have pointed out as something you're actively working on, trying to -- maybe Lein wants to respond.

But I know there have been active efforts to see if there are more ways to get information back to hospitals more quickly. Lein?

DR. HAN: This is Lein Han from CMS. We got this feedback all the time from the hospitals. And it's understandable that they do need most data for quality improvement.

So, what we're working on is that we're not providing the risk-adjusted rate quarterly, but we would like to see if we can

just get the data to them. But the raw data, not really the calculated data. So when they have the raw data at least they can look at the cases.

So, and we plan to do this quarterly, hopefully that we can get to the hospital quarterly this type of data. But raw data like we provided in hospital-specific report, those cases. Thanks.

CO-CHAIR KAPLAN: Thank you. Karen, one of the issues that's come up before is that in that backwards look for three years time, for example, the tensions between getting a precise estimate and so you get more cases means that you lose on the other end in terms of usability for quality improvement issues. So there are some tensions and tradeoffs in these different kinds of calls.

DR. BRIGGS: So in the reporting is this going to be reported out by the anatomical buckets? Or is this going to be vascular readmissions altogether?

DR. MCNAMARA: As the measure specifies right now there will be an overall vascular readmission.

Whether in the future it can be done on a procedure level for some of the high-volume procedures or in different buckets can be done based upon how it's set up. But the measure was developed as an overall.

DR. CURTIS: Just to follow up on this. Jeptha Curtis. I think the way that you could use it, I think you'd report out all vascular readmission rate and that's useful for public reporting.

To the hospitals we could try and create buckets that make it more usable for them for actually driving quality improvement processes so they can know where they are maybe not in a risk-adjusted fashion but at least what's driving their hospital-specific readmission rates.

CO-CHAIR KAPLAN: Okay, we're coming up on time. I don't want to cut this

short. Any other comments or questions? We're ready to vote on usability.

MS. SHAHAB: Voting for usability and use. One high, two moderate, three low, four insufficient information and the time begins now.

We have all the votes for usability and use. One high, eleven moderate, four low, four insufficient information.

CO-CHAIR KAPLAN: Thank you and I think we're ready to vote on endorsement. Any comments?

Okay, ready to vote.

MS. SHAHAB: voting for overall suitability for endorsement. One yes, two no. Time begins now. Just one more vote. Can you please just press your votes one more time?

We have all the votes for overall suitability for endorsement for measure 2513 Hospital 30-day All-cause Risk-standardized Readmission Rate Following Vascular Procedures. The votes are 14 yes, 6 no.

CO-CHAIR KAPLAN: Thank you very much to the developers for coming and for CMS coming as well.

And we finished within three minutes which is measurement error in my view on time. So, excellent. We we have a break until 10:15. Thank you.

(Whereupon, the foregoing matter went off the record at 10:03 a.m. and went back on the record at 10:13 a.m.)

CO-CHAIR KAPLAN: Can we ask our developers to briefly introduce yourselves? And make sure when you're commenting that you state your name and briefly give us a two-minute brief discussion of the measure.

DR. NAKAMURA: Thank you. My name is Mari Nakamura. I'm a pediatric infectious diseases doctor and health services researcher at Boston Children's Hospital.

DR. ZASLAVSKY: I'm Alan Zaslavsky. I'm a statistician at Harvard Medical School.

DR. NAKAMURA: And as you heard joining us on the phone is our principal investigator for our center Mark Schuster who's joining us from Vancouver today.

Measuring and reducing readmissions has become a widespread focus in pediatrics, but to date no readmission measures developed specifically for use in children and adolescents have been publicly available.

We were therefore assigned to develop readmission measures by CMS and AHRQ as part of their pediatric quality measures program for which we serve as the center of excellence.

Hospital readmissions within 30 days occur for 2 to 6 percent of children. These rates are certainly lower than the rates of about 20 percent that we often hear for Medicare beneficiaries over age 65, but overlap with rates for adults under age 65.

As a point of comparison pediatric

30-day readmission rates are equivalent to pediatric inpatient adverse drug event rates.

Hospitals, payers and other stakeholders are actually already actively working to reduce pediatric readmissions even in the absence of a publicly available measure.

Our all-condition measure evaluates readmissions following an index hospitalization for almost any condition.

We were encouraged by CMS to develop an all-condition measure to correspond with the adult measure that they've now rolled out. And in addition, our national stakeholder panel supported an all-condition measure because it includes the broadest range of children and hospitals.

Furthermore, we found that very few specific pediatric conditions are common enough to serve as a focus of a readmission measure.

We've prioritized harmonizing our

measure with the NQF-endorsed adult readmission measures while still making it appropriate for pediatric use.

One important way in which our measure corresponds with NQF-endorsed adult measures is that we choose to focus on evaluating unplanned readmissions. Based on our own research as well as other studies we don't think that the preventability of readmissions can be assessed using billing codes.

The main data set we used to develop and test the measure consisted of Medicaid claims for about 400,000 hospitalizations at 2,000 hospitals in 26 states.

We also used AHRQ HCUP all-payer data from 2 states and NACHRI case mix data from 72 children's hospitals.

We developed a case mix adjustment model for the measure that includes patient age, gender and chronic conditions on the

index hospitalization and we found that the model performed similarly to those used in other readmission measures with regard to discrimination and calibration.

A challenge not just for calculating pediatric readmission rates but for all pediatric quality measurement is small sample sizes at some hospitals with resulting low reliability of measure scores.

However, because pediatric patients are not distributed across as many hospitals as adult patients we found that the majority of pediatric hospitalizations occur at higher-volume hospitals whose readmission rates have good reliability.

Because the measure uses claims data that are already collected for other purposes we anticipate that implementing it will be highly feasible.

We believe that the measure fills an important need for publicly available readmission measures and think that it could

serve as a valuable tool to assess health system quality and motivate improvements in pediatric care delivery. Thank you.

CO-CHAIR KAPLAN: Thank you very much. Kathy?

DR. AUGER: So, to speak to evidence. Certainly there isn't as much evidence around pediatric readmission as there is in the adult world.

However, certainly it meets the overall construct that this could be an important measure. So I think it's high in that sense.

CO-CHAIR KAPLAN: Karen, do you have anything to add? Are we ready to vote evidence? Any other discussion? Go.

MS. SHAHAB: Voting for 1(a) evidence. One is yes, two is no. And time begins now. We need one more vote, please.

We have all the votes for evidence. Twenty-one yes, one no.

CO-CHAIR KAPLAN: Thank you.

Performance gap. Kathy?

DR. AUGER: So, as the developers note the prevalence of pediatric readmission rate is between 2 and 6 percent.

What I was just looking for and couldn't find is what the range in interquartile ranges for the risk-standardized rate. Do you guys have that available?

DR. NAKAMURA: This is Mari Nakamura again. To give you a sense for an all-condition measure based on the variance component of the hospital random effect in our mix model a hospital that's two standard deviations below the mean would have a readmission rate of 2.4 percent whereas one that's two standard deviations above would have a readmission rate of 10.3 percent. So about 4 times greater.

CO-CHAIR KAPLAN: Karen?

DR. JOYNT: We're actually looking for this information together. And I think that there's actually impressive performance

gap when looked at that way.

The graph -- I didn't find a distribution on this. The distribution in the Berry paper from last year suggests that most hospitals actually don't fall that far outside the mean. So do you know the 25th and 75th percentiles?

DR. NAKAMURA: I don't have that with me, no.

DR. JOYNT: In that one it looked like there were really very few that were past about between 5 and 7. But that may be because it was the NACHRI hospitals.

DR. NAKAMURA: This is Mari Nakamura again. So, it's a good point that in that study it was a quite homogenous set of hospitals. And so you might expect that there wouldn't be as much difference among them in terms of the range of readmission rates.

CO-CHAIR KAPLAN: I might point out that the magnitude of the difference is about in the range of other measures that

we've looked at before. In fact, it seems to be somewhat broader than some of them that we looked at.

Other comments? Ready to vote performance gap?

MS. SHAHAB: Voting for 1(b) performance gap. One is high, two moderate, three low, four insufficient and the time begins now. One more vote, please.

We have all the votes for 1(b) performance gap. One high, twenty moderate, one low, zero insufficient.

CO-CHAIR KAPLAN: Great. Priority?

DR. AUGER: So, the developers mentioned cost of readmissions at six months which is certainly a longer window than the measure in front of us which was $136 million.

So, of course pediatric costs are not anything what adult costs are. Having said that, they do also present data on disparities that exist in pediatric

readmission, including race/ethnicities disparities and payer disparities. So that goes certainly to priority.

And then also of course as they mentioned this would be the first pediatric metric. So that again speaks to priority.

CO-CHAIR KAPLAN: Karen? Other comments? Ready to vote?

MS. SHAHAB: 1(c) high priority. One high, two moderate, three low, four insufficient. Time begins now. One more.

We have all the votes for 1(c) high priority. Seven high, thirteen moderate, two low, zero insufficient.

CO-CHAIR KAPLAN: Thank you. Scientific acceptability. We'll talk about reliability first. Kathy?

DR. AUGER: Sure. So, as the measure development experts had mentioned previously they used ICC to assess reliability. And it's very much dependent on the volume of cases seen at hospitals.

So, with the reliability of 0.5 there were only 607 out of the 2,011 hospitals that had an ICC greater than that but that did a count for 88 percent of the index hospitalizations. So the majority of the hospitalizations are at higher-volume hospitals which had the higher reliability with the low-volume hospitals are the issue.

CO-CHAIR KAPLAN: Karen? Other comments? Ready to vote reliability?

MS. SHAHAB: Voting for 2(a) reliability. One is high, two moderate, three low, four insufficient and the time begins now. One more.

We have all the votes for 2(a) reliability. Three high, seventeen moderate, two low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Validity. Kathy?

DR. AUGER: So, this is an examination of unplanned hospitalization or readmissions. So the developers have a

somewhat novel way of determining planned or unplanned which is worth mentioning.

The way that they did that was through expert opinion panels of which codes could be consistent with a planned procedure. So they also went through some validation of that algorithm in and of itself using chart review at Boston Children's which seemed reasonable.

Then in terms of the risk adjustment validity they use the number -- age and the number of chronic conditions as well as gender.

There was some concern in the public reporting comments that they hadn't used primary diagnosis so that might be something worth just asking about.

And then in terms of how the model performed the C statistic was 0.69. In terms of calibration there was good observed-to-expected graphs.

And then the other question I had

was again about how many hospitals or outliers, whether or not they have done that assessment.

And finally, other threats to validity would just be in terms of missing data because of the MAX chart. The MAX data system in and of itself has a lot of issues with handling admission data.

CO-CHAIR KAPLAN: Karen, I'm going to ask you to hold on a second and let the measure developers respond.

DR. NAKAMURA: Thank you. Regarding the question of --

CO-CHAIR KAPLAN: State your name, please.

DR. NAKAMURA: Sorry. Mari Nakamura. Regarding the question of including the reason for admission, the primary diagnosis, this was something that we thought a great deal about and had done some exploratory analysis of and ultimately concluded didn't make sense to include.

One reason is that we found that actually performance in a given hospital tended to track across different types of diagnoses so that it does make sense to be able to aggregate all of the patients into a single measure.

We also found that there's a challenge in using pediatric diagnosis codes with a good grouping system. We even experimented with trying to devise one of our own.

Because patients don't have just a sort of few common diagnoses but really in pediatrics diagnoses are quite variable we found that for any given category that for some hospitals there were problems with cell sizes.

I don't know if Alan has anything further he might want to say? No? Okay.

Regarding the question of outliers what we had provided in our submission was one way of looking at outliers which is to use the

method that's currently used for the adult measure pay-for-performance program.

And using those methods we found that for the all-condition readmission measure that about 47 percent of hospitals have a higher than expected readmission rate. So their predicted readmissions exceed their expected.

And that the median readmission rate for those hospitals that were above 1 in their ratio was 1.15 suggesting that they have about 15 percent in excess in terms of the median readmissions.

We haven't looked at the question in terms of outliers using confidence intervals, but recognize that there are different ways that one could apply the measure and choose to identify outliers.

And then I think your last question was about missing data in MAX. So we definitely acknowledge that MAX is a very messy data set as we learned once we delved

into it.

For those of you who may not be as familiar it's assembled by collecting Medicaid claims from all of the 50 states and the District of Columbia and then trying to make them into a uniform data set for researchers to use.

And as you might imagine that process is a difficult one. It seems to be definitely improving over time. But that is one reason, for example, that in our MAX data set we felt that not all states had good enough data quality for key variables such as hospital identifiers to be able to include them.

So all that said while our test data set was the MAX data and we were actually pleasantly surprised that the percentage of records that had to be dropped based on data quality or completeness issues was actually 10 percent.

That doesn't necessarily mean that

the real data sets that would be used, meaning the actual claims data available to, for example, state Medicaid agencies would have as many issues with missing data as the MAX data.

CO-CHAIR KAPLAN: Thank you. We're going to go Karen and then Frank.

DR. JOYNT: Yes, just had a few additional questions to the ones that were brought up before.

I think one of the threats to validity is just that there aren't other pediatric measures with which to compare this one. And so it's a little bit difficult to know exactly what we're measuring.

Certainly hospitals can differ on things like socioeconomic status and access to care for kids. And that may be the difference that we're seeing driving this. It's hard without having anything else that we would sort of consider to be "quality" to compare this to to know exactly what we're measuring here.

I think that's probably the case with most of the readmission metrics to some degree but it's particularly problematic if we don't have comparisons. This is not the developer's fault and kudos to you for trying to develop a quality metric in what is often a very data-free zone. But I think it is an important threat to validity.

The two other -- well, I guess the one other question is really if you could just explain this a little bit, how this model is similar or different to the ones that we're used to hearing for the adult metrics. Just so that we are clear on whether or not this is the same method as is being used for the other measures or if it's different and in what ways it differs beyond the exclusion of procedures. Or, sorry, planned admissions.

DR. NAKAMURA: Thank you. This is Mari Nakamura again. Karen is absolutely right that a challenge we faced in trying to assess the validity of our measures is the

fact that there are not other pediatric inpatient measures we could use, or widely available data sources for which we would be able to even evaluate such measures.

We agree that this is a really important question and one that we think will need to be evaluated as more measures are developed. We felt like it was good to start somewhere and acknowledged that this is a limitation currently in our field in pediatric measurement.

Regarding the question of how our measure compares to the adult measures in terms of our statistical approach. Overall the approaches are really similar in that we use hierarchical modeling.

We do as a result have the shrinkage effect that I've heard discussed quite a bit here. And we've talked about both the advantages of that.

It is relevant to pediatrics because of course we do have many small-volume

hospitals in terms of pediatric volume.

One difference with the all-condition measure is that the all-condition measure for the adult Yale measure uses five different service line models and then combines the outputs of that to end up with a single readmission rate.

We considered such an approach but our worry was that at many hospitals given that pediatric volumes overall are low that we would then have trouble with the sample sizes for splitting our sample among different models.

Another difference which Alan may want to speak a little bit more to in terms of the statistical implications is that we use direct rather than indirect standardization.

Meaning that in our approach of standardizing we hypothesize what the rate would be at a given hospital assuming that the entire cohort of reference data set was cared for at that hospital. In the indirect method

instead it's this approach of using predicted to expected readmissions.

But the end result again is that it tends to pull the small-volume hospitals closer to the mean. So in that way the outputs are similar.

DR. ZASLAVSKY: Alan Zaslavsky. I would just add that while the direct and indirect standardization look on the face of it pretty different the underlying models actually work out to be pretty similar.

The direct standardization we're doing uses the logistic model. Indirect uses usually a ratio of observed to expected which implies a multiplicative or log linear model. But in the range we're talking about the two models are pretty close to each other. So, that doesn't make a big difference.

And the shrinkage effects as Mari said are handled in pretty similar ways. The sufficient statistic for the performance of a particular hospital is essentially the total

number of readmissions. And that's the same in both models. So the two should give pretty similar results.

We're another group. We set things up a little differently in a way that we found to be a little bit more direct, not just because it's direct standardization but because it's all done on one model. But I think the two are similar enough that we wouldn't have found anything terribly different if we'd done it exactly the way Yale did it.

CO-CHAIR KAPLAN: Thank you. Any other comments? Frank, and then Leslie, and then Tony.

DR. BRIGGS: So, two quick questions. First just being a definition. You said you excluded specialty hospitals. I was wondering if that was the same cancer hospitals that you see in the adult realm.

And then the other was hospitals, if the readmission was in the discharge was

readmitted to an area out of the state because of limitations on the data set. Although you might have some data I was wondering how often that happened, especially for border states and rural care and things of that nature.

DR. NAKAMURA: So for the question about specialty hospitals, in pediatrics that designation tends to capture, for example, hospitals that do deal with specific conditions. Cancer hospitals are one. Another would orthopedic hospitals.

The other sort of major category is more -- and another category, excuse me, is the Shriner's hospitals that deal with burns and trauma.

And then another sort of group of hospitals that we excluded is hospitals that don't provide acute care, that are more long-term care such as for rehabilitation.

For the question about what percentage of readmissions are to an out-of-state hospital, so the good thing is that

because Medicaid claims go back to the state of residence of the patient there's actually complete data for the patients who are in a given state about where they were readmitted.

Where there is a challenge is for the states that weren't in our data set. Because for those we have probably a minority of information. We would only know about he claims that happen to come through the particular states that were in our data set.

And so that's why we exclude the hospitals that are outside of the states that we include in our data set. But for those that are in the data set we would be able to tell if patients are admitted outside of their home state.

CO-CHAIR KAPLAN: Thank you. Leslie?

MS. HALL: So, I have a question. Where you might have two hospitals in the same community, one hospital is known as a brand for children's hospital. It's not a

children's hospital, it's just where most children go.

Both hospitals would be very high-volume community hospitals. But one would display as a low-volume hospital even though it wasn't in this case. Does this create any challenges where we might have the same sort of problem we've mentioned over and over again about low-volume hospital size and low-volume overall when we just simply have a patient mix that's very different in an otherwise high-volume hospital?

DR. NAKAMURA: This is Mari Nakamura. I want to make sure I understand your question.

So you're wondering about the issue of community hospital that has done pediatric patients but overall not a very high volume of pediatric patients compared to another hospital?

