NATIONAL QUALITY FORUM

+ + + + +

PATIENT OUTCOMES ALL-CAUSE
READMISSIONS STEERING COMMITTEE

+ + + + +

MONDAY
DECEMBER 5, 2011

+ + + + +

The Steering Committee met, at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 10:00 a.m., Sherrie Kaplan and Eliot Lazar, Co-Chairs, presiding.

PRESENT:
SHERRIE KAPLAN, Co-Chair
ELIOT LAZAR, Co-Chair
TANYA ALTERAS
BRENT ASPLIN
RICHARD BANKOWITZ
JO ANN BROOKS

PAULA FOLTZ
FRANK GHINASSI
LAURENT GLANCE
JEFFREY GREENWALD
BRUCE HALL
LESLIE KELLY HALL
ASHISH JHA

MICHAEL LANGBERG
PATRICIA McDERMOTT*
DAVID POLAKOFF
BRUCE POMERANZ
MARK SCHUSTER
CHRISTINE TRAVIS

NQF STAFF PRESENT:
TAROON AMIN
HEIDI BOSSLEY
HELEN BURSTIN
JANET CORRIGAN
ALEXIS FORMAN MORGAN
ANN HAMMERSMITH, General Counsel
KAREN JOHNSON
ADEELA KHAN
LAURA MILLER
KAREN PACE

ALSO PRESENT:
DAWN ALAYON, NCQA
ELIZABETH DRYE, Yale University
JENNIFER FAERBERG, AAMC
NANCY FOSTER, American Hospital Association
JEREMY GOTTLICH, NCQA
JEPH HERRIN, Yale University
LEORA HORWITZ, Yale University
RABIA KHAN, CMS
HARLAN KRUMHOLZ, Yale University
KAREN NAKANO, CMS
BOB REHM, NCQA*
MARA RUBIN, UnitedHealthcare
ROBERT SAUNDERS, NCQA
GRAEME SCANDRETT, UnitedHealthcare

RON STETTLER, UnitedHealthcare

*Participating via telephone

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

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

10:13 a.m.

MR. AMIN: Good morning, everybody.

Thank you all for joining us this Monday morning in Washington for, hopefully, a very interesting two days' discussion on readmission measures.

I would like to introduce Helen Burstin for some quick introductions.

DR. BURSTIN: Hi, everybody.

I am Helen Burstin, Senior Vice President of Performance Measures, here at NQF.

I recognize a lot of familiar faces. We did pull on many of you who have been through committees before, understand our process, since this is our first expedited review. We have never done a project quite this fast before or potentially, I was going to say, "or potentially as high-profile," but every time I say that, the next project that

comes along, we say that one is the most high-profile.

(Laughter.)

I think certainly at the moment cost and resource use is consuming much of that energy at NQF, but I think readmissions is going to follow right behind.

I am really thrilled to have you here. This is a bit unusual as well because we also don't usually have the luxury of a day and a half for three measures, but these are quite complex, and we recognize that. We are also very much invested these days in not putting out competitive or duplicative measures that just confuse the field.

So, we are trying very hard, and, hopefully, as a result of your efforts today and primarily tomorrow, to have you actually give us a sense of which measure is best in class and could be used for the broader set of applications. And so, that will be a good exercise tomorrow.

You are in great hands. Taroon has been fabulous, really understands this work. Also, Karen Pace, our lead methodologist, is here with us as well and will offer insights on risk adjustment and other issues.

And with that, I will turn it back over to Taroon.

We actually are waiting for Eliot, but I think we can go ahead and sort of get the usual stuff taken care of while we are waiting for him.

MR. AMIN: Also, I would like to introduce our Chairs, Dr. Sherrie Kaplan and, also, Dr. Eliot Lazar, who is not here yet, but I am sure he will be joining us soon, if you want to have any welcome/introductions to the Committee.

CO-CHAIR KAPLAN: I would just like to also add my welcome to everyone and invite a rich discussion.

Just to lead off, I am a

psychometrician by training, full disclosure. So, I am a measurement scientist. My mother has no idea what I do for a living, but, happily, Helen knows what I do.

(Laughter.)

But I am a measurement scientist, and I am a professor of medicine at UC-Irvine, Assistant Vice Chancellor for Healthcare Measurement and Evaluation at the UCI School of Medicine.

In order to sort of frame this quickly, a couple of housekeeping details. I have chaired a few of these discussions in the past. What would help us all move this thing forward, if everybody could keep their questions and comments concise and focused, we can get a lot richer discussion, and especially responses, if there are any responses from measure developers. If you can keep it focused and concise, we can pedal a lot faster.

So, thank you again and welcome.

MR. AMIN: So, I will turn it over to our Chief General Counsel here, Ann Hammersmith, to do the disclosure of interest.

MS. HAMMERSMITH: Thank you, Taroon. I like that "Chief General Counsel".

MR. AMIN: "Chief," yes. I just added that on.

(Laughter.)

MS. HAMMERSMITH: Good morning, everyone.

If you recall, when we invited nominees for the Committee, you were given a conflict-of-interest form, a disclosure form, to fill out. What we are going to do this morning, in the interest of openness and transparency, is go around the room and have you make any disclosures that you might have.

To give you some guidance about the kind of thing we would be looking for, it is any consulting work that you have done that is relevant to what is before the Committee, grants or research support that you have,

again, relevant to what is before the Committee, and also any speaking engagements that are relevant to the work that will be done here.

We don't expect you to recount your CV, and really we would prefer it if you do not because that will eat up a lot of your meeting.

I want to remind you of a few things before we start. You sit as an individual. Sometimes Committee members will say, "I'm here representing the American Association of...," fill in the blank. You actually are not here representing anyone but yourself. You are here because you are an expert. We want to know what you think, and you don't represent anyone, including your employer, anyone who may have nominated you to sit on this Committee.

The other thing that I just want to give you a quick reminder about is sometimes I hear people say, "I have no

financial conflict of interest." We are interested in a potential financial conflict of interest, but because of the nature of the work we do, there is a possibility that there could be a real or apparent conflict of interest, even where you haven't been compensated for what you are doing. For example, if you sit on a committee with a measure developer that is relevant to what is before the Committee, that is likely uncompensated, but that could be something relevant that we would want to know in connection with your service on this Committee.

So, I am going to start with Dr. Kaplan, and then we can just go around the table.

CO-CHAIR KAPLAN: In the spirit of full disclosure, my colleague and partner, Dr. Sheldon Greenfield, and I have developed a total illness burden index, patient-reported comorbidity, total comorbidity, for outpatient

use with multiple different funding sources in the past. However, we have never used it to date in readmissions for hospitals, and we don't have, I really don't have any experience with all-cause readmission.

In addition, I just want to clarify I have chaired many committees of this nature, but I have never chaired an NQF committee. So, apologies in advance if I make some rooky moves on sitting in this seat.

But that is the probable end of my conflict.

MEMBER KELLY HALL: Leslie Kelly Hall. I have no conflict.