MS. HALL: That's correct.

DR. NAKAMURA: So, this is

certainly a challenge about trying to measure pediatric readmission rates or any sort of hospital-based quality measure. You're right that there are big differences in the types of hospitals that care for children.

What we've found is that children tend to be really concentrated at relatively small numbers of hospitals. So they tend to be high-volume, they tend to provide a full range of care. And then the community hospitals tend to provide care for relatively small numbers of children.

To give you a sense, children's hospitals make up about 5 percent of all of the 4,000 hospitals in the country but actually care for about one-third of pediatric inpatients.

And so I think where this is a challenge is that as we've alluded to for these community hospitals on the one hand we didn't want to ignore them. But we also acknowledge that the reliability of their

readmission rates is limited because of the low volume.

And so one way that we think these hospitals could still be included is of course to be very responsible about explaining the limits and what we actually know about them. And also perhaps to compare like hospitals with like.

CO-CHAIR KAPLAN: Thank you. Tony?

DR. GRIGONIS: Yes, I just have a quick question about your databases from you said 26 states I believe. Were they consistent in terms of the state populations? That's the first question.

The second follow-up would be did you see any differences in states that have higher populations.

DR. NAKAMURA: So to answer -- this is Mari Nakamura again -- to answer the question about population, the states actually varied quite a bit in population. Because we

found that the states with the best data quality fortunately were geographically distributed. And so some of them were lower population states and other high.

We didn't actually do an analysis to evaluate how the volume of the state related to readmission rates. That's something we certainly would be able to do.

CO-CHAIR KAPLAN: Thank you very much. Other comments?

I would remind the group that as with the previous measure on all-cause readmissions following vascular procedures this would be a new measure. So if endorsed the 3-year period would ask that the developers generate some of the data that have been raised and issues of concern and so on.

So if approved it would be a new measure that we would ask the developers to consider some of these -- response to some of these kinds of issues.

Are we ready to vote validity?

MS. SHAHAB: Voting for 2(b) validity. One is high, two moderate, three low, four insufficient. Time begins now.

We have all the votes for 2(b) validity. Zero voted high, nineteen voted moderate, three low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Feasibility. Kathy?

DR. AUGER: Certainly these are claims data so potentially feasible.

I think it is worth mentioning, as Mari already did, that what's in Medicaid claims does vary from state to state. So that's a little bit of an issue.

Having said that they did test the measure on the New York database and were able to see a little bit of model fitting issues with rare values. But then they were able to offer some troubleshooting to address any issues of feasibility. So I think that's a strength.

CO-CHAIR KAPLAN: Karen?

DR. JOYNT: Yes, I think feasibility is a real concern here because kids, unlike adults over 65, are covered by a whole bunch of different insurance plans.

So yes, there's a chunk in Medicaid but they certainly don't represent a randomly selected group of kids. And so understanding how this measure might act in all-payer claims databases versus a Medicaid database versus Aetna or something like that I think is something that would really need to be thought through as this is rolled out.

Are we trying to build a Medicaid quality metric? Are we trying to build a pediatric quality metric? And I think those two things are probably different. And so I think the model and the measure would probably have to take that into account given the limitations of the pediatric data and the ways in which the populations that underlie those data might actually differ quite a bit.

CO-CHAIR KAPLAN: Other comments

or thoughts?

DR. NAKAMURA: Thank you. This is Mari Nakamura again.

The issue about the variation in Medicaid claims from state to state is certainly a true one. We anticipate that one of the likely uses for the measure will be by Medicaid programs. We know that's something that CMS is interested in.

And at least initially it seems that what would be most feasible is to evaluate readmission rates within a state rather than trying to compare across states.

That said it may be that as more measures come out for pediatrics that this will perhaps drive the creation of a national data set that's available in a more timely way than the MAX data set.

We also note that there are several states that are developing or already have all-payer claims data sets that could be useful for this measure.

Again, that wouldn't necessarily allow national comparisons but at least within-state comparisons.

One thing that we tried to do with our measure and have provided the SAS program for is we've provided a way for a given state Medicaid program to be able to calculate nationally comparable rates using our MAX data set as a reference data set.

And so this is something that if a state Medicaid program really wanted to have a sense of how they compared to another state that they would have the option to use.

Regarding Karen's point about the fact that there are different insurers for children and that unlike the adult over-65 program there's not a Medicare for children. This is very true.

Medicaid covers about one-third of hospitalized children. So it is a sizeable portion. But she's absolutely right that those children are not necessarily directly

comparable to children covered by other insurance plans.

We've looked at insurance disparities and found in fact that Medicaid-insured children do have a higher risk of readmission.

We know that there has been a lot of discussion about including such sort of sociodemographic factors. And if there was interest in including something like insurance status for pediatrics we would be very interested in discussing that further.

CO-CHAIR KAPLAN: Thank you. I'm going to -- we're shading into usability here so I'm going to ask Tom and then Larry.

DR. SMITH: Well, that was my question. Is this a feasibility issue or a usability issue? Are all your data from MAX data? And are you putting this out there as a Medicaid performance measure or a children's performance measure?

DR. NAKAMURA: This is Mari

Nakamura again. As a pediatric performance measure but recognizing that one probably common use will be as a Medicaid performance measure. We were aiming to be as inclusive as we could.

An example of another data set that could be used with our measure would be an all-payer claims data set at a state level. And so as part of our measure testing we did work with the State of New York as Kathy mentioned to test the measures on their all-payer data set as well.

CO-CHAIR KAPLAN: Again, this is a usability issue. We're still on feasibility.

DR. SMITH: Well, I was going to say I now agree with Karen that there's a significant feasibility issue here based upon your response.

CO-CHAIR KAPLAN: So feasibility for application across all -- I'm concerned about your issue. Let me have you guys re-frame that as what is the actual nature of the

concern about feasibility?

DR. SMITH: I'll defer to Karen.

DR. JOYNT: Well, I just -- I think it depends what you're trying to define it as. As a Medicaid measure it is feasible for states to use because it's based on claims data.

CO-CHAIR KAPLAN: So is your concern the target of inference here?

DR. JOYNT: As a pediatric measure it's not super feasible because there are entire swaths of the population that are in no claims data set. Which makes it very difficult to apply a claims-based metric to them. So I think it depends a little bit on the population that we're trying to assess whether or not it is actually feasible to do.

If you had a kid in, I don't know, Kansas -- I don't know if Kansas is an all-payer claims data set -- covered by Aetna I'm not sure that you actually could include that child in a metric. But for Medicaid it's

highly feasible because it's a claims base. So it just depends on the group.

CO-CHAIR KAPLAN: Larry? Taroon?

MR. AMIN: It's an interesting sort of characterization of the feasibility question.

The way that we generally think about it is so if a state had an all-payer claims database, let's say California for instance, and they wanted to be able to run this measure for their pediatric patients in their state, would they be able to do that? I think that's generally the way that we would look at it. So if they had the data can they take these specifications and run them?

And so I'll just open that. And if they can I think that makes it a very feasible measure. I mean, if there's no -- if a user doesn't even have the data that's not necessarily a feasibility question for the measure developer.

So, maybe I'll just -- with that

framing maybe the measure developer can respond.

DR. NAKAMURA: Yes, thank you. This is Mari Nakamura.

So with that framing we certainly think that the measure is highly feasible, that it's actually quite easy to implement. We have very detailed specifications and we've also provided or can provide programs to actually be able to do a lot of the data preparation and running of the model.

So in our experience working with New York and in working with the programs ourselves we feel like if you have the data that it's highly feasible.

CO-CHAIR KAPLAN: Thank you. Larry?

DR. GLANCE: A quick comment though. Was the model validated in all-payer data? And if not, although it may be feasible to use with all-payer data it probably would not be a very appropriate use if it has not

been validated using all-payer data.

DR. NAKAMURA: It's true that our primary data set for development was Medicaid claims data. And so we've done the most testing absolutely in that data set.

We did also, however, use a couple of states, HCUP state inpatient database data to test the measures as well. And to be able to also in the case of New York to compare our findings with their findings on their all-payer claims data set.

CO-CHAIR KAPLAN: So now I would like to ask NQF to comment on how we frame this. Because the issue has become the target of inference here.

And if we're trying to infer what the all-cause readmission rates for pediatrics are using the Medicaid database that's a different issue -- I'm hearing that's a different issue than if you're trying to extrapolate to the entire pediatric population.

MR. AMIN: Well, I'll answer but with a question I guess. I mean, in some ways when we look at -- I mean, not to draw comparisons with other measures, but you know, the Yale CMS measures that are -- the all-cause hospital readmission measure, has been developed using Medicare claims data, then tested with California all-payer claims database. So, and it's specified as 18 and older. So the unit of inference could be any group of people that are over that population. In this case it seems very similar to me.

So I would say that -- I mean, I'm not judging this but I think you should keep that in mind as you make your decision about the feasibility of this measure.

CO-CHAIR KAPLAN: Leslie?

MS. HALL: So, just a clarification though. Didn't you already state that the research showed that there actually was a difference with the current Medicaid covered population, and that you saw

right away that there was a difference in that insured population versus the other insured population?

So therefore we're starting to game with a difference in class of service I guess based upon the insurance. And so do we have mixed messages?

DR. NAKAMURA: Thank you. This is Mari Nakamura. So yes, you're correct that Medicaid-insured children have a higher readmission risk than privately insured children, for example.

This would be one argument for why it might be important to eventually adjust for insurance status as part of the case mix adjustment.

At the same time, to go back to the analogy that Taroon made, we know that Medicare-insured adults over age 65 are at a higher risk of readmission than other adults.

And so I think as you're pointing out very well the population that you're

evaluating is a really important question. And it's important to be careful to consider who you compare.

But we don't feel that that limits the ability to use a measure either in a specific payer population or more broadly by taking into account the fact that there are differences among different payers.

CO-CHAIR KAPLAN: Okay, so I'm hearing some confusion about -- that I think probably although it does touch on feasibility for certainly the data you've run, you've run it. So you can get access to that data.

And I think the issue now is shading over into usability and that's causing a little bit of concern around here, although the feasibility issue -- does anybody else want to make a comment on feasibility? Are we ready to vote? All right, let's vote.

MS. SHAHAB: Voting for feasibility. One high, two moderate, three low, four insufficient. Time begins now.

We have all the votes for feasibility. Three voted high, eighteen moderate, one low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Now we're into usability. And does anybody have further comments on the potential usability of this issue? Kathy, you want to go first?

DR. AUGER: I think it's just worth noting the potential differences between what is an unplanned readmission, an actual preventable readmission.

I think there's a decent amount of angst in the pediatric hospital medicine community that there may not be a lot we can do broadly to prevent readmission.

Although I would personally comment it's still early in the game. It's still a little hard to actually come down on whether or not that's true or not.

There's certainly large groups trying to reduce readmission rates nationally. And frankly, the studies that they have done

haven't yet been published so I think it's still just -- it's just early to assess. But I think it's a consideration is whether or not how much this is actually preventable versus planned. Unplanned, sorry.

CO-CHAIR KAPLAN: Paul?

DR. HEIDENREICH: Along similar lines I do have significant concerns that there's not a credible rationale for improvement. I know looking through the background that there have been almost no studies of interventions but it doesn't seem like there's either been -- or maybe you can tell me if there are some studies that at least correlating certain hospital practices with better readmission rates even though there wasn't an obvious randomization or experimental design.

And without that it seems -- and given that pediatrics I think is significantly different from adult with multiple chronic diseases, a large number of medications, it

would seem that your interventions you would do would be different.

And so I'm very concerned that it's not clear how hospitals will be able to use the information once it's released.

CO-CHAIR KAPLAN: Developers want to comment?

DR. NAKAMURA: Yes, Your Honor. This is Mari Nakamura.

To start with Kathy's point about preventability versus unplanned readmissions. We agree that ideally it would be desirable to be able to assess preventable readmissions rather than unplanned readmissions.

We just don't think that that's possible using claims data, using the data that we currently have widely available to us. And so we took the approach in harmony with what has been done in adult measures to instead focus on unplanned readmissions.

That said, research that we're currently working on as well as conducted at

other institutions has indicated that a sizeable proportion of unplanned readmissions actually are potentially preventable.

Of course, how preventability is determined is highly controversial. It's something that I think will probably always be a subject of debate because it's so subjective.

We have a study that's still currently underway in which we've talked with patients, families, nurses, outpatient doctors, inpatient doctors to try to get a sense from all of them of how preventable they thought readmissions were.

What we found is the more information you get the better sense you get that it's a very complicated question. And so knowing that we are using claims data for this measure we're choosing to focus on unplanned readmissions.

Regarding the rationale for whether readmission rates can be improved in

pediatrics it is indeed true that there are far fewer studies in pediatrics than adults looking at interventions for readmission.

There are a handful that have focused on pediatric patients or that have included pediatric patients and evaluated them as a subset that have found, for example, better adherence to practice guidelines has been associated with reduced readmissions or improvements in criteria for discharge.

All that said we acknowledge that in this space for pediatric measures that more likely what we're dealing with is a rationale that makes sense rather than a lot of evidence already in the literature.

And we believe that given that processes such as discharge preparation and education, making sure that there are good transitions to the community are felt to be equally important in pediatrics that improvements in those processes could also lead to improvements in readmission rates.

The point that there are not as many very chronically ill complex patients among pediatric patients as adults is true.

That said, one of my colleagues at our center conducted a study a couple of years ago examining specifically the population of patients with chronic complex conditions who are frequently readmitted and found that while they're a small percentage of all patients they account for a high percentage of readmissions in cost.

And so in terms of a population that could be a focus for readmission improvement that seems like a natural one. Thank you.

CO-CHAIR KAPLAN: Thank you. Taroon, do you have something specific to this? Then let's go Leslie, Karen and Taroon.

MS. HALL: So, this is an interesting area in usability and maybe should be questioned or comments aimed towards implementation.

But communication and education around this measure could not be stressed more. Because readmission is something that is known in the public I guess consciousness. And now we apply the highly emotional area of pediatrics. And without good explanation and the fact that we have data sources coming from Medicaid patients that were already felt that are disenfranchised by other payer groups we have the potential to have a good deal of angst associated with this measure release.

And so I would just caution and encourage that along with the implementation guide we include communication plans and education to the public.

CO-CHAIR KAPLAN: Certainly this is one of those ones that if approved that MAP is going to have something to say a lot about. So, Karen?

DR. JOYNT: I just have two additional thoughts about the usability. Both again perhaps not things that would keep the

measure from going forward but things that I think would be particularly essential in this population to consider.

And one is socioeconomic status. I don't think you can get away from that in this population. And especially because you've shown data that shows us that there's a big difference there. Not necessary to adjust for it but it's got to be part of the discussion about how we think about this and how it sort of rolls out. Because if we're just identifying hospitals that differ by socioeconomic status I don't think we're doing anyone a favor.

And the second is administration rates. We've talked about this with the adult measures as well. If a community works to reduce their admission rates for asthma they could potentially increase their readmission rates. And I think that would be equally true for the complex chronic condition patients. If you put a good intervention in place to

improve their care as outpatients they might not go into the hospital as much.

And I would hate to see hospitals look worse on readmissions because they have done such a good job of providing good outpatient care for their patients.

I just don't think the long-term quality here is just about readmission. It's also got to be about admission. That's where this stuff is moving. And I would really encourage as we think about usability think about some sort of companion way to examine that piece.

CO-CHAIR KAPLAN: Thank you. Taroon?

MR. AMIN: Just to follow up on the question of the all-payer claims and the MAX data.

I just wanted to clarify from the developer is the information that's provided in the testing form, that's a testing from the MAX data set, right?

DR. NAKAMURA: This is Mari Nakamura. Yes, that's right.

MR. AMIN: Is there information that you can share with the committee around the testing from the New York State all-payer claims database that you've done?

DR. NAKAMURA: What we would be able to provide is a sense of what the readmission rates look like, how they compared to the Medicaid readmission rates.

We also specifically used our all-payer data sets to be able to evaluate some questions that we couldn't in MAX. So, specifically insurance status. And also the relationship between insurance status and race/ethnicity and risk factors for readmission.

That said, we did not do all of the testing in the all-payer data set that we did with the MAX data set in part because we were limited in the all-payer data set to not as many states and we felt that the MAX data

set was actually far larger. And although it was limited to Medicaid patients based on the power that we had the more desirable data set to use as our primary development data set.

CO-CHAIR KAPLAN: Thank you. We're over time for this measure by a considerable amount.

On the other hand, since this is the first of its kind I feel like this was a very good discussion of the kinds of issues that are problematic.

With respect to Medicaid status and Karen's issue about socioeconomic status, if you use it as an adjuster you're adjusting away the thing you're trying to explain.

So it's one of these problematic areas that once again we are faced with here's new ground and we're faced with the measure we have in front of us, not the one we would like to develop 5-10 years hence. So, staring at the thing you're staring at, this issue we're now preparing to vote for. Any other

comments?

CO-CHAIR HALL: I do. I'm not an expert in gates data and I'm still not sure what I've heard. So a couple of the experts maybe could comment.

The extrapolatability from Medicaid to the general population sounded like there were concerns about the -- resolved in my head.

Those still exist. Anybody want to?

DR. FIELDS: Yes, I think the tough thing about this, I'm just trying to summarize what I've heard and what I know is that ideally the first measure for pediatric care would probably be more related to access to care and especially primary care and surveillance. It wouldn't be about hospital services. Children have the lowest overall admission rate of any cohort seen in the emergency department, for example, in terms of unscheduled admissions or readmissions.

So, in a perfect policy development world you wouldn't start with the hospital piece. You'd probably end with it.

The other fundamental problem which we touched on yesterday is that the place you need to move the rock with the most urgency are the very states where you're going to have the poorest Medicaid data and the smallest number of children who should be getting better primary care eligible for Medicaid services.

So it's a fundamental conundrum. Because as with other applications of administrative data this is an easy place to start. It may be the only place to start, but it's not the best place to start.

CO-CHAIR KAPLAN: Thanks. Kathy?

DR. AUGER: I would comment that I agree with you that certainly some aspects of pediatric readmission are -- hinge upon things like access to care although I don't think that that's all of it.

And certainly I think in terms of -- I think it's a real question of whether or not hospitals can truly move this metric although it seems like with the new metric the standard is to reassess in three years and see whether or not -- if there was any change in time.

But then I would just also -- I think to me the bigger question is what Bruce raised. It's primarily based on Medicaid data and how it extracts beyond Medicaid is still a little bit of a question mark in my mind.

CO-CHAIR KAPLAN: I would just add from a non-physician, non-provider standpoint it isn't a good idea if children are readmitted to the hospital. So the idea that readmission to the hospital is a quality measure is not a stretch for me.

I mean, the question is rather as Bruce alluded to who are we extrapolating to from who are we actually able to measure this on. And that remains a difficulty.

On the other hand we are again, right Taroon? Stuck with -- we're not stuck with --

(Laughter)

CO-CHAIR KAPLAN: -- we have what exactly we have in front of us and that's what we're considering.

CO-CHAIR HALL: We're looking where the light is. Leslie.

MS. HALL: I guess my concern about the Medicaid data is somewhat helped by the thought that if we help the poor we help everyone.

CO-CHAIR KAPLAN: Okay. Go ahead.

MS. TRAVIS: I guess -- this is just a clarification because I might have gotten confused along the way.

If this measure is fully tested in the Medicaid population with the validity testing and not fully tested in the all-payer or in other settings, when we approve it it will say it was tested in the Medicaid

population how does that affect how it is used out of the portfolio?

Because usually if I remember correctly we've always said that it's endorsed based on how it's specified and how it's tested I think was the second part of that. So I'm just trying to ask for clarification on that.

CO-CHAIR HALL: It's not specified to be exclusively used in Medicaid. It is -- the details reveal that it was tested on Medicaid so my interpretation is that it would not be limited to application in Medicaid. But then in follow-up somebody would see how it performs. Taroon, do you want to add to that?

MR. AMIN: Yes, and I think that seems to be the conundrum that you've raised which started this conversation which is that it's not specified -- the testing doesn't match completely the specifications.

So if it were to be used outside

of Medicaid, you know, it's specified to be able to do that. Although the testing doesn't demonstrate how it would perform outside of Medicaid.

CO-CHAIR KAPLAN: So, are we close to ready? It's my understanding that if you did nothing else you would overestimate the amount of -- the readmission rates by hospital. And then the question is whether or not that overestimation would compromise its usability for when measures got less and less stable because the numbers got smaller and smaller for readmission rates. So, that could be addressed in a use if approved.

DR. NAKAMURA: Thank you. This is Mari Nakamura. You're correct that readmission rates will tend to be higher in a Medicaid-only population we found than in an all-payer population.

There's some different sort of considerations that work in going from a Medicaid-only to an all-payer data set.

So, for example, we would anticipate that for many hospitals the sample sizes used would actually get better because while Medicaid is a sizable portion of all hospitalizations it's only about one-third. And of course that will differ depending on the proportion of Medicaid at a given hospital.

But if anything we would expect that reliability would improve and that the position of rates would actually get better.

CO-CHAIR KAPLAN: Gotcha. The numerator would -- okay.

CO-CHAIR HALL: I think the numerator would get better but fundamentally you are -- you're drawing coefficients out of a higher-risk population which means you're less likely to hold a hospital accountable that has a lower risk population.

DR. NAKAMURA: This is Mari Nakamura. Thank you for that question because I hadn't realized -- I apologize -- part of

the concern here.

So the way that our measure is run is that actually we don't provide the beta coefficients as set numbers to be used in the model. Instead the model is actually run on the data set to which it's being applied and new coefficients for that particular population are generated.

So I agree that the choice of covariates we made certainly was based primarily on the Medicaid data set. We did test the very same model on our all-payer data set and we believe that in terms of the relationships between things like age and gender and chronic conditions that one can generalize from the Medicaid population to an all-payer data set in terms of those fixed effects relationships while recognizing that at the same time having Medicaid insurance as a child is, it appears, an additional risk factor on top of those patient characteristics.

CO-CHAIR HALL: Thank you.

CO-CHAIR KAPLAN: Fair enough. Kathy.

DR. AUGER: Do you have a sense of what the C statistic was in the all-payer model? Is it the same?

DR. NAKAMURA: This is Mari Nakamura. No, I'm sorry, I don't have that although we would be able to easily determine that.

CO-CHAIR KAPLAN: Okay. I think we are ready to vote on the issue of, where are we, usability? Usability.

MS. SHAHAB: Voting for usability and use. One is high, two moderate, three low, four insufficient information and time starts now.

We have all the votes for usability and use. Zero high, fourteen moderate, eight low, zero insufficient information.

CO-CHAIR KAPLAN: Thank you. And

drum roll, drum roll, we are now to endorsement. Any more further thoughts? Okay.

MS. SHAHAB: Voting for overall suitability for endorsement. One is yes, two no. Time begins now.

We have all the votes for overall suitability for endorsement. Measure 2393 Pediatric All-Condition Readmission Measure, the votes are 17 yes, 5 no.

CO-CHAIR KAPLAN: Thank you very much and thank you to the developers. Now we are 20 minutes behind. And so we are going to have to sort of make tracks.

I assume much of the discussion will now sharpen up and condense itself around many of either not similar issues but possibly easier to deal with issues.

So, would you reintroduce yourselves for the record and then describe the measure.

DR. NAKAMURA: Thank you. This is

Mari Nakamura from Boston Children's Hospital.

DR. ZASLAVSKY: Alan Zaslavsky from Harvard Medical School.

DR. NAKAMURA: And we have Mark Schuster on the phone. I will keep my introduction brief because by design these two measures are very similar and so for the most part the same considerations apply.

Our pediatric lower respiratory infection measure evaluates readmissions following an index hospitalization for bronchiolitis, influenza, or community-acquired pneumonia.

We decided on a measure focusing on lower respiratory infections because they're among the most common reasons for hospitalization in children.

In addition, they're among the diagnoses with the most prevalent readmissions. We found an overall 30-day readmission rate of 5.6 percent which corresponds with the large absolute number of

readmissions given the high number of initial hospitalizations for lower respiratory infection.

We prioritize harmonizing this measure with NQF-endorsed adult readmission measures as well as with our all-condition measure. And as a result the approaches we used in developing the measure are very similar.

We used the same case mix adjustment model because we found that it performs very well for LRI readmissions with regard to discrimination and calibration.

The issue of limited reliability due to small sample sizes is even more of an issue for any condition-specific rate -- and this is true for LRI -- than for all condition rates. But again we found that a majority of children were cared for at higher-volume hospitals than as a result were at hospitals with good reliability.

The New York Office of Safety and

Quality also tested this measure on its Medicaid and all-payer databases and noted that having implemented one measure it required minimal effort to be able to implement the other.

And so we do think that this measure could be a useful tool to evaluate quality and encourage improvements in care for an important pediatric condition. Thank you.

CO-CHAIR KAPLAN: Thank you. I think Mari is the only person in the room that talks faster than I do so thank you very much for that quick summary.

Jo Ann, do you want to talk to us about evidence?

DR. BROOKS: I'll go ahead and get started with this.

As was said this is a companion measure to the one we just discussed. And looking at lower respiratory infections, accounting for a large number of the readmissions we see it is an important area.

And there's good support here on how these readmissions, we may be able to look at things to improve the readmissions, looking at our key processes, our discharge planning, care transitions, appropriate follow-up, et cetera.

And disparities exist for many of these differences we see in patients being readmitted for lower respiratory infection.

CO-CHAIR KAPLAN: Thank you. Kathy, nothing to add? Others? Vote evidence.

MS. SHAHAB: Voting for 1(a) evidence, one yes, two no. Time begins now.

We have all the votes for 1(a) evidence. Nineteen voted yes, two voted no.

CO-CHAIR KAPLAN: Performance gap. Jo Ann?

DR. BROOKS: There exists a performance gap for this. This is also a newly commissioned measure by CMS and AHRQ. There's disparities in care across populations

in many different ways and there are strong data to support that there is a quality gap and a need for this measure.

There is support in the application talking about some of the pediatric data that's out there although it is not as rich as the adult data. But the appropriate rationale is there.

CO-CHAIR KAPLAN: Kathy?

DR. AUGER: I'd just comment though that the readmission rate for the lower respiratory tract infections is actually lower than the all-cause.

And so it's hard to know what the -- like how much of a range we're actually dealing with at the different hospitals, like what the interquartile range would be for the risk-standardized rate and whether or not that's significant.

CO-CHAIR KAPLAN: Do the developers have any information for us?

DR. NAKAMURA: This is Mari

Nakamura. I guess -- thank you.

So, for our lower respiratory infection measure to give you analogous numbers to what I provided for all-condition, a hospital that's two standard deviations below the mean would have a readmission rate of 1.7 percent.

A hospital two standard deviations above would have a readmission rate of 12.6 percent. So, there's actually a wider range for lower respiratory infection.

CO-CHAIR KAPLAN: Thank you for that. Others? Okay, ready to vote? Performance gap.

MS. SHAHAB: Voting for 1(b) performance gap. One high, two moderate, three low, four insufficient. And the time begins now.

We have all the votes for performance gap. Three voted high, eighteen moderate, zero low and zero insufficient.

CO-CHAIR KAPLAN: Thank you.

Priority. Jo Ann?

DR. BROOKS: And this continues to be a high-priority measure as it was commissioned by AHRQ and CMS. Also because it relates and impacts a large number of pediatric patients and accounts for a large number of the readmissions in hospitals.

CO-CHAIR KAPLAN: Thank you. Kathy?

DR. AUGER: Yes, just lower respiratory tract infections are one of the most common indications for hospitalization in pediatrics so to me it's high priority.

CO-CHAIR KAPLAN: Others? Ready to vote?

MS. SHAHAB: Voting for 1(c) high priority. One high, two moderate, three low, four insufficient. Time begins now. Just one more vote.

We have all the votes for 1(c) high priority. Twelve voted high, eight moderate, one low and zero insufficient.

CO-CHAIR KAPLAN: Thank you very much. Scientific acceptability. Start with reliability. Jo Ann?

DR. BROOKS: Yes, this measure was tested the exact same way as the previous measure using the MAX data. And when we looked at reliability it ranged between 0.5 to 0.77.

The measure was also found to be highly reliable at hospitals with an adequate sample size, but obviously it did not perform as well in those with lower sample size.

And one of the questions was will exclusions make the reliability across time and place an issue for this measure. Because there's a large number of exclusions in this measure.

DR. NAKAMURA: This is Mari Nakamura. Regarding exclusions the -- one difference with our all-condition measure is the case definition requirement for the index hospitalizations.

So, speaking more generally about both measures I think that we've created an appearance of maybe more exclusions than typical because we listed everything in terms of being exclusion rather than some as inclusions.

So, some of these are, for example, based on limiting to the pediatric age range. Some other common exclusions that we saw have been used quite uniformly in adult measures such as excluding patients who leave AMA or certainly patients who die in the hospital.

Our other set for the exclusions that are quite similar to the adult readmission measures are certain data quality exclusions for key variables with missing or what looked like poor quality data. For example, a discharge date that occurs before a date of birth.

The main clinical exclusions that we apply for both measures are, first of all,

patients who are receiving obstetric care with the rationale that these patients are typically not under the purview of pediatrics even if they fall in the pediatric age range. And so we felt would most likely be better included in obstetric measures rather than a pediatric measure.

We also exclude patients with a primary mental health diagnosis. That is consistent with other measures but in our own testing as well we felt it was justified because we found that readmission people for a given hospital does not track for mental health conditions as compared with other conditions.

And then finally we exclude newborns who are in the hospital for their birth admission. The clinical rationale for that is that they are among all patients in the hospital not actually there because they're ill, but rather for another life event. And so we felt that it made sense not

to include them in a readmission rate.

CO-CHAIR KAPLAN: Thank you. I'm going to ask the committee to hold that consideration for the validity discussion as opposed to -- unless those exclusions change the reliability we need to keep focused on reliability. Kathy?

DR. AUGER: I would just comment as Mari already acknowledged the reliability in this metric is not quite as good as the previous measure in that only 229 of the 1,743 hospitals actually had a readmission rate reliability of greater than 0.5.

But that again these hospitals accounted for 62 percent of the lower respiratory tract infection hospitalizations. So it's still -- even though it's a smaller number of hospitals it's still a majority of the hospitalizations.

CO-CHAIR KAPLAN: The precision of those estimates once again sort of tracks with everything else we're seeing. In fact, it

lands a little bit on the higher side compared to other reliability estimates we've been seeing in some of these other measures. Anybody else? Vote reliability.

MS. SHAHAB: Voting for 2(a) reliability. One is high, two moderate, three low, four insufficient. Time begins now. We're still waiting on two more votes. One more.

All votes are in for 2(a) reliability. The results are 1 high, 18 moderate, 2 low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Validity. And I'll ask you to keep in mind the exclusion discussion we just had.

DR. BROOKS: On the validity, the construct validity was demonstrated as associated with the literature and the quality of care. The processes related to reductions and readmissions that we're all aware of.

Criterion validity was shown using a data set from Boston Children's Hospital

over a 1-year period. And the sensitivity and the specificity were 87.0 and 99.7 percent respectively.

They also did face validity on the planned procedure algorithm that was completed and also received public comments from the Federal Register. And they then took those comments and put those into the methodology to improve it.

And those would be my comments.

DR. AUGER: So, very similar issues with validity in terms of identifying unplanned readmissions as appropriate.

The couple of questions -- well, so I'd say that the model calibration is good and the C statistic for predictive ability was 0.71 so that's in a reasonable range.

The one question that we were just talking about is how CF exacerbations come into play here, whether or not the model would adequately account for CF as well.

DR. NAKAMURA: This is Mari

Nakamura.

So, cystic fibrosis in particular would end up being included in the model as one of the chronic condition indicator variables for chronic respiratory infections.

But this is also an opportunity -- I'll keep it very brief -- to mention that we thought about whether we needed to have more symptom variables for chronic conditions for the LRI measure versus the all-condition.

And in reflecting decided not to because it's actually a wide range of conditions that place patients at higher risk of severe lower respiratory infections. They're not just respiratory, but for example, cardiac, neurological. And so we felt that it made sense to keep all of the chronic condition indicators and found indeed that the model actually performs very well.

CO-CHAIR KAPLAN: Thank you. I want to ask for a point of clarification about criterion validity because that implies a gold

standard. You weren't using Boston Children's as the criterion. Did I misunderstand that? How is what you did criterion validity?

DR. NAKAMURA: This is Mari Nakamura.

So we were using the electronic health record data as our gold standard based on having performed detailed chart reviews of the cases that we evaluated with the idea that such a chart review is at least a better standard than what can be found from claims data in terms of being able to assess both whether readmission occurred as well as whether it met our definition for an eligible readmission meaning that it wasn't for a planned procedure or chemotherapy.

CO-CHAIR KAPLAN: Thank you. Some of us would call that convergent validity because it converges with a different data source as opposed to criterion validity. But that's okay.

DR. NAKAMURA: Thank you.

CO-CHAIR KAPLAN: Other comments on validity? Let's vote.

MS. SHAHAB: 2(b) validity. One is high, two moderate, three low, four insufficient. Time begins now. One more vote.

All votes are in for 2(b) validity. Zero voted high, twenty moderate, one low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Feasibility?

DR. BROOKS: And on feasibility I would say that since it's claims data it's easily feasible for us to get these data.

The one question -- some of the concerns we've already had is it's based on Medicaid data. But is it feasible as claims data.

CO-CHAIR KAPLAN: Others? So we're burned out on feasibility.

(Laughter)

CO-CHAIR KAPLAN: Ready to vote.

MS. SHAHAB: Voting on feasibility. One high, two moderate, three low, four insufficient. Time begins now. Still waiting on two more votes.

All votes are in for feasibility. The results are 3 high, 17 moderate, 1 low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Onto usability.

DR. BROOKS: Usability, since this measure is really a subset of the previous measure all the issues we've discussed with usability in the previous discussion are the same here.

CO-CHAIR KAPLAN: Kathy?

DR. AUGER: So yes, again, just the whole issue of preventability versus unplanned is a consideration. But that's the same as the other metric.

The one thing that I would comment might be different for this metric compared to the other metric is this is -- lower

respiratory infection is a very seasonal illness. And so therefore as it's written I think it's not an issue because it's an annual evaluation.

But it's something that it would not be appropriate for it to be used as a quarterly evaluation because of the seasonal variability in this. But I see it as written as fine.

CO-CHAIR KAPLAN: Want to respond to that?

DR. NAKAMURA: Thank you. This is Mari Nakamura. Kathy is correct that if we were to try to report these results as quarterly that season would be a really important consideration.

We chose to make it annual, taking care of that fact but also recognizing that quarterly rates would be an even greater challenge with regard to sample size.

CO-CHAIR KAPLAN: Thank you. Other comments? Ready to vote.

MS. SHAHAB: Voting on usability and use. One high, two moderate, three low, four insufficient in. The time begins now.

All votes are in for usability and use. The results are zero high, 17 moderate, 4 low and zero insufficient information.

CO-CHAIR KAPLAN: Are we ready to vote on endorsement? Any discussion?

DR. BROOKS: My only comment is that I think this one is very specific to sociodemographic data as well and we need to consider that as we move forward with this down the road.

CO-CHAIR KAPLAN: Yes, the issue of sociodemographic data as we've discussed yesterday and today is going to be, you know, one of these things that is going to be guided by a committee that hasn't yet kind of given its guidance to us. So we are going to either endorse or not endorse the -- right.

CO-CHAIR HALL: Understanding that we can all as a group express that we want our

NQF colleagues to capture that for these two measures the group in particular felt very strongly that this question needs to be addressed in the future.

Do people feel that's a relative consensus? Anyone would object to attaching that comment?

CO-CHAIR KAPLAN: No, but and the issue of Medicaid and SES adjustment is one of these things that's also going to present a rather dodgy problem. Because socioeconomic status and some of the adult measures actually use Medicaid status as a proxy for socioeconomic status. So it's going to be a more complex issue with some of these measures.

CO-CHAIR HALL: All I'm suggesting is we create a bit of a flag for our NQF colleagues that if someday comes where the white paper recommendations have changed on this that this is an easy flag to spot.

CO-CHAIR KAPLAN: So flaggage

approved and associate recommendations.