MEMBER SCHUSTER: Mark Schuster.

MS. HAMMERSMITH: Oh, excuse me. I am just going to jump in for a minute.

I neglected to ask you to identify yourselves and who you are with.

MEMBER KELLY HALL: I am Leslie Kelly Hall with HealthWise.

MEMBER SCHUSTER: Mark Schuster

from Children's Hospital, Boston, Harvard Medical School.

I am heading up an AHRQ-funded Center of Excellence for pediatric quality measurement. We have two first assignments, one of which is to create a pediatric readmission measure, and the other, since readmissions keeps coming up in the discussions about it, I will also mention to develop a pediatric hospital CAPS measure, family experience-of-care measure.

MEMBER POLAKOFF: David Polakoff with MassHealth, the Massachusetts Medicaid Agency, and the University of Massachusetts Medical of School, Commonwealth Medicine Division.

I have no conflicts of interest, but I am doing some related work that I will mention. I co-chair the State's Expert Panel on Performance Measurement, where we are in the process of evaluating and, hopefully, selecting soon a readmission measure for

statewide transparency reporting.

And I am also a member of the Medicaid Medical Directors Learning Network, where all of the states' Medicaid agencies are working together on figuring out how to measure readmissions in the Medicaid population.

MEMBER ALTERAS: Tanya Alteras with the National Partnership for Women and Families and Associate Director of the Consumer-Purchaser Disclosure Project.

And I have nothing to disclose.

MEMBER TRAVIS: Christy Travis, Memphis Business Group on Health. I do serve on the NCQA Purchaser Advisory Council, but there is no conflict of interest relative to this particular measure. I also serve as Vice Chair of the Leapfrog group.

MEMBER GHINASSI: Frank Ghinassi, University of Pittsburgh Medical Center, Western Psychiatric. I am on a committee for NCQA on the behavioral health MAP Program.

MEMBER JHA: Ashish Jha from the Harvard School of Public Health and the Boston VA.

And I don't believe I have any direct conflicts, but I will just run through a few things. I have done consulting work for a few different companies, including UpToDate, which is a medical textbook, an electronic medical textbook company; MedCo, which is a pharmacy benefits management company, and I am on the Scientific Advisory Board of a company called Umedica, which does analytics. But none of them really relate to readmissions.

I have written a lot about readmissions. I have gotten grants studying readmissions from Commonwealth-funded. I have a couple pending from the NIH. And as I already said, I am from the Department of Veterans Affairs, which obviously is very interested and involved in a lot of work around readmissions.

Thank you.

MEMBER LANGBERG: Hi. I'm Michael Langberg. I am from Cedars-Sinai in Los Angeles.

I don't think I have any conflicts, but, again, full disclosure, I serve on an AAMW task force looking at variations in various outcomes, including readmissions. We haven't actually done anything yet. So, I am not sure that counts as a conflict.

(Laughter.)

And some people who are direct-reports to me are principal or co-principal investigators in an AHRQ-funded grant along with five other University of California hospitals, looking at outcomes in heart failure, including readmissions. But I don't have that directly.

MEMBER GREENWALD: I'm Jeff Greenwald. I am from Mass General Hospital. I am a hospitalist in internal medicine there.

I have nothing to report in terms

of conflicts. I work with Project BOOST as a clinical investigator, formerly with Project RED, and I am the physician lead for readmissions for Partners HealthCare.

MEMBER ASPLIN: Good morning.

I am Brent Asplin, President of Fairview Medical Group in Minnesota, part of the Fairview Health Services System. I also chair the Quality and Performance Committee for the American College of Emergency Physicians, but do not have any direct conflicts.

MEMBER BROOKS: My name is Jo Ann Brooks. I am the Vice President of System Quality for Indiana University Health in Indianapolis, Indiana.

I have no direct conflicts regarding readmissions. I am a speaker for Cadence Pharmaceuticals as well as a quality consultant, and I serve as Chairperson of the Quality Improvement Committee for the American College of Chest Physicians.

MEMBER FOLTZ: I am Paula Minton Foltz. I am the Assistant Vice President for Education Quality and Patient Safety at Harborview Medical Center, which is part of the UW Health System in Seattle, Washington.

I have no conflicts of interest.

MEMBER POMERANZ: Good morning.

I am Bruce Pomeranz. I am Medical Director at the Kessler Institute for Rehabilitation and Chief Quality Officer for Rehabilitation for Select Medical.

And I have no conflicts to report.

MEMBER GLANCE: Good morning.

My name is Laurent Glance. I am from the University of Rochester. I am a professor of anesthesiology and community and preventative medicine, as well as Vice Chair for Research.

I don't really have any direct conflict of interest. I do have funding from AHRQ looking at performance measurement in trauma and I sit on a number of committees,

including the Committee on Performance and Outcomes Measures for the American Society of Anesthesiologists, as well as serving on some advisory committees, one for trauma, and also the Anesthesia Quality Improvement Panel.

Thank you.

MEMBER BANKOWITZ: Good morning.

I am Richard Bankowitz. I am the Chief Medical Officer of Premier, which is an alliance of about 2500 hospitals.

I will note that Premier has its own proprietary risk-adjustment methodologies, including one for readmissions. And I don't believe that is a conflict of interest.

I will also note, since one of the measure stewards is CMS, that Premier has quite a few applications for contracts pending before CMS for the CMMI, but I don't believe that is a conflict of interest.

MS. PACE: I am Karen Pace on the NQF staff.

CO-CHAIR LAZAR: I must have found

the only slow cab driver between New York and D.C.

(Laughter.)

I am the Chief Medical Officer and Senior Vice President at New York Presbyterian in New York City.

And I don't think I have any conflicts of interest, either, but belong to a number of groups, sit on a variety of committees, probably like many of us in the room, in Greater New York, the American College of Physicians, and so on.

CO-CHAIR KAPLAN: Okay. Let me just add that I forgot something. I am on the same project that Dr. Langberg is on that has yet to produce anything. So, apologies for that.

MS. HAMMERSMITH: Okay. Thank you.

I understand there are some members of the Committee on the phone. Is Patricia McDermott on the phone?

MEMBER McDERMOTT: Hi. Yes, I am.

I work for Aetna. I work in cooperation with in metrics development. We have had an readmissions measure that we have used and developed over the last 18 years within Aetna. It is not anything that is using logic from any other vendor. We have also looked at other vendors' logic, but our logic is its own. And I don't think there is any conflict of interest.

MS. HAMMERSMITH: Okay. Thank you.

Is Mark Williams on the phone?

(No response.)

Okay. I guess he is not on the phone yet.

Is Jim Bellows on the phone?

(No response.)

Okay. Thank you for those disclosures. Do any of you have anything that you want to discuss with each other or any questions of me regarding the disclosures that

have been made?

(No response.)

Okay. Thank you. Have a good meeting.

MR. AMIN: Okay. Thank you.