(Laughter)

CO-CHAIR KAPLAN: Are we ready? Any other comments before we vote for endorsement or non-endorsement?

MS. SHAHAB: Voting for overall suitability for endorsement. One yes, two no. Time starts now.

All votes are in for overall suitability for endorsement for measure 2414 Pediatric Lower Respiratory Infection Readmission Measure. The results are 18 yes and 3 no.

CO-CHAIR KAPLAN: Thank you very much. And now we are ready for NQF member and public comment. And we will invite public comment. Thank you to the developers for a nice summary and discussion.

We're onto the NQF member and public comment. I'm going to turn over to Adeela.

MS. KHAN: Kathy, can we compile

the list, please?

OPERATOR: Yes, ma'am. If you would like to make a public comment please press * then the number 1. No, no public comments at this time.

MS. KHAN: Do we have any public comment in the room?

DR. SCHWALENSTOCKER: Good afternoon -- or I guess it's good morning. And I realize I'm standing between you and lunch.

My name is Ellen Schwalenstocker. I'm with the Children's Hospital Association.

And I just wanted to highlight some of the discussion among the committee that I just think is really important some of which is outside your purview.

But we've been saying for a long time that it's really important to develop good pediatric measures. And I want to thank the Center of Excellence at Boston Children's Hospital for their work in developing these

measures.

I'm not sure, and this may be a question for you, what the 3-year experience sort of guidelines are. I do strongly agree with the recommendation on should the NQF change its policy on adjustment for SES and other factors, that it would be really important to look at these measures as well as adult readmissions measures in that light.

I also think because of the lack of a large set of good pediatric measures it will really be important to have other measures available before these kinds of measures are used for, say, accountability purposes or pay-for-performance.

And so I'm wondering if that's sort of part of the 3-year process to give us more chance for validation of measures against other measures? I guess that's convergent validity.

And I know there are some other measures that will be coming forward from the

CHIPRA Centers of Excellence which we just think is a really important program and are glad to see the measures coming to NQF.

MR. AMIN: So, I can address that, Ellen, right now.

So, the NQF reevaluates all of its measures that are recommended for endorsement in a 3-year maintenance cycle.

In that 3-year maintenance cycle we look for experience on the measure. In particular in the importance to measure criteria we're looking for some actual, at the measure performance gap information we're not just looking conceptually whether there's a measure performance gap, we want to see the performance gap in the measure itself and to see the distribution and if there's any overall gap in performance or overall less than optimal performance.

And then also in the use and usability criteria we want to see if the measure has been implemented for use in

quality improvement applications and then a further time line is ready for accountability applications which we include public reporting in that domain.

So the purpose there is that we have measures that are actually picked up by the field and we can demonstrate that they're actually being used for the purposes of quality improvement and reporting the information to the public.

DR. BURSTIN: I'm just wanting to add again if this measure is endorsed and goes through the entire process, again, the SES issue I think outstanding. We don't know how that's all going to play out.

But I think certainly we'll have to make a decision with our CSAC and our board depending on how that lands in terms of whether we'll bring measures back sooner than the 3-year time frame for maintenance if in fact there are additional issues to address.

But you know, a very important

piece of this. As these measures are going out there NQF endorsement does imply they are appropriate for any accountability application and could get picked up for any of those certainly by CMS or others.

We would really encourage the field to be really vigilant about keeping track of measures that are really helping to drive improvement, measures for which there may be unintended consequences and bring that forward, that information, in realtime rather than just waiting for the 3-year limit.

MR. AMIN: Are there any other public comments?

OPERATOR: Once again to make a public comment please press *1. You have a comment from John Muldoon with 3M Health Information Systems.

MR. MULDOON: Are you able to hear me?

MS. KHAN: Yes, we can hear you.

MR. MULDOON: Several comments.

And I submitted a lengthy quite detailed set of comments about a month ago.

But in terms of the concept of readmissions, the risk adjustment and the testing and the evaluation work. And just highlights some concerns that we had identified. And some of these came up during discussion as well.

The approach focuses on unplanned readmissions as opposed to trying to identify potentially preventable. And realize that that's a tough challenge to identify.

But there's a number of consequences. Because there are a number of very low preventability readmissions such as malignancy-related admissions, neutropenia, classic anemia, just an easy one to highlight. And those kinds of kids are treated at certain centers. So it creates numbers that may not be that real in terms of preventability and distortions in terms of comparisons across different settings.

Also, when you go out to 30 days and the number of unrelated and low preventability readmissions increases. So that compounds the concern and the approach to readmissions.

On the risk adjustment concern that often the principal diagnosis and the acuity of the admission and really complex prior conditions that have a very big impact on admission rates and readmission rates.

And that's not very specifically addressed. It's more of a generic approach to any chronic condition from a list of about 4,500, mild, moderate and severe, many of which have very little influence over admissions and readmissions.

So, we're concerned that that's not picking up the readmission factors very well.

And in terms of testing and evaluation I don't think we really saw it tested across different subgroups of pediatric

patients and across different hospitals who serve different populations.

So I think there's a lot that we don't know and need to be concerned about as it gets rolled out if it's endorsed.

And just to illustrate, for lower respiratory infections for the pediatric population those with major chronic conditions.

We've done a lot of research and analysis on this such as cystic fibrosis as the committee discussed, bronchopulmonary dysplasia, ventilator-dependent patients, chronic respiratory failure. They tend to have readmission rates in the 10 to 15 percent range compared to otherwise healthy children in the 3 to 5 percent range.

And we just don't see how that can be teased out with the more generic risk adjustment methods. So those are the thoughts that we'd like to share.

And if the committee does endorse

I think there's a lot of cautions to go forward. Thank you.

MR. AMIN: Thank you. Is there any other comments on the phone?

OPERATOR: At this time there are no public comments on the phone line.

MR. AMIN: There are no other public comments in the room.

CO-CHAIR HALL: I think we can break for lunch then. Just a quick note for those who weren't here yesterday. We would ask the audience members to wait for the committee members to grab their lunch first before helping yourself.

CO-CHAIR KAPLAN: So we'll reconvene at 12:15.

(Whereupon, the foregoing matter went off the record at 11:41 a.m. and went back on the record at 12:13 p.m.)

CO-CHAIR HALL: We have three measures left this afternoon, one from the American College of Cardiology and two from

Yale CMS.

And then we have a very important brief update on measure 1789. So we're hoping we'll still have faces around the table when we get to that point.

We'll ask the American College of Cardiology representatives to introduce themselves and briefly introduce their measure, please.

DR. CURTIS: Hi, this is Jeptha Curtis from Yale also representing the ACC today.

CO-CHAIR HALL: Hang on, I'm not sure we heard that.

DR. CURTIS: Sorry. Jeptha Curtis from Yale representing the ACC today.

MS. SLATTERY: So Dr. Curtis will be speaking primarily to the measure methodology but I'm Lara Slattery, ACC senior director for scientific reporting and I can address any questions related to implementation.

DR. CURTIS: So, for endorsement maintenance today under review is the Hospital 30-day Risk-standardized Readmission Rates Following Percutaneous Coronary Intervention.

It is a measure that was endorsed I believe 3 or 4 years ago. And identifies unplanned readmission rates for hospitals that perform percutaneous coronary interventions on Medicare fee-for-service patients greater than 65 years old.

As I mentioned this is unplanned readmissions and we updated the measure from the initial endorsement to include a modified version of the hospital-wide readmission algorithm that identifies unplanned readmissions. Basically including a larger number or considering a larger number of readmissions planned.

In addition, we've updated the measure to include the use of direct as opposed to indirect identifiers to match the CathPCI registry data with CMS information

about readmission which is used to identify the risk-standardized rates.

And has also in the interim actually gone into implementation. And last March the hospitals received their reports as to their risk-standardized readmission rates. And then there was a voluntary public reporting of those rates on Hospital Compare as well as the ACC's internal websites.

So that this information has -- or this measure has progressed from development to approval to implementation in a relatively short time frame. And I think it's open for questions.

CO-CHAIR HALL: Thank you very much. We will start with the category of evidence and our lead discussants are Mae and Larry. So I invite them to kick it off.

DR. GLANCE: I'll go ahead and start. In terms of evidence to support the measure focus the measure developers present evidence that there is a high level of

variation across hospitals which to me represents strong evidence.

When you have variation in outcomes that means that there's the potential to improve your outcomes assuming that you have properly adjusted for differences in case mix. So I think the evidence is strong.

CO-CHAIR HALL: Mae said she agreed but I didn't -- is your mike working?

MS. CENTENO: I agree.

(Laughter)

CO-CHAIR HALL: Okay, great. Any other comments on evidence? Okay, we'll move to vote.

MS. SHAHAB: Voting for 1(a) evidence. One is yes, two is no and the time begins now.

Twenty yes and zero no.

CO-CHAIR HALL: Moving into performance gap, opportunity.

DR. GLANCE: In terms of the performance gap without risk adjustment the

difference between the lowest decile and the highest decile was zero percent readmission rates versus 28 percent readmission.

After risk adjustment the difference between the 10th percentile and 90th percentile was 13.5 versus 10.1 percent. So again, evidence of a significant performance gap.

CO-CHAIR HALL: Additional comments? Okay.

MS. SHAHAB: Voting for 1(b) performance gap. One high, two moderate, three low, four insufficient and your time starts now.

All votes are in for performance gap. Seventeen high, four moderate, zero low and zero insufficient.

CO-CHAIR HALL: Priority.

DR. GLANCE: In terms of priority this is one of the conditions that was identified as a high-priority condition by MedPAC.

The overall incidence of readmissions within 15 days was 10 percent which represents roughly about 44,000 readmissions in 2005 at a cost of $360 million annually. So, I would suggest that this is a high-priority condition.

CO-CHAIR HALL: Additional comments? Seeing none.

MS. SHAHAB: Voting for 1(c) high priority. One is high, two moderate, three low, four insufficient and time begins now. Three more votes.

All votes are in for 1(c) high priority. Eighteen high, three moderate, zero low, zero insufficient.

CO-CHAIR HALL: Scientific now. Reliability and validity.

DR. GLANCE: In terms of reliability -- data reliability. This is based on clinical data which is audited using annual onsite chart reviews and data abstraction.

In terms of model reliability the intraclass correlation coefficient was 0.37 which is indicative of fair agreement and very much in the zone of the other measures that we have looked at over the past two days.

CO-CHAIR HALL: Additional comments? Seeing none.

MS. SHAHAB: Voting for 2(a) reliability. One high, two moderate, three low, four insufficient and the time starts now. One more vote.

All votes are in for 2(a) reliability. Five high, sixteen moderate, zero low and zero insufficient.

CO-CHAIR HALL: Validity.

DR. GLANCE: In terms of validity testing, in terms of looking at the statistical performance of the model the discriminate ability, the C statistic was 0.66, actually 0.67 in the validity data set which is very good for this kind of a model.

Model calibration was also very

good. They looked at this using both graphical techniques and other approaches as well.

In terms of threats to validity 29 percent of the observations were missing data on ejection fraction. Ejection fraction is considered to be a very important clinical risk factor for these types of models.

They imputed the missing data but they used a very rough approach I guess for imputation. And I quote, "We stratified by gender and imputed the missing values to the median of the corresponding groups."

So state of the art for imputation is to use multiple imputation. I was a little surprised that this approach was used. And I was wondering if the measure developers could maybe comment on this.

DR. CURTIS: Yes. I mentioned we've been on this journey for 5 or 6 years so sometimes it's a little hard to reconstruct exactly what the logic was from that long ago.

Nevertheless there are two aspects to, first, why EF is missing so frequently and second, our approach to accounting for the information that's conveyed by the missing data.

So first off, the LVF is specified to be an ejection fraction that -- an information by the ejection fraction that is available prior to the performance of the PCI. So we obviously don't want to -- if a PCI goes wrong and the patient has a large MI, has a low EF we don't want to account for that in the model. So it has to be LVF prior to the PCI.

The patients who don't have information about an ejection fraction before the PCI typically are those that are being done on an urgent or emergent basis. So, it is very much colinear with patients who have an ST elevation MI, patients with cardiogenic shock or other highly morbid conditions. And for that reason it's certainly not missing at

random.

To account for that, in addition to imputation we actually also have a dummy variable in the model for missing LVEF. So that aggregates both as patients in whom it's missing at random as well as those in whom it is missing not at random, i.e., that they had an emergent procedure.

But it does account for the information that runs or is colinear with missingness. And so we don't ignore the information that's conveyed by that.

In terms of the single versus multiple imputation, that's something I think -- a decision that we made a long time ago that was consistent with the inpatient mortality model for ACC. And so for that reason I think we were trying to be harmonized in terms of our approach.

It's something that we could re-look at or modify going forward. There's no reason that we couldn't do it from a

statistical standpoint.

DR. GLANCE: So, in my mind this remains a bit of a threat to validity. Some people would contest that if you're missing this much data on a particular covariate that it shouldn't even be included in the model.

And usually people typically, if you're missing more than about 10 percent, maybe 20 percent of the data that's grounds for sometimes not including it in the model.

The reference was made as to whether it's missing at random or missing not at random. Without getting into a lot of the technical details what lots of folks will do in this particular setting is they will create a regression model based on all the other available risk factors to predict the missing values for patients who have missing values.

If you can do this, if you can predict a missing value using available covariates then it is missing at random. And I would suggest that probably in many of these

cases you could do that, that the data actually is missing at random as opposed to not missing at random. And I do think that is a bit of a threat to validity.

Missing completely at random would be if it was just -- well, I don't want to get into too much of the details on this. But I think it is an issue.

CO-CHAIR HALL: Jeptha, you indicated that it was -- or did I misunderstand that you said it was somewhat colinear with emergency status.

DR. CURTIS: Right. I mean, it's been a long time since we did those analyses, but the patients in whom LVEF is not available, it's not really necessarily missing but it's not -- the test of ejection fraction has not been performed, that is typically those patients in whom you're under the gun to perform a primary angioplasty.

So the STEMI patients for whom we're trying to do a door to balloon time in

less than 90 minutes oftentimes we'll forego doing an ejection fraction before the procedure. So from that standpoint it is colinear with the urgency or emergency procedures.

And that it sort of fits in that it's missing in about I'd say 10 to 20 percent which is -- I'm sorry, I can't remember the exact number. But that's the lion's share of what's missing.

It's not necessarily missing the information was available and just not captured. I don't think that's what we're looking at here.

DR. GLANCE: So by imputing it to essentially a normal value which I believe is what you're doing you are to some extent disadvantaging hospitals that are taking care of more emergencies compared to fewer emergencies.

DR. CURTIS: Because we have a dummy variable for missing information my

understanding -- I said that before, yes.

We have a category for LVF of -- and I miss the actual specifications, but it's low EF, moderately low EF, normal EF, and missing EF. I think there are four or five categories for EF in the model. So missing information is not ignored.

CO-CHAIR HALL: So the continuous variable is moved to median and there's an indicator as well. And we think it's probably redundant information somewhat with at least one other variable as well.

So, although it may not be perfect it's probably less of a threat than it might at first sound like it is.

DR. GLANCE: I will take back my original comment. I did not realize, or I didn't hear you when you said that you had an indicator variable for missing. So I would no longer qualify that as a significant threat to validity.

CO-CHAIR HALL: Paul.

DR. HEIDENREICH: I'd also say that we have all this clinical data here that we have in a couple of other measures. But most of the measures you've been talking about have claims data. And we're talking about potentially even better risk prediction. So I think in the big scheme of things any threats to validity are probably pretty small.

CO-CHAIR HALL: I'm not seeing any other cards raised so I would just like to highlight if for my own understanding to make sure.

The real significant updates to the measure are that there were some changes in the CathPCI registry variables which warranted remodeling.

There's an improved strategy for linkage, so that should again improve. I'm just thinking in terms of overall validity. The first issue, the update to Cath registry data fields you would expect to improve the validity overall as Paul just hinted at.

The improved linkage we would expect to be an improvement overall.

The planned readmission algorithm is an interesting aspect. You updated and more or less expanded the planned readmission algorithm such that at least in summary the end result was that the overall crude planned rate fell from 12.3 in the percent according to the prior algorithm to 11.8 percent according to the new algorithm.

In other words, you're giving more institutions and providers more credit for readmissions being planned. And so we're minimizing -- hopefully we're minimizing or reducing the chance of falsely throwing a flag at them. Perhaps at some cost but that's what we're doing.

The ICD crosswalk, not really an issue yet.

And then updating your cohort code again appears to be an improvement to the overall validity of the measure.

Any other -- yes, Sherrie.

CO-CHAIR KAPLAN: So, you're not opposed to doing multiple imputation if we suggested that that might be an enhancement to the missing data problem?

DR. CURTIS: I don't see it as being a major barrier. I think that the vast majority of the information that's conveyed by missing LVEF is captured in that categorical variable of missing EF. It is one of the more powerful coefficients in the model. But yes, we could certainly do that.

CO-CHAIR HALL: Well, I think as a small ask maybe you could show in the future what some of the modeling looks like with and without in order to make the case of whether it's really worth the effort.

Paul?

DR. HEIDENREICH: It just seems like that would be an NQF sort of policy if they want to make that, that one does multiple imputation whenever you're doing imputation.

I would not recommend that but it seemed like rather than have the different groups suggest it here and there that should be a standard policy.

CO-CHAIR HALL: I understand that comment. I think NQF has shied away from saying that models have to be performed in a particular specific way as we've seen. We've seen modeling done in a number of different ways already.

So I understand your comment but I would hazard a guess that it will remain the judgment of the group as to whether a model is appropriately specified. Larry?

DR. GLANCE: I think the issue really isn't single imputation versus multiple imputation. I think they're both fairly straightforward to carry out. There's not a huge difference in terms of doing that mechanistically.

I think the issue is whether or not you should use an indicator approach

versus an imputation-based approach. And I think there's a lot of literature which will show that estimates based on multiple imputation are less biased that are based on the indicator variable approach.

CO-CHAIR HALL: At the same time, Larry, with all due respect that might be true of the imputed variable itself. But in this case, the variable probably is as part of a large number of variables in this model the information value may be redundant. And in fact, there could be no value to multiply imputing versus the current approach.