I just want to quickly introduce our staff here. Adeela Khan will be taking over starting the project introductions.

Clearly, you guys have heard from Alexis Forman Morgan, who has been great in getting materials prepped for our meeting as well.

And I am Taroon Amin.

So, I will turn it over to Adeela.

MS. ADEELA KHAN: Good morning, everyone.

So, we want to know why we are here today. The Affordable Care Act, under Section 3011, directs the Secretary of HHS to develop a National Strategy for Quality Improvement, which is called the NQS. The NQS is used as a guide and includes a strategic

plan on how to increase access to quality affordable healthcare for all Americans.

So, the NQS identified reduction in preventable hospital readmissions as an opportunity for success and uses all-cause readmissions within 30 days of discharge as an illustrative measure for effective care coordination.

And some of the other goals that readmissions relate to with regard to the NQS were effective care coordination, prevention and treatment of leading causes of mortality, and safer care.

So, just a little bit about our project scope. Right now, we are seeking to identify and endorse a cross-cutting, non-condition-specific measure for accountability and quality improvement that specifically addresses all-cause readmissions to hospitals.

And we are also going to be reevaluating under the maintenance process any measures that were endorsed by NQF before June

2009.

So, our meeting objectives today are member introductions and disclosure of interest, which we just finished. We are going to be evaluating the measures to determine if they meet the measure evaluation criteria, make recommendations regarding endorsement as a voluntary consensus standard, vote on the rating for each of the four major criteria and overall on whether to recommend each measure for endorsement.

Discuss related and competing measures to facilitate measure harmonization of related measures. We are probably going to get to discussing related and competing measures tomorrow, I believe. And from among related and competing measures, we are going to select which measure is the best.

MS. FORMAN MORGAN: Okay, and as Adeela said, we will be evaluating the measures based on our NQF criteria. These are our major criteria.

The first one is importance to measure and report. And so, we need to make sure that the measure focus is evidence-based, there is opportunity for improvement in quality, and it demonstrates a high impact where there is variation or overall less-than-optimal performance.

In order for the measure to continue through the review process, it must pass all of its three subcriteria under importance to measure and report in order for it to be evaluated against the remaining criteria.

The second criteria is scientific acceptability of measure properties. This is looking at the measure. Is the measure consistent and credible? Does it present credible results when implemented? So, looking at reliability and validity. And in order for the measure to continue through the evaluation and the criteria process, it must pass the reliability and validity criteria.

The third criteria is usability. Is this measure understandable to those who will use it, to the intended audience? Can they understand it? Can they understand the results of the measure?

Looking at the fourth criteria, usability, this is looking at, is the data available or can it be collected without undue burdens.

Again, as Adeela stated, our next criteria is we do have competing measures, and in order for us to evaluate competing measures, the measures must pass all of the above criteria. So, all of the above criteria must be met in order for the measures to be considered competing, if there are any, and in this case all of the measures are considered competing.

So, today we will mainly focus on the four criteria, making sure that the measures meet all those criteria. And then, tomorrow we will focus on the competing

discussion.

So, our process, it is an endorsement/maintenance process. What that means is we look at measures that are currently in our NQF portfolio and they have been endorsed for three years. And so, now they are going through their maintenance cycle. We evaluate those measures based on the same criteria that we evaluate newly-submitted measures.

As we are evaluating the maintenance measures, we do solicit for new measures in that topic area as well. So, that is what we mean by endorsement/maintenance.

As Helen stated earlier, this is our first expedited review process. And so, in order for a project to have an expedited review process, it must meet these three criteria. The three criteria are the measures must have been tested or currently in use. That looks at our usability criteria. The second, a measure or project must be narrow.

In this case, it is readmissions to the hospital, pretty narrow. And there must be a time-sensitive legislation mandate for the measures.

The role of the Steering Committee, we expect the Steering Committee to evaluate the measures, and we want to thank everyone for submitting their preliminary evaluations.

You will make recommendations to the NQF membership. You will work with the project staff to achieve the goals of the project, and you will respond to comments received during the public and member comments. We have that conference call scheduled.

The Co-Chairs will represent the Steering Committee on the project webinar which occurs at the beginning or during the first part of the member voting period -- and we will go into that in more detail -- as well as the CSAC meeting.

This is just an illustration of our endorsement process, and you can see what is highlighted in yellow/green. This is where we currently are in the process with the Committee reviewing the measures.

And this is our project timeline. So, today we are here to discuss and recommend measures. And then, following this meeting, the project staff will compile all the rationale and all the discussion points and put it into a draft report. We in that draft report will list the voting. We will list everything, all the major discussion points from this meeting, and we will send it to the Committee to review before we post it for member and public comment.

Member and public comment is a 30-calendar-day comment period. Anyone can comment, NQF members, not NQF members. Anyone can comment on the Committee's recommendations and rationale behind the measures that were recommended or not recommended.

Following the close of that period, the Committee will meet via conference call on January 31st, and we will send this information out so everyone can have it. You will via a conference call, for a two-hour conference call, discuss all of the comments that came in regarding your recommendations.

MR. AMIN: So, I will give a brief overview of the discussion format for the Committee deliberations over the next two days.

I will ask if we know where the microphone is. Okay.

After I go through the discussion format, it would be great to have all the measure developers and members of the audience introduce themselves.

So, we will go through each measure individually. The first measure of the day will be Measure No. 1789, the hospital-wide all-cause unplanned readmission measure.

We will ask each of the measure developers, as we move through the measures, to give a brief, three-to-five-minute introduction to the measure, noting that each of the measure developers submitted supplementary materials post the initial submission of the measure. We will ask that the measure developers focus on the updates to focus the Committee.

We will then move to Measure No. 0329, the risk-adjusted 30-day all-cause readmission measure submitted by UHC, and then Measure 1768, the plan all-cause readmission measure.

The first measure, staff will guide the Committee through the subcriteria and the measure submission form to familiarize the Committee members on the evaluation process.

For the second two measures, the NQF staff will not guide the Committee members, the evaluation Committee. We expect

that by the time you will get through the measure, you will get the hang of it, and if there are more process-related questions, feel free to ask as we sort of move through this process.

During the Committee deliberations, we will have the preliminary evaluations, which you should have received in your materials this morning, and also before attending today's session. This will help to guide the discussion on where there are major concerns for the measures that were submitted. Also, the comments that were submitted by Committee members prior to attending here are also listed in these materials.

This is really provided to help focus the discussion. So that, if there are no major considerations for various subcriteria, we can help facilitate moving this day along.

While we only do have three measures for discussion today, we expect each

of the measure discussions to be very robust, particularly around scientific acceptability, potentially even the risk-adjustment models. So, if there is not a major discussion around importance, because of the nature of this call for measures, we can move pretty briskly into the voting procedure.

And I will just say that we will also stop for various public and member comment periods at various points throughout the Committee deliberations, which is listed on the agenda. We have Nicole on the line, who the Chairs will ask to open the lines for public and member comment.

CO-CHAIR LAZAR: Taroon, Sherrie and I thought that it might be helpful for the group for you to give a little bit of a framework in terms of what is congressionally-mandated. In other words, what flexibility does the group have, you know, and so on?

MR. AMIN: Okay. Actually, Helen, if you can help me out with this as well?

DR. BURSTIN: Sure. Again, there are folks in the room, like Nancy Foster, who really know it even better than me, from the American Hospital Association.

There is an expectation that an all-cause readmission measure would be used as part of the hospital readmission program. It is an ACA. We have agreed to have a measure, hopefully, endorsed by April 1.

DR. BOSSLEY: But I think we should emphasize this process is agnostic of whatever is out there. So, if you all -- it hasn't happened yet, but it can happen -- if all of you determine that none of these measures meet the criteria, you can say that none are endorsed.

We do say we are agnostic of what is occurring. Ideally, we would be putting that forward, but it is very important to just remember you truly need to evaluate the measures against the criteria.

MR. AMIN: Can we have

introductions of --

DR. BOSSLEY: I am Heidi Bossley, Vice President of Performance Measures at NQF.

MR. AMIN: And could we have introductions for the members who are in the audience?

MS. RABIA KHAN: I am Rabia Khan from CMS.

MS. NAKANO: My name is Karen Nakano. I am one of the medical officers also from CMS.

MR. GOTTLICH: Jeremy Gottlich from the National Committee for Quality Assurance.

MR. SAUNDERS: Robert Saunders from the National Committee for Quality Assurance.

MS. ALAYON: Dawn Alayon from the National Committee for Quality Assurance.

MS. FOSTER: Nancy Foster from the American Hospital Association.

MS. FAERBERG: Jennifer Faerberg,

Association of American Medical Colleges.

MR. SCANDRETT: Graeme Scandrett, UnitedHealthcare.

MR. STETTLER: Ron Stettler, UnitedHealthcare.

MS. RUBIN: Mara Rubin, UnitedHealthcare.

MS. NEWELL: Alexa Newell, NQF.

MR. HERRIN: Jeph Herrin, Yale University.

MR. AMIN: So, with that, are there any other procedural questions as we begin? I know there is a lot of information presented. I think we will get to know the process as we sort of move along. But if there are any preliminary questions, we are happy to entertain them. If there are any questions, just please raise your placard, as such, and the Chairs will lead the discussion in that fashion.

I guess we will just start. Shall we give a brief introduction from the measure

developers on this measure and any updates that you have?

We will begin discussion of the measure, and as your colleague comes, I think we can have additional discussion at that point.

So, moving on to the first criteria, overall impact, opportunity, and evidence, the importance to measure and report. What this criteria is really aiming to evaluate is the extent to which the measure focus is evidence-based, important in making significant gains in healthcare quality, and improving the health outcomes for a specific high-impact area of healthcare.

This will be evaluated through three subcriteria, high impact, gap in performance, and evidence to support the measure focus.

Specific evaluation considerations for this project that should be kept in mind is that readmissions is considered an outcome

measure, a proxy for an outcome measure. So, evidence will not be required, measure developers will not be required to submit a body of evidence. However, they should submit a rationale that supports the relationship between this health outcome and at least one healthcare structure or process, intervention or service. Evidence, however, would make the submission stronger, but is not required. As I noted, readmission is considered a proxy for healthcare status.

And just one more slide, actually. That one.

So, the importance to measure and report is, again, a must-pass criteria. The two subcriteria that we should consider here is the performance gap, the distribution of performance scores, the number and representativeness of the entities included in the measure performance data, data on disparities, the size of the population at risk, the effectiveness of interventions, and

the likely occurrence of the outcomes, and also the evidence, since the measure focuses on outcome.

Again, the rating of the quality, quantity, and consistency of the body of evidence is not required. However, a rationale that supports this relationship should be presented.

So, I will turn it over to the Chairs to facilitate discussion, if there is any, on this particular measure. And then, we will move to voting once you are ready for that.

CO-CHAIR KAPLAN: I think the first issue is whether or not people have burning issues about the importance of the measure and the performance gap and the evidence given, not the scientific acceptability, but sort of high-impact performance gap and evidence, given the congressional mandate.

So, if this is a burning issue, I

mean, we should have a discussion about this, if people feel strong about some of these issues. But this is clearly from, and the reason it is expedited is, there is this congressional mandate out that has a very tight timeline.

Has anybody got any issues around either -- well, let's start with performance gap.

(No response.)

Hearing none -- I think that is a safe window of silence -- what about evidence? If the measure is a health outcome, then the rating and the quality, this body of evidence is not required by NQF.

Comments? Yes, go ahead.

MEMBER LANGBERG: I have to say I am a little puzzled by the process. I certainly understand there is a congressional mandate for the purpose of Congress, and ultimately CMS, deciding what to do with a measure. So, I certainly understand that

there is a mandate there.

But I am back to the comment that was made earlier that we are supposed to be agnostic and look at the measures themselves. And so, I am not sure what the direction is here for us. Are we supposed to be kind of agnostic when we look at the measures themselves, looking at their capacity for accountability and performance improvement? Or are we supposed to acknowledge that there is a mandate, so we just --

DR. BURSTIN: I don't think there is any reason to say there is a mandate and acknowledge it and just move on. I think, though, importance is threefold, the first of which is, how important overall is the issue of readmissions? I think that may be the issue that may be a bit of a fete accompli in and of itself.

The issue about whether the second -- two parts of importance I think are especially important for you to consider. The

second is, do you believe that there is a significant gap in performance and variation among the entities being measured? That is the second part of it.

And the third is whether you believe there is evidence for the measure focus. This is a little bit more complicated because this is what many would consider an outcome measure. And so, we really want a rationale for the evidence, you know, evidence for the rationale for the evidence.

But having a congressional mandate in and of itself doesn't really answer our criteria. So, that is somewhat separate. I think it was just in the bill; obviously, it is an important general area, but it still doesn't change the fact that we need to fully evaluate and rate each of our criteria.

Brent?

MEMBER ASPLIN: Yes, I think what we are testing here is, can we move off this category relatively quickly? I think the

general answer is yes.

The only area where there was some discrepancy is within the evidence criterion and the subcriteria within that. Is this measure a health outcome? And it looked like most of the Committee said yes, and I said no. So, I don't if now is the time when I have a discussion about that. It is an outcome measure. It just didn't strike me as a health outcome.

DR. BURSTIN: Brent, do you want to --

MEMBER ASPLIN: If you don't want to have this dialog now, that's fine.

DR. BOSSLEY: I actually think it is a good one to have. Actually, Karen was going to give you our thoughts on how we would handle this today.

MS. PACE: Right. I think it is a good question, and it has come up in other projects where we have had readmission measures.

I think the reason we have kind of put it in the overall bucket with health outcomes is that it is really a proxy for deterioration in health status.