DR. GLANCE: So, agreed. We've actually done some of that research and have shown that there is a significant difference, at least in the population that we looked at. Whether or not that's generalizable to this population I don't know. It's an empirical question.

CO-CHAIR HALL: Agreed. We've done the same thing in NSQIP so it's a common

topic that people spend money on beer on.

I think that, correct me if I'm wrong, so far what we've stated is that the current implementation seems to be a minimal threat.

Other concerns or comments in this category of validity? Not seeing any.

MS. SHAHAB: Voting for 2(b) validity. One is high, two moderate, three low, four insufficient and your time starts now.

We have all the votes for 2(b) validity. Two high, eighteen moderate, zero low and zero insufficient.

CO-CHAIR HALL: Feasibility.

DR. GLANCE: So this measure is based on a hybrid of clinical data and administrative data. The administrative data is just to identify which patients were readmitted.

The clinical data is based on the CathPCI registry. I would ask the measure

developers to tell us what percentage of U.S. hospitals that are currently performing PCIs are in this particular registry.

MS. SLATTERY: So it's hard for us to get to an exact number but based on what we --

CO-CHAIR KAPLAN: Could you state your name, please?

MS. SLATTERY: Sorry. Lara Slattery. Based on what we -- when we look at our participating facilities against American Hospital Association we estimate about 85 percent of current PCI hospitals are participating in the registry, but that it probably represents more about 90 percent of the patients because it tends to be smaller facilities, and usually smaller facilities where they have a state reporting requirement and no other incentive for them to join our registry that are the facilities that are not participating in.

DR. GLANCE: It's in line with the

STS measure that we looked at previously.

CO-CHAIR HALL: I don't know whether feasibility is the right category, but could the developers comment on improvements over time? This has been in play now for at least 3 years, isn't that right?

DR. CURTIS: It's really only been publicly reported last year. And so we don't really have good information about improvements over time.

And in fact, as we've made these improvements to the model moving to direct identifiers there's been enough changes that I don't know if we have a good way of surveillance as to what's been going on even prior to public reporting. So it's probably still a little bit early for us to comment.

CO-CHAIR HALL: Fair enough. But there's a very clear implementation plan, ongoing implementation. So it seems acceptable from that perspective.

Other comments on feasibility from

other group members? I don't see any.

MS. SHAHAB: Voting for 3 feasibility. One is high, two moderate, three low, four insufficient and your time starts now.

We have all votes for feasibility. Six high, thirteen moderate, one low and zero insufficient.

CO-CHAIR HALL: Usability.

DR. GLANCE: So, I ask the measure developers to provide us with information on how this information is reported to participating hospitals.

Do you report it both as a continuous measure and also as a categorical measure, meaning high-quality, low-quality and average quality?

And if so in the most current reporting period what proportion of the hospitals were reported as being quality outliers?

DR. CURTIS: So in the hospital-

specific reports and actually in the public reporting it's put both in buckets as well as the overall adjusted risk-standardized rate. So they receive that information.

In addition, in the hospital-specific reports they receive information about what hospital -- to what hospital the patient was readmitted, the principal diagnosis for that readmission and the time frame of the dates of it.

So we're trying to encourage through this hospital-specific report sort of cross-fertilization and crosstalk across hospitals so that they can try to work and improve these rates.

In terms of the buckets, the calculation of the outliers was done completely in line with what's been done with the other CMS measures. And I believe it was about 2 percent of hospitals that were either high or low outliers.

So relatively small number in part

because the data that we had available for the measure with the Social Security numbers in it that we needed to identify we only had 2 years of data. So it was a little bit of probably a challenge in terms of having enough hospitals with enough volume that we could get really good discrimination as to put them into categories, but not that far out of line for other publicly reported measures.

CO-CHAIR HALL: Was that 2 percent at each tail, or 2 percent total?

DR. CURTIS: I think it was 2 percent total.

MS. SLATTERY: Lara Slattery. I just wanted to clarify. Hospitals received the feedback report by virtue of participating in the registry regardless of whether they opted in for the public reporting component. So they all received the feedback reports.

And the hospitals for which we had no data still received a benchmarking report just for information purposes only.

CO-CHAIR HALL: Additional comments or concerns on usability? Karen?

DR. JOYNT: I know someone just said this, but is there any way to -- or should we be thinking about ways to update the measure such that more than 2 percent can be identified as something? Or do you feel that the value is not about the identification as an outlier but about the benchmarking?

DR. CURTIS: I think we've had this discussion in different forms all morning.

DR. JOYNT: Do you know what happened to those that were identified? Like can you look over the last 3 years and see what prior to the reporting or identification as outliers, sort of how your measure helps people move? Do you have the information on that?

DR. CURTIS: Again, we've only had one year of public reporting which is last March. And so we have not seen what's

happened over time.

And I think your work and others has questioned whether or not it does make a dramatic change as opposed to incremental change.

And I think the jury's still out on how any of these measures can be used to change national performance. I think we are seeing encouraging trends but nothing definite yet.

But I think of it as an implementation question as to where do you draw the buckets and where do you draw the line. Are you very restrictive or are you more permissive in terms of categorizing hospitals as better than or worse than.

And again, we in this case tried to be as consistent as possible with other CMS measures.

CO-CHAIR HALL: Could the developers comment briefly on the unintended negative consequences remarks in the measure

materials?

DR. CURTIS: I think as we've discussed for other measures there's always the possibility of unintended consequences where you could have at worst case aversion on the basis of readmission rates. And that's really nothing that you can necessarily prophylaxe against.

I think it's probably lower stakes for readmissions in procedures than it is perhaps for mortality. Just that would be my initial impression. But I think it's something that has to be monitored and can be monitored.

And we've looked at for other, specifically for mortality we've been looking at whether or not expected case mix and predicted risk has changed over time in states that publicly report PCI mortality and have not seen dramatic changes in sort of -- I'm sorry, not the predicted but the expected mortality at the states that do have public

reporting.

CO-CHAIR HALL: Other thoughts or concerns on usability? Seeing none.

MS. SHAHAB: Voting for usability and use. One high, two moderate, three low, four insufficient information and the time starts now. Three more votes.

CO-CHAIR HALL: Can everybody do it one more time? Sorry.

MS. SHAHAB: We have all the votes for usability and use. Three high, fourteen moderate, three low and zero insufficient information.

CO-CHAIR HALL: Any comments in summary before an overall vote? I see Karen.

DR. JOYNT: I just have another clarifying question. You can tell I've not been on this committee before.

This came up with another measure yesterday that we were re-approving or whatever. Is the burden of proof that it has been implemented, or that it has made X amount

of change? Or what is the sort of -- what is our responsibility in terms of looking at its longitudinal performance in continuing it forward?

CO-CHAIR HALL: I'll give my opinion on that but I'll be corrected if I'm wrong.

I think when there's a clear plan for ongoing implementation that's kind of set one. And if that plan has been and is being carried out in good faith but may not have results yet I think that should meet our adequacy threshold. That's my opinion.

DR. BURSTIN: Just to add to that, last year -- Karen had to leave, but she led an effort with a task force to update our use and usability criterion.

It used to just be usability which basically was it in use. And I think what we really heard from the community, that's not enough.

So in use is really important.

And by the 3-year window we want to see that it's being used in an accountability application within two of those cycles. We want to see evidence of public reporting.

But I think increasingly what we're trying to see is in addition to that, and this is not a must-pass like evidence and scientific acceptability, but we want to in fact be able to see over time that measures in use have helped to move the needle hopefully positively.

We also want to be cautious as we've talked a lot about over the last couple of days that we also haven't seen any unintended consequences as a result of that use.

So it's something we'd love your input on. This is really just I think the first year that we've actually implemented the broader lens on use and usability. Bruce got it right.

DR. CURTIS: Can I just follow up?

I mean, I think there are different ways that you can measure impact.

I think when we developed this measure nobody was talking or thinking about readmissions after PCI. And certainly the whole healthcare system has really evolved in their consideration of the importance of this particular aspect for pros and cons and differences of opinions. It's out there.

I can say that over the past 5 years since we published our first paper in JACC just describing the readmission rates there has been an abundance of literature coming out examining the issue.

So I think we have moved the conversation on this. The next will be to see can we move the actual rates.

CO-CHAIR HALL: The notion that you're only calling out -- going back to one of Karen's earlier remarks, that you're only calling out a percent or so at each tail. Do you have any internal plans or deliberations

around changing that as potentially a way to drive change faster?

DR. CURTIS: So let me throw that out. I think we are going to be cautious as an organization in terms of how we call out hospitals.

And I think that the real value for me is in reporting this information back to hospitals, that they get their risk-standardized rates, that they have -- that they're valid, that they're believable and that they are usable.

And the usable is I think what needs to evolve most rapidly. I have spent the last four years of my life as opposed to -- sorry, in addition to working on this measure working on developing the evidence that will support reducing these readmission rates.

And we're in the last stages of a mixed method study understanding both the qualitative and quantitative strategies

associated with lower readmission rates.

So I think what the ACC and what we're talking about, and I'm not going to guarantee that this will be done because I don't speak for them in that regard, but we're trying to create an environment where we will take that evidence, combine it with this information and create a campaign or an effort to try and systematically reduce these rates at hospitals.

And I think much more so than identifying high and low outliers that's the way we're going to push things forward. You have to develop the evidence. You have to package it up in a toolkit or some other change -- some process that promotes change in a positive force much more so than putting it in buckets.

MS. SLATTERY: And so if I can just add to Dr. Curtis' remarks which are totally in line and I agree with.

This is our first foray out with a

measure that the hospitals were offered the opportunity to publicly report on.

The College has a larger portfolio within PCI but also this year we will be implementing it in the ICD implantable cardiodefibrillator. We do have plans to expand public reporting opportunities for all hospitals across our registry initiatives.

So, once we start to look at that contextually we want to ensure consistency in how we report the information out. So we actually -- it's a separate workgroup that is taking that on and will be working through those approaches. And we will likely evolve it over time.

But I agree, it's meant to drive quality improvement and be engaging a dialogue and useful for consumers. So it's hard to balance out all of those.

CO-CHAIR HALL: Thank you. Larry?

DR. GLANCE: I'd just like to add my thoughts on this discussion.

I think that the burden on showing that there is an improvement in population outcomes with a quality measure should not rest primarily with the measure developer.

I think that as long as the quality of the information that's being provided in a quality metric is deemed to be acceptable or high then it really is up to the end user to make those -- to use that information and improve population outcomes.

So I think we should be hesitant about looking for improvement in population outcomes as a determining factor in whether or not a quality metric is usable or not usable.

CO-CHAIR HALL: Cristie.

MS. TRAVIS: Just a clarifying question on the public reporting because you said it just a couple of times. Is this a voluntary public reporting on behalf of the hospitals? And if so, about what percentage of those who are reporting are actually being willing to be voluntarily reported publicly?

DR. CURTIS: That's a great question. Yes, it is a completely voluntary public reporting. And if hospitals opted out there was no indication that they had opted out. So it was simply those hospitals that opted in, at least on Hospital Compare.

Of the I think 1,200 hospitals that had met our reporting thresholds 350 roughly decided to participate. I will say personally that was about 320 more than I thought were going to be willing to do it. So I thought for an initial year's effort for voluntary public reporting it was very successful.

MS. SLATTERY: I'll also -- just since we talked a little bit about implementation and you want to understand where we're going with it.

I think that that's incredibly admirable, the hospitals that opted in. Because one of the things for you to understand is we went from providing the

reports to the hospitals to giving them a window of about 6 to 8 weeks to make a decision to voluntarily report.

The lion's share of those 300 and some odd hospitals opted in at that point. There was only one other opportunity for those hospitals to opt in to have it publicly reported and that's why the number is stable.

Dr. Curtis also referenced the fact that if you chose not to report there's nothing reflected on Hospital Compare.

Moving forward the ACC has made the decision that we will be making it more obvious to consumers if a hospital is participating in our registries and had the opportunity to be able to report that and elected not to. And then differentiating some of the categories of decision-making in displaying that information out. So moving forward there will be more.

And moving forward we will continue to partner with CMS on reporting this

data on Hospital Compare because we think that that is a very valuable resource for consumers.

We also will have a complementary effort within our organization. And the reason for doing that is it allows for a mechanism for more rapid dissemination of information.

So, right now data gets reported to Hospital Compare on a quarterly basis. The infrastructure we will put in place on our website will allow that as soon as a hospital makes the decision to publicly report it will be made available to the public. And we can control the infrastructure better to be able to do that.

CO-CHAIR HALL: Do you know if the performance distribution in the reporting group is different than in the overall group?

DR. CURTIS: Yes.

CO-CHAIR HALL: Is there a concern that the public is getting the wrong

impression?

(Laughter)

MS. SLATTERY: So we did have some of the low-performing hospitals opt in for voluntarily publicly reporting.

I mean, again, I do think that in fairness to them the window of opportunity for which to make a decision was incredibly tight.

So, even for the second sweep they at best had about 12 weeks to make a decision, get leadership buy-in within their organization, get legal counsel to review it, and get the paperwork back to us for it to appear.

So we think that moving forward over time it will be easier for hospitals to get on board with this.

CO-CHAIR HALL: Thank you. Sherrie?

CO-CHAIR KAPLAN: To follow on to Larry's comment. And that is sensitivity to

change is one of these awkward things where you don't know what actually manipulations were going on that your measure should be responsive to. So, the attribution to sensitivity to change implies efforts to quality improve where you don't know what those are. And especially at the tails where we are the worst at estimating where a hospital actually might be.

Regression to mean pops to mind. And you think the trouble around the tails without actually having a response to chase.

So, is there -- is there any effort -- and distributional scoring will always have this property. Somebody always loses. So is there any push towards trying to find the mutable point beyond which a threshold one would declare a hospital as either better or worse or whatever?

Or is there any effort along those lines to shift away from distributional scoring?

DR. CURTIS: I think it's a great question and something that the field as a whole I think continues to struggle with. And I think we always default to the distributional because it's comfortable and familiar.

I think for readmissions specifically we don't know what the floor is. And I think until we kind of know and they start bunching up on the lower side I think it's reasonable to use the distributional but be attuned to the fact that since we know that there are unplanned readmissions that are not preventable that the goal is not to go to zero.

CO-CHAIR HALL: Larry?

DR. GLANCE: One last comment. Since our discussion now is about endorsement I'd like to point out that this like many of the other measures that we have looked at over the last two days is I believe a very robust measure methodologically speaking.

What differentiates this measure from many of the other measures that we've looked at is that this particular measure is largely based on clinical as opposed to administrative data.

And I think that's really a very, very important qualifier because clinical data is believed by most to be much more accurate and therefore have much greater face validity compared to administrative data.

CO-CHAIR HALL: Thank you. Same comments we had about the STS discussion earlier on. Jeptha?

DR. CURTIS: Not to shoot chart-based measures in the foot, but I do think it's important to recognize that they have different strengths.

And I think that, yes, we're really good at knowing whether or not a patient has diabetes or what the creatinine was and things like that. And that is important in terms of risk adjustment.

We're not as good as an administrative model I don't think at measuring frailty and accounting for frailty in a robust fashion. So, I think it's -- I appreciate the support and I hope that it's taken into account but I think there are two different schools of thought and there are competing strengths.

CO-CHAIR HALL: Thank you. I see no cards up so let's move to vote overall.

MS. SHAHAB: Voting for overall suitability for endorsement. One yes, two no and time starts now. Two more votes. One more.

CO-CHAIR HALL: There's some empty chairs down at the end of the table.

MS. SHAHAB: All votes are in. For overall suitability for endorsement for measure 0695 Hospital 30-day Risk-standardized Readmission Rates Following Percutaneous Coronary Intervention, 20 yes, zero no.

CO-CHAIR HALL: We thank the

developers and we'll ask the Yale CMS teams to come back to the table for the next measure.

CO-CHAIR KAPLAN: Okay, welcome to the developer. I'd ask you to do what we've been doing which is briefly introduce yourself.

And also as soon as you speak remind you please say again your names because the recorder back in the corner can't see you. And then briefly introduce your measure.

DR. BERNHEIM: Hi, this is Susannah Bernheim. I'm a director of quality measurement for the Yale Core Team many of whom you have met today.

Nihar, do you want to introduce yourself?

DR. DESAI: My name's Nihar Desai. I'm a cardiologist at Yale and an investigator at the Center for Outcomes, Research and Evaluation.

MR. AMIN: Do you have anyone on the phone?

DR. BERNHEIM: I don't know that we have anyone on the phone. Do we have anyone on the phone?

(Laughter)

DR. BERNHEIM: We have some support in the back and CMS here.

MR. AMIN: Okay.

DR. BERNHEIM: But I don't believe that there's anybody on the phone for this measure.

Okay, so I will just say a couple of words about this measure. This is a 30-day all-cause unplanned readmissions for hospitalizations -- following hospitalizations with an acute myocardial infarction.

This measure originally came to NQF in 2008 and this is its first time back for full re-endorsement which is a little longer than usual just because of the cycles of projects.

It has been in public reporting through the inpatient quality reporting

program since 2009 and last year was included in the first year of the hospital readmission reduction program.

It is designed much like our other measures where you look at a cohort of AMI patients. We look at readmissions 30 days later. I will talk a little bit about the planned readmission algorithm in a moment to identify unplanned readmissions. It uses the same hierarchical modeling approach as our other measures and it is a claims-based measure.

There have been a number of changes over the years that are detailed in your -- in the application. I'll just highlight the important ones.

When it first was reported we moved from a one-year measure to a three-year measure because many hospitals do not have a huge volume of AMI cases and that allowed us to report on a greater number of hospitals.

Other key changes. Early on we

excluded patients who were discharged against medical advice. That's true now with all of the readmission measures.

In reporting I believe two years ago this measure was expanded to include data from VA hospitals. So in the publicly reported measure it's now all CMS hospitals and patients hospitalized originally with their AMI at a VA hospital. So that was a neat collaboration between CMS and the VA which took a fair amount of work.

And then the one other big change has been that as part of the development of our hospital-wide readmission measure which this committee is going to be talking about later we created an algorithm that used claims codes to try to define readmissions that were planned or likely to be scheduled in advance, largely procedural readmissions that were not associated with an acute diagnosis code. And there's a long algorithm that does that.