So, although it is not really considered -- you know, it could be considered a utilization measure, but we are really looking at it in terms of patients have a reason, you know, there is something that happens that brings them back to the hospital. And so, from that standpoint, it is viewed more in the vein of the health outcome because of the health status relationship.

MEMBER JHA: That sounds like a decision that has been made. I guess the question is, is that a decision we are interested in revisiting or have we decided readmissions are a health outcome? And if we have, then I don't need to sort of revisit that, but I just want to know if that is on the table for a discussion or not.

MS. PACE: I think that is how we

have viewed readmission measures up to this point. So, we certainly could entertain additional discussion about that. You know, it will have implications beyond this project.

So, if we want to be consistent with precedent, then we would consider it a health outcome, but that doesn't mean that we can't have new discussion and a new way of looking at things that would perhaps alter our approach to these.

CO-CHAIR KAPLAN: Does that require that, then, the measure developers go back and provide the evidence of quality, quantity, and consistency of somebody that had been suggesting that this is an outcome measure? Because it does have implications for what was submitted.

MS. PACE: So, I guess, you know, I don't want to just cut it off and say, no, we are not going to revisit that. I guess my question would be, what is it that you are --

MEMBER JHA: So, I am not

convinced it is a health outcome. At the same time, I am not trying to create trouble on it, meaning I know there has been a long process here. This is not like a burning platform that I want to revisit this issue. So, if there is broad consensus that this is a health outcome, I am comfortable with that; I can live with that, and we can move on.

So, I guess I am just trying to test the waters of how much room is there for discussion around that, knowing if you say there is not a lot of room, I am not going to feel like I've been shut down. I got my two cents in. But I personally am not convinced it is a health outcome.

CO-CHAIR LAZAR: Well, I guess I would ask the group, is there strong sentiment one way or the other? In other words, are others having a problem considering this a health outcomes measure? Or are we simply in the sort of shades-of-gray nuance region?

Perhaps this is one to speak up.

I am personally okay with it, although I understand that there could be two sides to it.

MS. PACE: And I think we should characterize it as a proxy for health outcome. Because I agree that people would look at this and say, why is that a health outcome? So, I think it is important that we kind of talk about it.

MEMBER GHINASSI: I am not sure if this is the right time, either, to discuss this. I am learning this process as well. But, as for whether it is a good proxy for health outcome, I think much of that is going to depend on the discussion about the level of adequacy of risk adjustment on these things, the level of adequacy for controlling in that risk adjustment for many of the variables, which some of these things have addressed in varying levels. Like other factors post-hospitalization, they are going to have a major impact on readmission, which may or may

not be connected to what we traditionally think of as health outcome.

SES chaos, a variety of different kinds of issues which post, you know, seven to fourteen days after admission impinge on an individual's health status, which may contribute to readmission, which is in no way directly connected to the outcome of a health intervention. I am not sure where that is being discussed.

MS. PACE: Well, the risk adjustment will be discussed under scientific acceptability and measure properties. In terms of a discussion about adjusting for things that happen after the hospitalization, the way NQF has viewed readmission measures is really an integrative measure and shared accountability measure.

And so, the readmissions measures are really unique, in that they have an element of utilization. They have an element of health status and what happened in the

hospital. And they also have an element related to transitions of care and care coordination and our health system.

That is one reason that they are considered really important measures. It is also a reason that they generate discussion and questions.

But our goal is to have measures that really do move us for better care, better care coordination, better care in the hospital, better transitions. And this is a very integrative measure in that respect.

So, when we talk about the scientific acceptability, risk adjustment, and what factors are included, and how that impacts the validity of the measure, I think that is where we would want to have those discussions.

MEMBER GHINASSI: So, that is going to happen in a different section?

MS. PACE: Yes.

MEMBER GHINASSI: Thank you very

much.

MEMBER ASPLIN: As one of the four rated as not a health outcome, I am perfectly fine moving on. Because I think we could spend two hours on the discussions; it is not going to change my conclusion about the overall importance to measure, which I think is the broader question. And so, I am very comfortable with it. Besides, I don't want to be labeled a troublemaker.

(Laughter.)

CO-CHAIR KAPLAN: We'll make sure that doesn't happen.

Also, in terms of all-cause, remember, the breadth of this is 30-day all-cause readmission. Just sort of remember that, and so as a proxy for everything wrong that could happen, and a suboptimal outcome attributable to a variety of different sources. The decision is made to view this as anything that puts you back in the hospital within 30 days. So, just to kind of keep that

in mind.

Yes?

MEMBER HALL: I would just like to say I agree very strongly with Frank and Brent that these are dilemmas, in the sense that all outcomes are either intermediate or incomplete. Even death is incomplete if your real concern is satisfaction or happiness or something else.

Then, I think we can agree this is at least an intermediate outcome. It is not a process in the sense that we are not asking whether this was applied for the benefit of the patient. So, we are not considering it a process measure in that sense. The process measure is where we incur this obligation to prove the evidence.

So, I would agree with what everyone said, that in that sense we have sort of met this mark. And I do think a lot of the difficult discussions here will be around the scientific acceptability.

CO-CHAIR KAPLAN: So noted.

MR. AMIN: Bruce, do you mind just introducing yourself and if you have any disclosures to the group?

MEMBER HALL: Yes. Sorry. I apologize for being a few minutes late.

I am Bruce Hall. I am at Wash U in St. Louis, and I also am a Director of the NSQIP for the American College of Surgeons in Chicago. At Wash U in St. Louis, I have a corporate position with BJC Healthcare.

So, my only disclosures would be that I have a corporate role for BJC Healthcare and that I am involved with the American College of Surgeons' National Surgical Quality Improvement Program.

MEMBER ALTERAS: Well, it sounds like this conversation is over now, but I just wanted to say, you know, from the patient and the consumer perspective, a readmission is a real health outcome. That is a real concern. It is obviously a proxy, but it is also a

matter of real-life issues that a patient and their family have to deal with. So, we do like to think of these as outcome measures.

CO-CHAIR KAPLAN: Thank you.

So, is there a consensus of opinion that at least in this instance it is acceptable to the group to declare this okay? And can we vote?

MR. AMIN: Yes, on that particular question, there is no need for a vote. But we can just move to vote on the importance to measure and report. Yes, so we can move to a vote on overall importance, if the Committee is ready for that.

CO-CHAIR KAPLAN: Okay. Is there a consensus that we are okay to vote?

MEMBER LANGBERG: Sorry for asking a procedural question again.

CO-CHAIR KAPLAN: That's okay.

MEMBER LANGBERG: This being the first time I have gone through this process and this being the first of three, so we are

going to cut our teeth on this.

If it turns out later on I think that a topic ought to be addressed elsewhere and it turns out it should be brought up here, is the process flexible enough that we visit that or, once we vote on this, it is done? Because I am not sure what is going to be on versus what should be talked about now. It is hard for me to really kind of parse it the way the experts in the room have.

DR. BURSTIN: I think, in general, we will walk through each of the criteria today. You have a lot of opportunity tomorrow, when you do the side-by-side comparisons, to rehash a lot of the discussion, I think.

CO-CHAIR KAPLAN: I think the question is, can he revote if he changes his mind somewhere along the line?

DR. BURSTIN: Not unless the Committee's desire is such that they believe new evidence is emerging and you agree, and

probably would have to take a vote to revote. If you decide to do that, that is certainly within your purview.

CO-CHAIR KAPLAN: I think that answer was probably not -- so, everyone is in agreement we should be voting on this part of the criteria right now?

Do you want to tell them how they vote?

MS. ADEELA KHAN: So, does everyone have a voting clicker? Everyone should have.

MEMBER HALL: I'm sorry, maybe I did miss this earlier, but so how are we going to work? Are we going to walk through all three measures before making final decisions? You mentioned, Helen, there is a side-by-side and all that. When will that happen, and will it be before or after a final decision is made?

DR. BURSTIN: The side-by-side is tomorrow. So, what you will be asked to do

today, if you will vote on each of the four criteria as well as an overall assessment of does the measure meet all criteria for endorsement. You will not make a recommended decision until after your side-by-side repeating measures decision tomorrow.

MS. ADEELA KHAN: Does everyone have a clicker?

MEMBER McDERMOTT: So, this is Patty McDermott on the phone.

We are voting on the first question? Is there a way that I will vote based on being on the phone?

DR. BURSTIN: We will just ask for your vote after the fact.

MS. ADEELA KHAN: Okay. So, the instructions are on the PowerPoint. So, voting won't start until -- there is a timer. We are actually going to do a test vote, so we know that the voting is working properly.

You have 60 seconds to answer the question. You only have to press the

corresponding button once. The last button you press is the one that we will be capturing. So, you can change your answer, if you would like. Whichever one you press last, that is the one that is sustained. You don't need to use the Send key. You can change an answer.

Once all the ratings have been captured, the timer will stop and the results will appear on the screen.

So, just as a test, we wanted to ask you, isn't the weather in Washington, D.C., great today?

(Laughter.)

So, you have 30 seconds. So, if we can have everyone enter their responses?

(Whereupon, a test vote was taken.)

CO-CHAIR KAPLAN: How do you know if the thing worked? Is there some light? Does it light up?

MS. ADEELA KHAN: No. It did work

because we got everyone's responses.

MS. PACE: When you press it, you will get a little green light, and that means it is working. If you would get a red flashing light, you need to let us know. That means the battery is low.

MS. ADEELA KHAN: So, the weather is great today.

(Laughter.)

Well, I guess those from up north think the weather is great here. Okay.

So, we are going to move on to then vote on importance.

So, again, we have the same procedure. Once I start the clock, you will have 60 seconds to enter your vote, and we will start right now.

(Whereupon, a vote was taken.)

We have two people holding out. So, if everyone could just try it one more time?

MR. AMIN: Final vote of 17 to 1.

And the member on the phone?

MEMBER McDERMOTT: I'll go with the majority.

MR. AMIN: Eighteen to 1.

So, what I will do is I will start to move the discussion into the scientific acceptability, reviewing the -- oh, we have the measure developer who joined us. We will offer you five minutes for an introduction to the measure, specifically focusing on areas that were updated post the initial submission.

MS. DRYE: Hi. I'm sorry to be late.

So, I am Elizabeth Drye from the Yale Center for Outcomes Research and Evaluation. We developed this measure for the Center for Medicare and Medicaid Services, as you know.

In this five minutes, I wanted to give you a quick overview -- I think you know the features of the measure -- and review the NQF application and the technical report. But

I wanted to highlight as well the goals that were behind the work that we did that explains why we put the measure together the way that we did.

So, in brief, as you know, this is a measure that divides into patients -- I'm sorry, can I use this thing?

Okay. So, we divided patients into five different patient cohorts: a surgical cohort, a cardiovascular, a cardiorespiratory, neurology, and medical cohort. The medical cohort includes patients not falling into the other areas.

And then, we developed a risk-standardized, a separate model for each of those five patient cohorts that estimates a standardized readmission ratio. We rolled those up into a summary score. Either the individual scores or the summary score could be reported, but our emphasis is on the summary scores, although we tested all the individual five models.

And then, I just point out a particular feature, which is we only count readmissions in the measure that are unplanned. And so, we have a specific algorithm that identifies planned readmissions and doesn't count those in the outcome.

And so, what were we trying to do in developing this? You have already touched on our main goal, which was to develop a measure that reflects quality of care. In addition, we wanted to develop a fair measure that could adequately characterize or purely characterize very diverse U.S. hospitals and, also, a useful measure.

To address the concern which we had many, many long discussions about as well, about how to capture readmission, the quality signal that readmission provides, we specifically shaped both the cohort of patients in the measure as well as the outcome. And let me just give you an example.

So, the measure is very inclusive.

It includes almost all patients admitted to the hospital, but we do carve out patients for which readmission we don't think is a quality signal. We identify, in particular, a group of cancer patients admitted for medical care of their cancer. These patients just looked very different. They seemed different potentially clinically, but lots of patients differ clinically, but they looked different in that they have very, very high mortality rates. When we developed a model for that group and compared it to the other groups of patients, it didn't move with the rest.

We hypothesize, and there is some evidence, that there is a hospital-wide signal of quality captured in a hospital-wide readmission measure, and we should see some relationship across these groups of patients in the quality signal. We, as we report in our application and our technical report, didn't see that for this group of patients. So, they are not in our measure, although many

cancer patients are in the measure; for example, cancer patients admitted for surgical care.

In addition, for the outcome, as I mentioned, we carve out planned readmissions. And here, we were really not trying to, we didn't want to create a measure that discouraged routine patient care just because it happened to be occurring within 30 days of a prior admission.

So, for example, a patient admitted for gastroenteritis three weeks earlier, but they had a gallbladder removal or something scheduled, we didn't want to discourage that. So, we identify those admissions that are for planned care, and we don't count them as readmissions.

Finally, we chose a modeling strategy, hierarchical logistic regression, that really tries to separate what part of the outcome is due to patient factors that we can risk-adjust for, what is due to case-mix

differences, what is due to chance, and what is due to quality. And this is particularly important for small hospitals where chance is a factor.

One thing that is great about a hospital-wide measure is there is a lot of volume. So, the measure functions really well.

But we did think it was really important to characterize the amount of the outcome that is due to random variation and not count that as a quality signal.

And then, in addition, our second principle was fairness. There, we were just trying to make sure that all the hospitals and patients we included we can truly risk-adjust for, and the hospitals, their case mix could be accounted for. And we have pulled out a couple of groups of patients there.

We do not apply the measure to PPS-exempt cancer hospitals because, really, their patients are just fundamentally

different. They don't have a representative mix of patients. In addition, we don't include admissions for patients who leave against medical advice. And there are some patients, like patients admitted for primary psychiatric treatment, that we don't include because they are variably admitted in acute care hospitals; in general, they are not taken care of in acute care hospitals.