And we developed that first in a

hospital-wide cohort and then we carefully checked it against the condition-specific cohorts.

And last year NQF held an ad hoc review of our condition-specific measures including this measure just to look more closely at this planned readmission algorithm. So that piece of this measure has come before committee at NQF previously, but the whole measure hadn't come back at that point for re-endorsement.

I think an interesting case was made earlier that improvement doesn't prove that the measure works or not. But I will say because we're really excited about it that in the 3-year cycle that was reported last December we are seeing for the first time declining national AMI readmission rates. And that's in the context of also decreasing admissions for AMI and potentially higher severity admissions. And big declines in mortality.

And so I think we're very reassured around questions of unintended consequences that the decline in readmissions is happening in the context of other improvements around AMI.

I think that that's probably enough of a quick overview of who know this measure well. But obviously we'll answer questions.

CO-CHAIR KAPLAN: Thank you very much. Paul, do you want to go first on the evidence?

DR. HEIDENREICH: Yes. I think there is -- we don't necessarily have a lot of evidence to know exactly how hospitals are improving. Clearly in 2012 there was a sudden drop in readmissions for MI as well as all. It seems to be most hospitalizations for Medicare patients. So I think there's clearly a strong rationale that one could make improvements.

CO-CHAIR KAPLAN: Larry? Nothing

to add? Other comments? Ready to vote?

MS. SHAHAB: Voting for 1(a) evidence. One is yes, two is no and your time begins now.

All votes are in for 1(a) evidence. Nineteen yes, zero no.

CO-CHAIR KAPLAN: Great. Performance gap. Paul?

DR. HEIDENREICH: So that has narrowed but I think still remains important. I think the 10 percent/90 percent went from 17.9 to 19.4 several years ago and it looks like with the last drop was down to 17.3 and 18.3. So the higher end clearly dropped although still a reasonable difference between the groups. And the overall rate some might argue is still too high. So I'd say there is a significant evidence performance gap.

CO-CHAIR KAPLAN: Larry? Other comments? Ready to vote?

MS. SHAHAB: Voting for 1(b) performance gap. One is high, two moderate,

three low, four insufficient. And the time begins now. Just one more vote.

All votes are in for 1(b) performance gap. Nine high, ten moderate, zero low and zero insufficient.

CO-CHAIR KAPLAN: Thank you very much. Priority?

DR. HEIDENREICH: It has been, I think probably still remains a priority for the government, for CMS to improve readmission rates for MI.

CO-CHAIR KAPLAN: Larry? Other comments? Ready to vote?

MS. SHAHAB: Voting for 1(c) high priority. One is high, two moderate, three low, four insufficient. And the time begins now.

All votes are in for 1(c) high priority. Fourteen high, five moderate, zero low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Scientific acceptability, reliability. Paul.

DR. HEIDENREICH: So, I think this has been felt to be moderate. I think they used a test/retest with a random sample and I see a reported ICC of 0.38 which is, you know, I think relatively common for this type of data.

CO-CHAIR KAPLAN: Larry? Other comments? Ready to vote? Oh, you have a comment.

DR. BERNHEIM: I just wanted to clarify it was brought up in the CABG just because people had taken some interest in this. We did do some work because when we create these ICCs we're not using six years of data to get a three-year sample size. So Lisa explained earlier we put together a correction factor. And we can share those details.

And when you do that for this measure it goes up to 0.48 just so people know. It's stronger. Our best estimate of what it would be with a full three-year sample size.

CO-CHAIR KAPLAN: Certainly well within what we've been seeing. So it's not exactly like this is a huge departure but thank you for that clarification.

Other comments? Ready to vote?

MS. SHAHAB: Voting for 2(a) reliability. One is high, two moderate, three low and four insufficient. And the time starts now.

All votes are in for 2(a) reliability. Three voted high, sixteen moderate, zero low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Validity, Paul?

DR. HEIDENREICH: Yes, there's been -- I think with the original submission I'm not sure if things were updated but the model, the overall model's discrimination had C statistics close to 0.6 and slightly under.

It sounds like interestingly it was -- I think when it was tested in the CCP project which had actual chart review it

sounded like there was a very similar C statistic. So it didn't seem like you were losing a whole lot from using, or if any at all of using administrative data.

So that seems to be I think very reasonable for this type of data. I think the exclusions as described are reasonable. I don't think there's been -- if anything the slight changes over time have improved -- improved the model since it was last approved.

I didn't feel there were any significant issues with missing data but we'll see if anyone else has concerns.

CO-CHAIR KAPLAN: Larry? Can you clarify the C statistic for us? Because I want to make sure everybody understands. What the magnitude of it.

DR. BERNHEIM: Sure. So you were right. When it was first developed, the technical report from development I think it was 0.58. In the most recent year of data -- we look at it each time it gets run. The most

recent three years it was 0.64.

DR. HEIDENREICH: Oh, I actually remember I did have one question. The -- I know it's been then tested in California data so now you can have an 18 and over measure. But I didn't -- it wasn't clear to me if there was a significant improvement or decrease in the model's performance.

DR. BERNHEIM: So that's a great question that I don't remember the answer to. In general our measures often do slightly better in the all-payer data sets. We think that's because the comorbidities are even more powerful predictors in younger populations that have fewer of them.

And I can quickly find you the answer to this for this particular one. So 0.67 was the discrimination for the 18 and over measure. And the correlation between the two was 0.998.

CO-CHAIR KAPLAN: Larry?

DR. GLANCE: So, just as a quick

point of clarification. As was alluded it's very common to see a model perform differently in data sets.

And specifically if you're going to use an all-payer data set you expect to see more heterogeneity in the patient population and therefore as a result of that you'll see better discriminatory power.

CO-CHAIR KAPLAN: Thank you. Other comments? Frank.

DR. BRIGGS: What was the impact of expanding the planned readmission algorithm?

DR. BERNHEIM: Great question. Again, numbers I don't have at the tip of my tongue.

It reduces the overall readmission rate very slightly. Although for this measure, let's see if I can find it rapidly for you.

CO-CHAIR KAPLAN: So refreshing to see people throwing pages as opposed to

scroll.

(Laughter)

DR. BERNHEIM: Exactly, not be on a screen. Okay. Planned readmission algorithm I don't see here.

If somebody on the phone from our team has this handy please speak up. And I will otherwise find it quickly but it'll take me just a minute.

Can the operator open the lines and make sure that somebody on our team is available to answer this question quickly? Because they have it at their fingertips quicker than I do.

MS. KHAN: What are their names? The people that are on the phone.

DR. BERNHEIM: Chanch Nabat, are you there?

OPERATOR: All lines are open.

DR. BERNHEIM: Okay, great. Anybody on the Yale team have this number handy quickly so I don't have to make this

poor, tired committee wait while I flip pages?

MS. GEARY: Hi, this is Lori Geary at Yale. Can you hear me?

DR. BERNHEIM: Yes.

MS. GEARY: We are pulling that up now. Bear with us one minute.

DR. BERNHEIM: The easiest place may be the NQF application where we brought it back last year.

MS. GEARY: Okay. From version 1 to version 2 it went from 19.7 to 19.0. I'm sorry, 18.7.

DR. BERNHEIM: Okay, so a percentage point. We should have gone with my guess, I was right. Thank you, Lori.

MS. GEARY: Okay.

CO-CHAIR KAPLAN: Thanks for that clarification. Any other comments? Ready to vote?

MS. SHAHAB: Voting for 2(b) validity. One is high, two moderate, three low, four insufficient and your time begins

now.

All votes are in for 2(b) validity. Four high, fifteen moderate, zero low and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Feasibility?

DR. HEIDENREICH: Based on claims data, highly feasible.

CO-CHAIR KAPLAN: Larry? Other comments? Ready to vote?

MS. SHAHAB: Voting for feasibility. One high, two moderate, three low, four insufficient. Time starts now.

All votes are in for feasibility. Eighteen high, one moderate, zero low, zero insufficient.

CO-CHAIR KAPLAN: Thank you. Usability and use. Paul?

DR. HEIDENREICH: Well, it already is being used. And you know, one could argue it's been successful given that it's been used both I think for public reporting as well as

for payment.

And we've seen the expected changes at least within 2012.

CO-CHAIR KAPLAN: I have one quick question. If it were possible, would it be possible to investigate actually the sensitivity to places where you actually knew there were efforts underway to improve this? And so responsiveness to change could actually be estimated.

DR. BERNHEIM: So I think what you're asking is could we focus in on places that we know are making a big effort around this and show that those efforts are playing out. I mean I think that's a great research question. It's not something we've done on our team, but I think it is an important link that the research is slowly building to show. And there are some trials out there that have shown particular interventions work AMI patients.

CO-CHAIR KAPLAN: Thank you. That

raises the specter of the R word so we're moving on really quickly. So, for usability and use are there any other -- sorry, Karen?

DR. JOYNT: I think this is a great example of ways in which the same measure can be used a lot of different ways.

In the way that it's used in public reporting the -- of the 4,464 hospitals in the U.S. 23 are identified as being better than average, 2,327 are no different, 29 are worse and 2,085 are number of cases too small.

That's a lot of effort for hospitals to make for this amount of discrimination. And I know they get more information than this.

On the completely opposite side as the weight of the readmissions penalty is calculated in which there's no uncertainty built into the model if you're 0.001 percent worse than predicted on dollars, not even on rates, that you will in theory get a penalty.

And I don't know that there's any

way in this committee to address the different ways in which something can be used, but it certainly points out that the statistical arguments that we have about the way that this thing is done can go any number of different ways when things are put forward.

And to me the way that things are used actually is really important to how the measure is going to work. And I might personally choose different ways of calculating this based on whether it was going to be used for public reporting or for pay-for-performance or whatever this is going to be.

So I just have concerns about this sort of blanket blessing of measures when they can be used in such vastly different ways. That may be a bigger problem in this particular measure but I think I'd be interested in hearing from the developers sort of how we should think about what your model can do.

The data to suggest exactly what you're saying which is what we know can and can't work to improve readmissions is not great. And I'd be interested in knowing from your work are the people that are identified as outliers doing something differently.

What's happening as a result of the way that this metric is being used?

CO-CHAIR KAPLAN: Hold on a second.

CO-CHAIR HALL: A couple of things. Karen, your sentiment is a longstanding one. It's been heard in many, many NQF forums over the years. And in fact I think -- I won't jump the gun, but in the upcoming white paper coming out from NQF there may be some commentary about NQF increasing its guidance around recommended uses or uses where a particular measure seems most appropriate. So we won't change that aspect of it today but your sentiment has been heard many, many times.

If there's --

DR. JOYNT: And if this needs to be tabled I'm totally fine with that. I'm struggling a little bit with sort of the questions around usability for us to sort of say as it's being used what are the negative consequences, what's happened. And we don't really get that data --

CO-CHAIR HALL: Absolutely. And that's been -- I think we all sympathize. That's absolutely been the case.

But so if there is something you would like our developers to state please rephrase that. I didn't want to cut you off, but I do want you to know that that is a longstanding concern and there may be some change in the air around that concern. But it won't happen today.

Is there anything you do want the developers to state? Okay. Any other comments or concerns then?

CO-CHAIR KAPLAN: Just to say that

some of us share your pain. And the hopefully MAP group will begin to address these kinds of issues.

So there's a safety clause built in but it's not sufficient for some of us to kind of feel cozy about these decisions.

Okay, Susannah.

DR. BERNHEIM: I will decline from commenting on the pieces that you guys want not commented on. I just wanted to make sure that people know that a part of the work that we do with CMS is explicitly monitoring for unintended consequences. And we do some surveillance work. And there is a constant measure or maintenance process. So just in terms of the unintended consequences piece that is a part of the expectation as part of the measure life cycle.

CO-CHAIR KAPLAN: Thanks for that clarification. Any other comments? Are we ready to vote usability and use? Okay.

MS. SHAHAB: Voting for usability

and use. One high, two moderate, three low, four insufficient information and the time starts now.

All votes are in for usability and use. Four high, fourteen moderate, zero low and one insufficient information.

CO-CHAIR KAPLAN: Thank you very much. Moving onto suitability for endorsement. Paul?

DR. HEIDENREICH: I think no additional comments. I'd say it meets endorsement in my opinion.

CO-CHAIR KAPLAN: Larry?

DR. GLANCE: I think this measure is also very typical of most of the CMS measures that we've heard in other measures. It's very robust in terms of the methodology and I would also vote for endorsement.

CO-CHAIR KAPLAN: This is Washington so I won't say "robustitude" in this audience but it does -- so we are voting on its suitability for endorsement. Ready to

vote?

MS. SHAHAB: Voting on overall suitability for endorsement. One yes, two no. Time starts now.

All votes are in for overall suitability for endorsement for measure 0505 Hospital 30-day All-cause Risk-standardized Readmission Rate Following Acute Myocardial Infarction Hospitalization. The results are 17 yes, 2 no.

DR. BERNHEIM: Thank you.

CO-CHAIR KAPLAN: Thank you to the Yale group.

CO-CHAIR HALL: We'll invite the next Yale group to the table.

(Laughter)

CO-CHAIR HALL: We'll be discussing 2539 Seven-day Risk-standardized Hospital Visit Rate after Outpatient Colonoscopy. I think in some sense there's a little bit of a twist compared to some other topics we've discussed. So we'll wait for our

developers.

No, that was not a pun.

(Laughter)

CO-CHAIR HALL: So we have 18 people in the room. No one's permitted to leave because we lose our quorum. So raise your hand if you have to do a number one.

(Laughter)

MS. KHAN: We should be done by 2 I think.

CO-CHAIR HALL: We'll continue to push on. Our colleagues from Yale, please introduce yourselves and your measure.

DR. DRYE: Hi, I'm Elizabeth Drye from Yale.

DR. RANASINGHE: My name's Isuru Ranasinghe from Yale.

DR. DRYE: I was just going to say following up on Bruce's point that this is a little bit of a twist. This is the first measure we're bringing to NQF that is for outcome of ambulatory care. So we're really

excited about it. It is new terrain. And Isuru led the work and he's going to walk us through.

DR. RANASINGHE: Okay. So it's Isuru here again.

I'll start off by doing a quick summary of the measure and the rationale for the measure.

So the measure is a measure of unplanned hospital visits following outpatient colonoscopy. And that is colonoscopies performed in hospital outpatient department ambulatory pre-surgical centers and physician office settings.

The denominator for this measure is low or moderate risk colonoscopy procedures. And based on our inclusion and exclusion criteria we actually capture about 94 percent of all outpatient colonoscopies performed.

The numerator for this measure is unplanned hospital visits within 7 days of the

procedure. And hospital visits include ED admissions, observation stays and inpatient admissions.

Now, that's a very broad patient-centered outcome that captures adverse events that are related to the bowel prep, the anesthesia itself and the procedure.

Now, this measure is really important because of four critical reasons and I'll outline them.

First is that colonoscopy is incredibly common. So this is the most common procedure performed in the outpatient setting. We see an outcome rate of about 16.2 per 1,000 procedures in the Medicare data and we see significant variation between 8 to 20 per 1,000 between providers. So there is a facility-level variation in quality.

And if you extrapolate that to national data that's about 27,000 hospital visits following colonoscopy procedure nationally.

And that's really important because most of these procedures, we know two-thirds to three quarters are screening colonoscopies. So by definition these are procedures occurring in relatively healthy people who don't have signs or symptoms of a disease. And ensuring monitoring quality in that group is -- there's a strong mandate for measuring quality.

We also know when we look at the top diagnosis that many of these patients come back in with serious and very mild things, and potentially preventable things. Things like abdominal pain, nausea, bloating, bleeding, perforation, syncope, aspiration because of the anesthesia. So we think these are really important things to measure and potentially preventable.

The key thing with this is many of these providers in the outpatient setting are completely unaware of these events. They're simply invisible.

So for example, there's a study from Boston that suggests that about 80 percent of these visits the providers are actually unaware of. And you can understand that. So if you're an ASC, ambulatory surgical center, you're legally not allowed to provide follow-up care. You can understand why patients would present to a different provider in the event of an adverse event. So we think this measure is really important for illuminating quality.

And if I can finish by very quickly saying that this measure was developed with input from working group that -- we had input from Dr. Ron Bender and John Allen who are heads of the American Gastroenterology Association and the American College of Gastroenterology.

We were extremely fortunate to have them and they had a huge input into shaping this measure.

We were also very lucky to have

the highly qualified technical expert panel which provide us great insights into this measure. And a period of public comment which we addressed many of the issues that came up.

And I must say overwhelmingly this measure has been supported by that group. So we think that provides a lot of rationale.

This measure uses Medicaid data so it's eminently feasible. It's risk-adjusted. It uses a hierarchical model. We can statistically determine the outliers.

And just lastly, we submitted this measure for approval to MAP during the development process and we have received conditional approval from them.

CO-CHAIR HALL: And just immediately clarifying, the accountable entity is the outpatient department or the ambulatory surgery center?

DR. RANASINGHE: That's right.

CO-CHAIR HALL: All right. So, thank you for that introduction to the

measure. We're in the category of evidence. We'll ask Ross and/or Cristie if they would like to open the discussion.

DR. EDMUNDSON: Sure, I can start here if that's okay with you, Cristie?

MS. TRAVIS: Please.

DR. EDMUNDSON: Okay. On evidence, so you've already alluded to some of the evidence. But there's not a lot. This is a new measure.

I think -- you mentioned four reasons why this is important. I'd add a fifth one. When I need my colonoscopy I don't want a complication. And that's important to everybody because this is our general population. If you live long enough you should have one of these.

So the evidence that is literally reviewed shows complication rates from 20 to 34 percent. And then you did have on HCUP data unplanned hospital visits ranging from 8.2 to 20.1 per 1,000 colonoscopies.

DR. RANASINGHE: That's right.

DR. EDMUNDSON: And I think that -- so there's some evidence out there although it's light at this point in time for the evidence.

Cristie?

MS. TRAVIS: No, just that you all did document at least some of the interventions that you thought would be possible to actually improve upon the measure. So from my perspective it meets the evidence for health outcome.

CO-CHAIR HALL: Any other comments or concerns? We'll consider this an intermediate outcome for now. No other comments. No cards. Let's vote on evidence.