And those differences that we felt we would really have a fairer measure, it was just easier to take them out. It is relatively few patients and very few hospitals.

And then, lastly, we really wanted to build a measure that was useful for both public reporting and consumer information, as well as for quality improvement. So, we tried to be very pragmatic in constructing the measure. We just used inpatient data to build the measure. We used one year of data.

We think that the measure's

feature where we have separated patients into five cohorts is really useful because it gives hospital subpopulations that are essentially cared for, often cared for by different clinical care teams. So, they can have more resolution in their measure results, if they want to look at that, or if CMS or others using the measures want to report, want to provide those separate results.

And then, more recently, actually, since we submitted this application, we tested the measure in all-payer data, and like another measure before you, it actually performs even better in this 18-and-over -- when we say "all-payer", we limit it to the 18-and-over population, so that it can be used in a much broader range of patients.

So, to summarize, we really were trying to develop a measure that captures a quality signal associated with readmission that is fair and that is useful, and we are really looking forward to your comments.

CO-CHAIR LAZAR: Any general comments or perhaps some questions for the presenter?

(No response.)

Okay. So, shall we move to the issue of scientific acceptability?

MEMBER JHA: I don't know if this is a reasonable question to ask the presenter, but I guess I was wondering, can you talk about the advantages or disadvantages of using risk-standardized readmission rates versus risk-adjusted?

MS. DRYE: The terms are used pretty interchangeably. We use risk-standardized here because we are using hierarchical logistic regression, which is technically risk-standardizing, but it is very, very much like a more traditional with just at the patient-level we are adjusting for patient factors using logistic regression, just like many other risks. So, that is our first step. We use logistic regression to

adjust for patient factors.

You will hear the word "risk-standardizing" in the other models; it is a very different approach. So, that is a good distinction.

And I meant to note, in this model, which is different than other models that we have put in front of NQF, which have been more condition-specific, we adjust for both patient comorbidities and for the condition for which the patient is admitted, because we are trying to account both for differences in how sick the patients are, but also in what they are coming in for, what kinds of services the hospital provides, since the proportion of those services varies so much across hospitals.

MEMBER JHA: Let me try it a different way. People do use those terms interchangeably. I think that the general -- I guess, let me get more specific.

Can you talk about the advantages

and disadvantages of using the hierarchical model versus a straight logistic regression?

MS. DRYE: So, I have Jeph here, and he is our statistician, here in the back, too. But I will give you a definition or an explanation I think that sets up the juncture between -- it is where most of us work between clinical and statistical considerations.

The patients that we are using to assess quality are clustered within hospitals, and their outcomes are related by the fact that they are being treated at the same hospital. So, that is called the clustering.

The data is, in a sense, hierarchical. We are evaluating hospitals, but we are using patients within those hospitals to assess quality.

And so, one advantage of the hierarchical model is that it accounts for clustering. There are other ways to do that, but that is one of its strengths.

Another advantage is that it

treats smaller hospitals different than large hospitals, in that it weighs the number of cases in calculating the risk-standardized rate. If there are very, very few patients, say five or ten, the model does not make as much of an inference from those patients' outcomes, whether they are readmitted or not, as it does if there are hundreds of patients in that hospital.

So, the uncertainty there, the hospital weights -- and this is different than some other approaches -- an assumption that, if we don't know anything about the hospital, we will assume that this hospital is a typical hospital. And all of the cases that we see will help us pull away from that prior assumption.

So, smaller-volume hospitals in our model tend to have more average estimates, and, also, the model allows us to characterize the uncertainty around the estimate, which is another very important thing. So that, when

we are reporting, we know how confident or how not confident are we in the rate that we are seeing.

If you have any more detailed questions, Jeph Herrin is here and can get into the mechanics of the statistics.

MEMBER JHA: I think we all understand this importance of clustering of patients within hospitals and you have got to account for uncertainty. There are lots of ways of doing it.

Without getting into the details of the modeling, which we might need to do in the scientific discussion, I was trying to get on a broad level what the advantages or disadvantages are that you guys see of using the hierarchical model versus just a straight-out logistic with clustering and showing uncertainty around that.

But I can wait until the scientific discussion.

CO-CHAIR LAZAR: We are right

about there. So, I think a number of us, just from the expressions around the room, would like a little more discussion about this and a little better understanding. So, I don't know whether it is your statistician or how you would like to get to this, but I think we need to really understand it. I'm sorry about that.

MR. HERRIN: So, you are correct, there are many ways to model these kind of data. I don't understand your specific question. I mean, do you want to know -- I mean, if you were to compare our method to some other specific method, we could adjust it for comparison.

But our assumption is that there is some latent signal of quality in hospitals to do a measurement. The approach that we use specifically models that sort of leading factor of quality.

I hope that is helpful.

CO-CHAIR KAPLAN: Jeph, can I

follow that up? In other varying shrinkage estimator-type models, the random effects model without volume is a poor estimate, especially of the lower-performing hospitals. Can you talk a little bit about how the approach -- and I understand hierarchical models that take account for variations in reliability associated with lower volume -- but can you talk about the specific effect of volume on this particular problem?

MR. HERRIN: The specific effect of the hospital's volume on which --

CO-CHAIR KAPLAN: On all-cause readmissions as in random effects in shrinkage estimator-type problems, random effects models.

MR. HERRIN: Well, in any such model, the volume of data, the number of observations you have within a cluster affects how well you know --

CO-CHAIR KAPLAN: Right.

MR. HERRIN: -- for that cluster.

So, certainly, whether or not you are using a random effects model, if you have a small-volume hospital, you know the readmission rate for that hospital with less precision than if you --

CO-CHAIR KAPLAN: Right.

MR. HERRIN: -- had a bigger hospital.

The random effects model acknowledges that, but also incorporates other information you have about readmission rates in general. So, we know that or we can expect that a hospital, without knowing anything else, has an average readmission rate.

CO-CHAIR KAPLAN: Right.

MR. HERRIN: We can combine that information.

CO-CHAIR KAPLAN: Right.

MR. HERRIN: So, I don't know if that answers your question or if I even understand your question.

CO-CHAIR KAPLAN: Well, what I

would like to ask is, more specifically, what is wrong with adding volume to your model as a hospital characteristic, for example?

MR. HERRIN: Oh, what's wrong? I don't know that it is specifically wrong to do that. I think it is a positive question whether you want to --

CO-CHAIR KAPLAN: Yes, why wouldn't you do it?

MR. HERRIN: -- look at that.

MS. DRYE: Just to kind of expand on that, because I think if you haven't been having this discussion recently, it may be hard to follow.

When we are building these models, we are trying to adjust for, we are trying to level the playing field across hospitals. So, we are adjusting for patient comorbidities, and in this case we are adjusting for differences in the kinds of services a hospital provides. Like a surgery has a much lower readmission risk than, say, you are

taking care of very sick heart failure patients or running an ICU. So, we are adjusting for those things.