MS. SHAHAB: Voting for 1(a) evidence. One is yes, two no. Time starts now. One more vote.

All votes are in for 1(a) evidence. Fourteen yes, four no.

CO-CHAIR HALL: Performance gap

opportunity.

MS. TRAVIS: There does appear to be variability in performance. The standardized range was from 8.3 to 20.1. So there seems to be quite an opportunity for improvement.

CO-CHAIR HALL: Additional comments?

DR. EDMUNDSON: I agree. No other comment.

CO-CHAIR HALL: Let's move to vote.

MS. SHAHAB: Voting for 1(b) performance gap. One high, two moderate, three low, four insufficient. Time starts now.

CO-CHAIR HALL: Let's try again.

MS. SHAHAB: All votes are in for 1(b) performance gap. Seven high, eleven moderate, zero low, zero insufficient.

CO-CHAIR HALL: Priority.

MS. TRAVIS: This is a U.S.

Preventive Services Task Force recommendation as was talked about since we're recommending that people go get this as a screening test that we have an obligation to be sure they're getting a high-quality. I think I was impressed by the 14 million colonoscopies that were done back in 2004 alone. So a very high-frequency procedure that needs to have the quality measured.

CO-CHAIR HALL: Additional comments? No additional comments.

MS. SHAHAB: Voting for 1(c) high priority. One high, two moderate, three low, four insufficient. And your time begins now.

All votes are in for 1(c) high priority. The results are 12 high, 6 moderate, zero low, zero insufficient.

CO-CHAIR HALL: Thank you. Moving into scientific acceptability, reliability.

DR. EDMUNDSON: The reliability was on a 2010 split population arm. Large numbers of colonoscopies.

And I believe your -- the ICC on that was 0.335 judged as fair. And I think that's the information provided. Is that correct?

DR. DRYE: Correct. We did, as you heard on our other measures, we have recalculated that with the Spearman-Brown prophecy formula. It's an interesting name. And it's at 0.43. That gives us a better estimate if we had a full data set.

I would just note I think as you're all aware the outcome rates for this measure are lower than they are for, for example, AMI readmission. And so we have to get a bigger sample size to get reliable measure score results.

And we're happy with what we're seeing here but this is an inherent challenge in a healthier population with a high volume of procedures but a lower outcome rate.

CO-CHAIR HALL: Any additional comments or concerns on reliability? Seeing

none.

MS. SHAHAB: Voting for 2(a) reliability. One high, two moderate, three low, four insufficient and the time starts now. We need one more vote, please.

All votes are in for 2(a) reliability. The results are 1 high, 17 moderate, zero low, zero insufficient.

CO-CHAIR HALL: Moving into validity. Opening remarks.

DR. EDMUNDSON: Okay, on validity you had your technical expert panel that drove a lot of I think the validity questions as I read through your information.

In addition, you had split sample two years of data. And on that C statistics of 0.67. And your conclusions were that this is good model discrimination.

The other piece of information that you provided that I'd like to have you comment on is the Charleston model. Better than Charlson model in that I believe this is

your risk stratification tool. And the Elixhauser model of your C statistics actually being better than those as a risk model. Would you comment, please?

DR. RANASINGHE: So, this is Ranasinghe here again.

So, we did two things. One was that we compared the C statistics between development and the validation sample, and then again in using data. So we developed our data using the 2010 sample and then we validated it again in the 2011 sample.

The other validation step we did was to construct the risk adjustment model using -- we constructed a risk adjustment model based on our conceptual and statistical understanding of what we total predict hospital visits in this population.

But we wanted to benchmark against a risk model that's already being used. And one option was to use the Charlson and the Elixhauser models which are widely accepted.

Please keep in mind Charlson and Elixhauser are unknown to gastroenterologists, specifically -- not colonoscopy-specific. In fact, the Elixhauser model was to predict mortality from what I understand.

But this would give us a good idea of where our model sits and our C statistics. And the model characteristics actually ended up being better than both those models.

DR. DRYE: This is Elizabeth Drye. I would just add that, again, we're in a novel data environment in a novel setting. We don't have other, you know, there are not like 5 or 10 other readmission measures, similar measures we can compare it to.

So even though we were hoping with thoughtful variable selection with a lot of clinical input that we would do better than these indices we felt we should at least use some other approach to make sure we were getting what we thought we should be getting.

CO-CHAIR HALL: Karen?

DR. JOYNT: I just have a few additional questions. One is I know there's not a lot of other quality measures in this group. But is there anything else you can compare this to? Do we know if procedure volume matters for colonoscopy? The rate of detection of abnormalities I know has been proposed as a quality. Is there anywhere else to sort of externally validate whether or not the rates that you're seeing are indicative of bad procedure as opposed to sick patient?

DR. RANASINGHE: It's Isuru here again. It's a great question and one that we find really challenging because really there is no outcome measures full stop for any of the ambulatory measures. And we didn't really know what to compare against.

And in fact we -- our conclusion we reached was that there is no measure that we could adequately compare against.

The only sort of measure that comes close is the NQF measure that -- NQF-

approved measure for ambulatory surgical centers which does an immediate transfer following the procedure. But at the time we didn't have the actual reports of those values. Individual hospital-level values were not reported for that measure for us to compare against.

DR. DRYE: I would just add also that -- Elizabeth Drye again -- we relied heavily on looking at the reasons for the ED visits, observation stays and readmission and just thinking clinically, you know, are these likely to be just sick patients, or are they, you know, did they look like they're related.

And also we know we're dealing with a patient group and a procedure that typically would not be done in an outpatient setting on patients who were acutely ill for other reasons.

We did pull out of the measure those patients who we felt might end up in the hospital for the reason for which they were

having the colonoscopy. So, IBD, inflammatory bowel disease patients, patients with history of diverticulitis are pulled out of the measure. We had a lot of discussion around that and comment on that issue.

We also pull out of the outcome admissions for planned care like colorectal resection. And that's one-third of all hospital admissions that we see following colonoscopy. So we tried to triangulate or whatever, get at that concern in a number of different ways.

We also looked at the baseline admission rates and we looked at the falloff in the first few days to try to pick the outcome time frame. And we're confident that we are zeroing in on hospital visits that are really related to the procedure versus patient factors that we're not adjusting for. But it took a lot of different strategies.

CO-CHAIR HALL: Sherrie?

CO-CHAIR KAPLAN: I was just going

to sort of notice, and this is not for slowing the conversation down or anything, but different from the dialysis measure where the attribution was back to the hospital, right, for readmission to the hospital.

CO-CHAIR HALL: No, it was to the dialysis.

CO-CHAIR KAPLAN: It was to the dialysis center. So then it does follow the same issue.

So the skill in that case of the colonoscopist so confounded with center. At that time there was some discussion about the interest of the physician-level stuff and the potential for estimating things, or attributing things to the outpatient center versus the providers.

And there was some conversation around that. And I didn't want to -- I don't want to cause a -- stir up a storm here, but I did want to kind of for fairness of comparison raise that issue.

CO-CHAIR HALL: Actually in a similar place because we're assuming that the facility has some impact on quality and it's not just the operator's skill and proficiency.

DR. DRYE: That's a great issue. I think it's come up before.

The reason that we put the measure at the facility level, there's a couple of reasons.

One, there is a component of facility care that we think contributes to the outcome. There's the anesthesia care, there's the post-op care, there's decision about when a patient is ready to go home.

These facilities, a lot of them, particularly ambulatory surgery centers are physician-owned and they tend to specialize and they may have a couple of physicians who regularly work there.

But in general the physicians do colonoscopies in multiple settings too. So if you wanted to look at volume at the facility

related to outcome you might not necessarily be getting at the volume that you were trying to capture which would be the per-physician, at least for the physician component.

We're confident given the nature of things that we see people go back to the hospital for that there is a facility component. And we expect that there's a consistency at which physicians are doing the care.

Also, as I mentioned before we need to get a certain volume of patients and outcomes to be able to get a reliable estimate. So for all those reasons we housed it at the facility level.

Like the other measures -- this measure would be, if it's implemented by CMS would be implemented in the hospital outpatient department. I mean, sorry, the hospital outpatient prospective -- OPPS system as well as under the ambulatory surgery center program.

So, they have not had this type of measure as far as I know implemented there. And we have encouraged CMS to follow the approach that is taken for the hospital-based measures in which facilities get patient-level data, they can see which patients are in the measure and where they ended up which are things they won't otherwise see. And we think that's going to be really critical.

And they won't be able to see in there who the physician was, who the anesthesiologist was, et cetera.

CO-CHAIR HALL: Paula.

MS. MINTON-FOLTZ: Does this include immediate admission to either obs? Or is there is a 24-hour lag?

DR. RANASINGHE: So this includes sort of direct admissions from ASCs or from HOPDs. Hospital outpatient departments.

And the rationale was that this is -- colonoscopy procedures should be straightforward procedures, range from 30

minutes to an hour. There's really no reason for the patient to stay for an extended length of time unless some sort of adverse event occurs.

MS. MINTON-FOLTZ: Well, there are those patients who have no ride even. We've seen those. But it's not a large percent.

CO-CHAIR HALL: Frank?

DR. BRIGGS: I was wondering if you had data in regards to the how did you come up with the seven days, seven days for this type of procedure. And the side effects that you're describing trying to capture seven days actually seems a little bit long. And you're thinking the first day, two days, maybe three days. Out seven days you're probably looking at continuation of symptoms and things like that. So I was wondering if you had data specifically to support your cutoff of seven.

DR. RANASINGHE: So that's a great question. So there's a range of side effects that could occur after a colonoscopy. And

they range from within a few days for things like perforation but up to 30 days for things like bleeding, GI bleeding. And so there's a definite phenomenon of delayed GI bleeding.

We know from the literature that a vast majority of those complications or adverse events occur within the first seven days.

And we can empirically test that by looking at the number of hospital visit per each day post procedure. And what we see is a curve which sort of levels off to after about seven days. And that's why we picked the seven-day time window because we thought that would give us the best sort of quality signal for our measure.

So it doesn't -- and that was supported by our technical expert panel. It does mean that we miss some bleeding events that are delayed. But we specifically excluded them because we did not want to capture hospital events that aren't related to

the procedure.

CO-CHAIR HALL: Wes?

DR. FIELDS: Yes, a couple of comments to Sherrie's point and then one question.

The comments. Hopefully we'll see many outpatient measures in the future. I think this one actually has a pretty elegant design.

But to Sherrie's analogy one of the things I like about this is that they work pretty hard at the development level to identify what's essentially a well population. So that's quite different than the standard dialysis patient who's by definition chronically ill.

The other is this is pretty much of a 1 to 1 linkage between the physician doing the colonoscopy than the procedure and the outcome. And that's not necessarily true of dialysis centers in terms of who is actually providing the service and whether

it's a passive or active physician activity.

But one question about the design measure. I'm just sort of curious why you chose not to include unscheduled follow-up with a primary care physician or clinic during the first seven days.

DR. RANASINGHE: So that's a great question. So, it's very, firstly, we considered all outcomes that are possible.

We felt the acute care user or visit to a hospital is something that's unexpected following a colonoscopy procedure. And so that we thought would reflect a clear quality signal.

Whereas visits to a primary care provider could be planned, could be appropriate care, could be scheduled care. And that's very hard for us to identify from claims data. And that is I guess the primary reason.

CO-CHAIR HALL: You okay with that, Wes?

DR. FIELDS: Oh, I'm very okay with it. And for Helen and many other people around the table I think the measure developers make the point that there are many situations where a component of primary care or first contact care can happen someplace besides a primary care clinician's practice.

And that's part of why I'm fundamentally uncomfortable along with 30,000 of my close friends in emergency medicine with the implied negative metric that goes along with ED visits in many of the measures.

CO-CHAIR HALL: Point well taken. Ross?

DR. EDMUNDSON: Yes, thanks for bringing up the subject here.

I wrestled with this a lot because I was one of the telephonic calls saying well, my first impulse was this should be attributed to a physician.

And as I wrestled with it I thought, no, I have to address the methodology

that's in front of me. And I think there's very good value to knowing what particular facility had what kind of outcomes of increased visits into the emergency room or admissions.

I think there's value there. I personally though, my bias is that if you did the study and you attributed it, it's the same information. And I agree with you it's a 1 to 1 relationship with a provider, a physician at this point in time.

That physician is the one who will meet face to face with the patient beforehand, gives the instructions, gives the prep, does the same prep and in fact will very often say, well, where would you like to have this done.

We can do it in the hospital, you can do it in my center or you can do it at this -- based on their whim, their preferences, the day of the week, the convenience to the physician.

So, I think it would be a better

measure of quality tied to the physician. But I do think it does have as it's stated and presented to us some value as a facility as well.

CO-CHAIR HALL: Thank you, Ross. Karen?

DR. JOYNT: Just a quick question. First, I commend you for again trying to reach into the outpatient setting and look at something that I think is probably under-studied because of the difficulties like that.

Is there anything we should be thinking about differently? Because this is a composite of things that are very different. Sort of a quick visit to the ED versus an observation stay versus a hospital admission.

Or are these do you think, given the healthy population that you've selected we should consider them all to be within seven days equivalent events?

DR. DRYE: So that's a great question. We like to think of them as like

they're above a certain threshold of acuity. You feel bad enough that you have to go to the ED.

And they may reflect different things. They may reflect that you can't get into your primary care doctor's office because they're just not accessible to you the next day or that evening or whatever.

And so we're just saying is this a threshold effect, either because of the serious problem or because of a problem that could have been prevented, or could have been cared for more efficiently in an outpatient setting. That's where the patient is going.

Are they all equal? They're not all equal. And I think, you know, again we feel really strongly the data has to be reported back at the patient level with the reason and the location of the hospital visit so that providers can look at that.

I think, my own view is if three, you know, gastroenterologists or general

surgeons are at an ambulatory surgery center doing this care and one is having outcomes that are really driving those scores, that's a lot of peer review and direct pressure back on one provider.

So, but it gets back to your point about how things get put into use. I think with that information available to facilities this works well as a composite and it gives us the volume that we need.

CO-CHAIR HALL: Thank you. Ross, another question? You all right? I have a possibly small question but I'm curious as to why you risk-adjusted for polypectomy.

DR. RANASINGHE: So that's a great question. We know from the literature that polypectomy is associated with bleeding, GI bleeding, and that's the strongest risk factor for bleeding.

At the same time there's a lot of debate about, you know, removing the polyp could be discretionary, that some providers

may remove unnecessary polyps and you might be -- you're adjusting away that quality signal.

But we felt removing -- identifying polyps and removing them is a quality indicator itself of colonoscopy. We did not want to -- you know, if somebody appropriately removes a polyp and that resulted in a high rate of hospital visits we didn't want to disadvantage these providers who are taking appropriate action to treat a condition.

CO-CHAIR HALL: I understand that, but you could also argue that there's perhaps proficiency involved in taking polyps out without causing bleeding.

DR. RANASINGHE: I just want to make clear that we do not adjust for the technique used in removing the polyp, the number of polyps. There's a number of specific techniques for removing polyps. So that technical component we do not adjust for. And that is up to the discretion of the

provider.

CO-CHAIR HALL: I understand. I guess it looks to me like a part of the therapy that could cause the event, the complication, and you may be adjusting for it.

Any other? Paul?

DR. HEIDENREICH: Well, just it sounds like that should be the next measure, a companion measure of polypectomy rates for those undergoing polypectomy.

CO-CHAIR HALL: Or adenoma detection rates perhaps. Other comments in the category of validity? Seeing none.

MS. SHAHAB: Voting for 2(b) validity, one high, two moderate, three low, four insufficient and time starts now.

All votes are in for 2(b) validity. The results are zero high, 18 moderate, zero low, zero insufficient.

CO-CHAIR HALL: Feasibility. Opening remarks? Anyone? Sorry, go ahead, Cristie.

MS. TRAVIS: Administrative claims and therefore feasible to collect.

CO-CHAIR HALL: Other comments? Seeing none.

MS. SHAHAB: Voting for feasibility. One high, two moderate, three low, four insufficient and the time starts now.

All votes are in for feasibility and the results are 14 high, 4 moderate, zero low, zero insufficient.

CO-CHAIR HALL: Usability. Opening remarks? Ross or Cristie, any specific remarks?

MS. TRAVIS: Nothing specific. It's just not in use yet. But they did talk about the fact that if it was publicly reported or later used in payment methodologies and talked a little bit about using the same methodology as for some of the other CMS measures.

CO-CHAIR HALL: Kathy?

DR. AUGER: I just wanted to comment on a potential unintended risk. As the developer mentioned a few minutes ago if you're in a group of three people at an ambulatory care practice and one of those providers is really driving the outcome and you end up in an outlier as perhaps a bad performing site.

As a member of the public knowing that might actually give you anxiety about your physician's ability which may be completely misattributed. You might actually be wrongfully attributing risk to a good doctor as opposed to one that has higher adverse outcomes. So, there's the potential for misattribution of risk. But I suppose you could also argue that you could then put pressure on that lower performing physician as well. So I could go either way.

CO-CHAIR HALL: Karen.

DR. JOYNT: I may just be reading this wrong so feel free to correct me if I'm

mistaken.

But in the long report it looks like from the HCUP data that one facility was found to be better than expected and four were found to be worse than expected.

And I feel like from what you told us at the beginning about the variability in outcomes that that either strikes me as the sample size is too small or I misunderstood the variability. Or this isn't the same you were talking about. So just a clarification on sort of how this plays out in the real world data would be helpful.

DR. RANASINGHE: Okay, that's a great question. So, we developed our measure using a 20 percent Medicare sample. And because that's a sample we needed to actually test the measure score. We need a 100 percent sample.

And for that we used HCUP data from four states. And that outlier analysis is data from four states only. And that's

about 992 facilities. So we expect many more outliers. If you actually included the national population I think there's about 8,000 facilities plus nationwide.

The other point I need to make is that that is using the 95th percentile and that's a policy decision. So if you wanted to capture more outliers you can change the cutoff interval.

CO-CHAIR HALL: So with the 5/95 interval you're working at about a percent, 1 percent roughly more or less? Okay.

Other comments? Usability, other comments or concerns? Not seeing any.

MS. SHAHAB: Voting for usability and use. One high, two moderate, three low, four insufficient information and the time starts now.

All votes are in for usability and use. And the results are 1 high, 16 moderate, 1 low and zero insufficient information.

CO-CHAIR HALL: Any summary

comments for overall voting? I do not see any. Ross?

DR. EDMUNDSON: I think this is a very -- a high-frequency outpatient procedure. And this is a start. This is shining the light on it.

And I think that there's a lot that we're going to probably have to learn out of this one. And I think like as I say, my prejudice is that I think it will be better to look at this from an individual provider. But I'll leave that for time and further information to decide.

CO-CHAIR HALL: Thank you. Any other comments. No? Let's move to vote.

MS. SHAHAB: Voting for overall suitability for endorsement, one yes, two no. The time starts now.

CO-CHAIR HALL: We're waiting for one more if you could just click it one more time.

MS. SHAHAB: All votes are in for

overall suitability for endorsement for measure 2539 Facility 7-day Risk-standardized Hospital Visit Rate after Outpatient Colonoscopy. And the results are 17 yes, 1 no.

CO-CHAIR HALL: We thank the developers for their effort and input.

And we will move into the review of 1789. I'm turning over to who? Who am I turning over to?

MR. AMIN: Okay, so --

CO-CHAIR KAPLAN: Before you start, is part of your start going to be what we're supposed to do here?

MR. AMIN: Yes.

CO-CHAIR KAPLAN: Thank you.

(Laughter)

CO-CHAIR HALL: And we welcome the Yale team back to the table. They'll introduce themselves in a moment.

MR. AMIN: Okay. So, as part of the 2011 evaluation of measure 1789 the

Hospital-wide All-cause Unplanned Readmissions Discussion there was a number of elements that the steering committee requested from the developer post dry run, especially trying to understand the results of the dry run, specifically an analysis of the distribution of performance between hospitals with varying proportions of low SES patients and the proportion of the measure result variation that is attributable to providers compared to patients.

This information was not available during the initial endorsement of this measure and so given that we -- this is the first time this committee has met since the evaluation of 1789 we've asked the measure developers to provide this information to the committee.

And I will turn it back over to -- and provide a quick update on the progress related to harmonization to measure 1768 which was provided to members of the committee that are returning.

So there are a number of members of the committee that are aware of this issue. And I would sort of invite you to participate in this conversation.

I'll just note who that is. Bruce, Cristie, Jo Ann Brooks, Paula, Larry Glance, Leslie Kelly Hall and Sherrie Kaplan of course.

So, that is the topic of the next half hour.

CO-CHAIR HALL: Thank you. And just to clarify, we will not actually be asked to vote on anything.

MR. AMIN: No.

CO-CHAIR HALL: We're being asked to review the information.

MR. AMIN: Yes, these are purely updates from the developer and conversations that the committee may want to have with the developer related to those topics. But this measure is not up for review and there will not be an endorsement decision from this

conversation.

CO-CHAIR KAPLAN: And the goal of this conversation is to generate?

MR. AMIN: Is to provide an update to the committee. This was --

CO-CHAIR KAPLAN: Not on their side, on our side. What are we to do? We're to provide what?

MR. AMIN: This was a request by this committee at the end of the last review. So this is an update on these particular issues. There's no action required by the committee at this point.

DR. BURSTIN: And just to add to that. So, when the decision was made to endorse this measure there was -- some of you may remember this was not without controversy. Susannah is still smiling, that's good.

But part of the agreement particularly with the NQF board as well as with the steering committee was that they wanted us to take a look back to see what the

experience has been, to see if there is any evidence of unintended consequences and sort of monitor the situation. This is essentially that monitoring update.

CO-CHAIR KAPLAN: So, essentially -- not to keep beating this horse, but essentially our role is to thank you for sharing and we are -- any comments we have for the developer will be communicated to --

MR. AMIN: They will be communicated in the report around the updates on these topics.

DR. BURSTIN: And certainly since the developers and CMS are here we want them to be part of this discussion. If there are issues that are raised that need further discussion we will encourage those discussions.

CO-CHAIR KAPLAN: Could the developers please re-introduce yourselves for the record and then present your findings.

DR. HOROWITZ: I'm Leora Horowitz

from Yale.

DR. BERNHEIM: And I'm Susannah Bernheim.

DR. HOROWITZ: So I led the development of this measure and presented to this committee in 2011. And thank you. I know it's been a long two days so thanks for bearing with us.

So, as you know this measure was endorsed in the spring of 2012 and at that time you had asked us to come back earlier than the three years to talk about the dry run and the harmonization. And so I'm going to just quickly summarize those. And I believe you've received those materials as well.

So, we had the dry run in the fall of 2012. Dry run means that CMS sends hospitals a confidential report of what the measure results look like but without publicly reporting them.

And so hospitals received the overall score. They also received the scores

for each of the five specialty cohorts that make up this measure. And they received a list of every one of their patients that was in the measure along with whether that patient had been readmitted and if so to which hospital and what date and what the procedures and the diagnoses were.

And there was a lot of interest in that dry run. CMS sent results to 4,652 hospitals. Seventy-two percent of them downloaded their data. We received 163 questions about the measure after that dry run and we had 2,400 people approximately register for two phone calls that we had to explain the measure and answer questions.

Most of the questions were about the methodology in various ways.

So based on the feedback and the questions that we got in that dry run we made several changes to the measures. And I want to make sure that you're aware of those now.

We updated our planned readmission

algorithm. At that time we added nine procedure categories to the list of things that could qualify you as having a planned readmission. We also removed a diagnosis code from the acute diagnosis list which would have otherwise disqualified a readmission from being called planned.

And those changes were already brought before NQF in the context of the re-endorsement of our other -- various other condition-specific measures.

For this measure, for the hospital-wide measure those changes increased the proportion of readmissions that we called planned. So before we called 5.1 percent of readmissions planned and with the changes we called 8.3 percent of them planned. So that decreased the national unplanned hospital readmission rate from 16.8 to 16.2 percent.

I should add that we subsequently conducted a chart validation of the planned readmission algorithm and have made several

more small changes that have also slightly changed the readmission rate. But those changes are smaller.

Based on the feedback that we got in the dry run we also altered the assignment of some patients from the surgical cohort to other cohorts, so about 200,000 patients overall moved. And we changed the way -- the unplanned readmission following a planned readmission was counted in the measure. Again, that change has been brought before NQF for other measures.

So, with regard to harmonization we did include in our materials an updated memo from both us and NCQA. NCQA is the owner of the Plan All-cause Readmission measure which is another all-cause readmission measure but targeted at the health plan level, not at the hospital level.

And at the time when we originally had the endorsement we were asked to talk about eight different areas in which we were

not harmonized to see if we could harmonize on them.

So, one it turned out we were already harmonized on so that was easy.

Two more NCQA is planning hopefully to harmonize on. But they have put those decisions out to public comment and then it needs to be voted on by their board of directors. So their plan is to harmonize on two other topics which is the planned readmission algorithm and counting readmissions as new index admissions. And so we will hear about that when that vote happens.

For two other areas in which we're not harmonized NCQA did several analyses. Those are the exact form of the risk adjustment variables that we used and using hierarchical versus non-hierarchical modeling.

When they analyzed those two areas they found it made almost no difference at all to the measure. And so for the purposes of

just sort of practicality and simplicity we are both going to continue with the approaches that we take knowing it doesn't really matter very much.

And then there are the remaining issues that we have so far agreed to disagree on still.

One is patients receiving medical treatment for cancer. We exclude those patients from the hospital measure. They have extremely high post-discharge mortality rates and we're worried about the competing risk.

On a plan level that's less relevant to the plan population. NCQA prefers to keep those patients.

The other area that we are going to continue to disagree on is psychiatric patients. We both include patients who are admitted with substance abuse or other sort of medical psychiatric problems.

But patients who are only admitted for a psychiatric disease, like for an acute

schizophrenia or acute bipolar disorder are not in our measure because we -- our measure doesn't have psychiatric hospitals.

And so psychiatric hospitals and psychiatric units are not in our measure. They are in the NCQA, sort of the plan measure.

And so there's a very small fraction, about 3 percent of all psychiatric admissions come into medical hospitals perhaps because we just can't exclude them well. And so we exclude that tiny fraction because they're just a very small fraction of all psychiatric patients. So we're going to continue to disagree on that point.

And lastly with regard to socioeconomic questions it's obviously extremely complicated and difficult.

So after the original committee meeting we put together a variety of analyses trying to understand what the relationship of hospital and patient kind of level variability

was.

And we sent those analyses to the board after the committee had met. But the committee never got a chance to see them. So we included those analyses in the materials we sent to you.

And the 30,000 foot view very quickly is just that hospitals that have higher rates of patients with lower socioeconomic status which in itself is a very hard thing to kind of get your brain around, but we did that in four or five different ways so we could try to capture that.

Those hospitals that have a lot of those patients do have slightly higher risk-standardized readmission rates. Although the overlap is really profound so there's just a huge amount of overlap in the rates.

And then -- but that didn't particularly help us much because that could have been because there's an intrinsic risk to having low socioeconomic status. Or it could

have been because patients with low socioeconomic status may cluster in hospitals with lower quality. And so it's so hard to disentangle what that means.

So, the last thing that we did is we just took all of those patients out as best we could. So, in one analysis we removed all Medicaid patients from the data. So we tried as best we could to take out those socioeconomic status patients altogether. And then we redid the analyses and still we find a slightly higher risk-standardized readmission rate in the hospitals that have a lot of those patients even though we're not putting them in the measure.

And so again, it's hard to exactly know what that means but it's at least suggested that this is not purely a patient-level problem, that there's some component relating to the hospital.

CO-CHAIR HALL: Thank you. Leora, would you mind just reminding us how you --

you touched on it, but you didn't state it in detail, how readmission is considered or not considered an index as well in your algorithm?

DR. HOROWITZ: So, in this measure every readmission is newly considered an index admission. And so a readmission counts as a readmission, and then it also counts as an index and we look to see forward if there's a readmission after it.

CO-CHAIR HALL: Thank you. Any questions from the group? Wes.

DR. FIELDS: Just one small one. I wasn't in the room for this prior discussion. So I'm not clear. Is this an all-plan analysis or analysis of CMS populations?

DR. HOROWITZ: This is endorsed for 18 and over but it's in use currently only for Medicare patients.

DR. FIELDS: So when you said you removed the Medicaid patients, you're talking about the dual eligible population?

DR. HOROWITZ: Correct.

CO-CHAIR HALL: But the removal was just a sensitivity test. It wasn't actually how it's -- yes, okay.

CO-CHAIR KAPLAN: So, I was out of the room as well. Just -- but recalling these data, no matter that it looked like robust across disproportionate share hospitals, whether or not your status was a public hospital, proportion of Medicaid patients, and I forget what the fourth one was.

But robust across about four or five different considerations of things that actually could be a proxy. We can argue about what they're actually a proxy for. But at least for robust across those treatments of potential differences in socioeconomic status. Non-random clustering by patients within hospital. Your findings are reasonably robust across pretty much everything you tried.

DR. HOROWITZ: Yes. As Sherrie said, we defined SES in every way we could

knowing we only had administrative data. And so the things that often people really care about, whether they are literate or are homeless or things like that, we just don't have that data.

But what we did was we looked at proportion of patients that are dual eligible that have Medicaid at the hospital.

We looked at whether the hospital was a safety net hospital which we defined as being more than a standard deviation above the state average for its dual eligible patients.

We looked at whether the hospital was considered a disproportionate share hospital by the government. And we looked at whether the hospital was a public hospital. So, those are all ways we tried to get at whether a hospital was going to see a disproportionate number of low-SES patients.

And of all of those the tightest, the most conservative definition we used was proportion of Medicaid. So we looked at

hospitals that had 30 percent or more of their patients having Medicaid. That was our smallest, most extreme sample. We only had 300-something, 331 hospitals like that. So that's where you're going to see the biggest differences. If we're going to see anything it should be in those hospitals.

CO-CHAIR HALL: Any other delving into any other geographic qualifiers?

DR. HOROWITZ: We did not look at any other geographic differences?

MS. MINTON-FOLTZ: Did you account for no-pay or undocumented?

DR. HOROWITZ: So again, because we did this in Medicare data all of our patients by definition have Medicare. And so all we really had was the dual eligible Medicaid and Medicare patients.

CO-CHAIR HALL: Any other comments or questions? I'll look back to our NQF colleagues. Do we just thank the developers or is there any other?

MR. AMIN: Yes, that's it.

CO-CHAIR HALL: Thank you for this update.

DR. HOROWITZ: Thank you for having us.

MR. AMIN: Okay, so we have two other items. We have public and member comment at 2:30.

I just wanted to point out a few next steps. I'm just going to turn it over to Adeela to talk through some of the next steps.

One of the immediate next steps that I want the committee to be aware of is that we have a call scheduled on May 16 from 2 to 4. And we will use that call, the majority of that call to discuss the harmonization of competing measures discussion.

So we have three sets of competing measures discussions that we're going to have. Just so that there's a few folks in the room that we're going to ask you to play a little

bit of a role during that conversation.

Those that have been lead discussants for the measures, Wes Fields and Pam Roberts, for the two -- there's the 2505, the ED use within 30 days of home health. There are two measures that are related to this measure. And we'll send you a side-by-side table to give you a description of what they look like. But just so you're aware. So again, Wes and Pam, we're going to ask you to lead this discussion during the call on the 16th.

So 2505 relates to 0173 which is the acute hospitalizations for home health patients. And 0171 which is ED use post home health without the 30-day qualifier. And so we're going to have to have a conversation related to how these are related and whether we should be selecting one of the measures for endorsement.

The other set of measures, Helen, we're going to ask you to take the lead on

since you were the lead discussant on is 2375 the SNF all-cause readmission measures and the 2510 also the SNF all-cause readmission.

And then the third set, you know, there are a number of lead discussants. Bruce, Paul, Ross and John, you guys were all part of the discussions around the CABG readmission measures. So we'll send a side-by-side table related to those two measures. But we'll have to have a conversation related to potentially selecting a best in class.

We'll also send along prior to that call a description of the decision logic of how we will go through the discussion around either selecting one as a best in class measure or potential harmonization. But the nature of that call will be to discuss harmonization or selecting a best in class measure.

Again, we'll follow up with much more detail in an email with some descriptions of those measures and who's responsible. But

before we leave today I wanted to make sure that we at least had a little bit of, you know, these recommendations are contingent on the fact that we have a discussion around selecting or at least addressing the question of harmonization or best in class.

I think there are a few questions on that topic so I welcome them.

MS. SHIPPY: Weill we be asked to vote on the phone call?

MR. AMIN: We will -- I don't know the answer to that yet. But likely we won't be -- we'll have to make a decision. So likely it will be through a follow-up SurveyMonkey and not voting on the call itself.

Again, we'll follow up with a lot more detail of exactly what will be kind of expected during that conversation. But it will be more of a lead discussant on the measures that you've already reviewed for the committee. So there shouldn't be anything new

there.

So, Adeela, I'll turn it over to you in terms of summary and follow-up in terms of next steps.

MS. KHAN: Sure. You can look for that email around Friday. You'll get it in your inboxes by then.

Just a quick summary of the measures that we've gone through today. 2515, the Hospital 30-day All-cause Unplanned Risk-standardized Readmission Rate Following Coronary Artery Bypass Graft Surgery passed.

2514 Risk-adjusted Coronary Artery Bypass Graft Readmission Rate also passed.

2393 Pediatric All-Cause Readmission Measure passed.

2414 Pediatric Lower Respiratory Infection Readmission Measure passed.

2513 Hospital 30-day All-cause Risk-standardized Readmission Rate Following Vascular Procedures passed.

0695 Hospital 30-day Risk-

standardized Readmission Rates Following Percutaneous Coronary Intervention passed.

0505 Hospital 30-day All-cause Risk-standardized Readmission Rate Following Acute Myocardial Infarction Hospitalization passed.

And 2539 Facility 7-day Risk-standardized Hospital Visit Rate after Outpatient Colonoscopy also passed.

In terms of next steps after this meeting we're going to have our post-meeting call that Taroon mentioned. It's scheduled for May 16 2 to 4. You should have that on your calendars already. If you don't let me know right away.

After the in-person meeting we're going to start writing the report. And we expect the report to go out to public comment in June, early June, June 6 through July 7. That will be about a 30-day public comment. We have to accommodate for the July 4 holiday. But it will be 30 days.

And we encourage all of you to pass the report around and get as many comments as we can.

We'll have a steering committee call to review and respond to the comments. That's July 30. Again you should have that on your calendars.

In August we expect the measures to go through the NQF member vote and to CSAC followed by endorsement by the board in September.

And we will have a 30-day appeals period we'll start in October. And we'll have exact dates for you once the time is closer.

That's all I have for today. I'll turn it back to Taroon.

MR. AMIN: So, I would just say from the NQF team a profound thank you very much to the committee for all of the hours that you spent reviewing all of these measures, all of the workgroup calls that you spent and obviously this very entertaining but

exhausting two days that I'm sure we've had here.

And a particular thank you to the co-chairs who have led us through this on time across the two days. You've saved us a lot of work in terms of scheduling follow-up conference calls to review measures. So thank you to Bruce and Sherrie for all of your work here.

And we're just -- I'll ask for some reflections from the chairs. But I also want to be cognizant that we have the 2:30 public comment period. So thank you.

CO-CHAIR HALL: Well, I will just briefly say that Sherrie and I are thrilled to preside over such a wonderful group of experts in these areas. So the privilege has been all ours, all mine. I always, always learn from this process so I'm always thrilled to take part. And being with such a great group is what makes it worthwhile.

CO-CHAIR KAPLAN: Ditto.

MR. AMIN: So, Operator, I want to see if there are public comments on the phone. And we'll also take any public comments in the room.

OPERATOR: Okay. If you'd like to make a public comment please press * then the number 1. There are no comments at this time.

MR. AMIN: Are there any comments in the room? No. Okay.

Again, thank you all very much. And we again, we appreciate all of your work on this. Look forward to the follow-up call.

(Whereupon, the foregoing matter went off the record at 2:25 p.m.)

H

N

Q

U

W

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: All Cause Admissions and Readmissions Steering Committee

Before: NQF

Date: 05-06-2014

Place: Washington, D.C.

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