There are many more things we could put into the model to make the model better predicative of readmission. And one of them would be potentially volume.

But if we did that, small-volume hospitals, this is a big area of demand, and there are small-volume hospitals that have a level worse. And so, we would be saying, well, gee, for you, we are going to make your -- you are going to look worse on this measure, essentially, just because you are small volume. And so, it is a fairness issue.

You could do that. You could separate hospitals out and say, well, we will set a separate prior assumption about how good you are, and we will weigh it. In our numerator, we do a calculation; we don't do a body count or a readmission count.

So, we could weigh a prior

assumption about their quality and that they are worse than an average hospital because they are small. I think you would hear back from small hospitals that that is not fair, but it is probably, if you are a consumer looking on the other end, you might say, well, gee, that may be more accurate.

And so, we are not trying to get the most accurate model. This will come up when we talk about adjusting for -- you know, you could throw in lots of things that make the model predict readmission better. We are just trying to level the playing field. That is why we don't do it. But it is not good or bad; it is just a policy decision.

CO-CHAIR LAZAR: Tanya?

MEMBER ALTERAS: This bleeds into, I think, the usability evaluation that we will get to later. But when it comes to what you were just saying, a consumer -- and there is a smaller hospital. They might look worse. The fact is, if you go into Hospital Compare,

all the hospitals look the same when you are using a hierarchical regression model. You know, they are all in that no different from the national average, which is useless to consumers.

And so, I also don't really know if it is very useful to the hospital, if they are trying to improve quality internally. They do have the actual data.

I know that, in theory, a consumer can drill down on Hospital Compare and see the actual readmission rates. They're not going to do that. You know, it is hard enough just to get them on Hospital Compare in the first place.

So, my real concern, and I think this is an ongoing issue that we are not going to solve today, is trying to get away from this hierarchical risk model, which doesn't really -- I don't know how to balance the fairness versus the providing useful information conflict. But I think we are

tipping very far into fairness. I know you are very far away from providing information that is useful, and we are not supposed to be thinking about the commercial message, but I think we can't really get away from that.

Then, how is this type of modeling going to play into creating useful information to fulfill the requirements? That is a big concern for me when I think about this measure.

MEMBER JHA: So, this is very helpful. I guess I have two kind of broad thoughts on this.

One is whether you choose to account for volume or not in a model is a decision that has real implications. But by choosing not to, the assumption is small hospitals are no worse than large hospitals, which we know from countless studies to be just not true. And so, by shrinking the rest of it to be a national average, basically, what you do is you take all the small

hospitals and make them look about the same.

I understand there are political considerations, but I guess my bias is, if I have to come down on one side or another, this as a quality improvement tool and this as a tool for consumers trumps any unfairness that you might introduce to some small hospitals that actually do pretty well.

And so, we are not going to come up with a perfect choice. But, on the one hand, it seems to be that the choice is give a number to consumers that is the actual number for the hospital. And even if you drill down on Hospital Compare and looked at the actual readmission rate, for a small hospital that is mostly coming from the national average, not the hospital's actual performance.

It also makes it sort of uninteresting for hospitals to focus on improving because they look at their numbers and say, "We're doing fine."

And so, if the goals are quality improvement and consumer engagement, the kind of model strikes me as problematic. If the goal is extra fairness to small hospitals, I think it succeeds quite well.

CO-CHAIR KAPLAN: Let me just sort of focus, refocus this. These are questions for the developer, and we are kind of moving into the group discussion about scientific acceptability.

So, are there any more questions for the developer?

MEMBER LANGBERG: I have a couple of questions not all statistically-related. The first question is whether or not any of the data is risk-adjusted for socioeconomic status.

MS. DRYE: We did not adjust for socioeconomic status. We present data in the application where we show that across hospitals with a different proportion of Medicaid patients there is a wide range of

performance on the measure.

We did look at and think about options, but didn't feel like it was necessary in this case.

MEMBER LANGBERG: I have three questions. Is that okay?

CO-CHAIR KAPLAN: That's okay.

MEMBER LANGBERG: The second question is that, if a patient is admitted to a hospital other than the one that did the original discharge, that will count as a readmission. So, that is a statement. My question is, is there any way in the methodology that you used that that specific information about readmissions to other hospitals would be available to the original hospital, so that they can use that for performance improvement?

MS. DRYE: That is a great question. CMS decides how to report that. To date, they haven't provided that in the publicly-reported, condition-specific

measures, but there is ongoing dialog about it. So, I would just encourage you to keep pushing CMS to give that data. Initially, it is just the barrier to privacy, reporting what happened to that patient and where they ended up.

I know many hospitals have said it would be useful information to have. It doesn't get reported in our rate, but it is potentially information CMS could provide.

MEMBER LANGBERG: My third question has to do with timeliness of the information available to the hospital for performance improvement and to the consumers for their engagement in whatever use they want to have. How soon after a discharge period would the information be available to the hospitals and the public? If it is the next month, let's say, that would be pretty current. If it is a year or two later, then it is pretty useless for performance improvement from the hospital point of view,

and I'm not sure what the consumers would do with it. So, it is somewhere between one month and two years, I'm guessing. Can you answer that?

MS. DRYE: Yes. We built the measure on Medicare fee-for-service claims data, and we built it using one year of data. So, typically, to have really good, quality data -- the claims are in, and CMS has done their quality checks -- you would want the most recent admission to be within a year. So, it would be basically a year to two years from the date of reporting, which is longer than people want. The tradeoff is, if you use more recent claims, the claims are a bit less complete.

So, I would just say that, for this measure, for really having confidence in the claims, it is probably between 12 months and 24 months prior, which is a lag. It would be great to identify ways to provide other kinds of information that show more current

trends. And our group is thinking about that, but it is not part of this measure.

MEMBER LANGBERG: I have a fourth question. So, given that we are dealing with data that is two years old, the focus of our view of the measure is on accountability and performance improvement. Do you have information about the utility of the measure in those settings that spur performance improvement and accountability?

MS. DRYE: I would defer more to measure users and hospitals, but I would just say at the Hayman Hospital these numbers get looked at and they are real and they are acted upon.

It has only been a couple of years that the results have been in place. So, there really hasn't been, for the condition-specific measures, I don't think we have studies showing how quickly people are reacting to them.

CO-CHAIR LAZAR: Okay. We have a

number of questions on the table. We just want to make sure that they are all questions for the developer.

So, Frank, Tanya, Brent, and then Richard.

MEMBER GHINASSI: Thank you.

Just two specific questions. One was about the decision to exclude psychiatric patients. I believe it quotes, the reason was that many or most are treated in, I think you said, rehabilitation facilities or specialized psychiatric hospitals.

I am not sure about the accuracy of that statement. I think if you look at the numbers, you may find that there are a lot of folks treated in single units within acute care hospitals. There are data available on that. I don't have them immediately at my disposal. But I am not sure I agree with that reason for exclusion, No. 1.

And therefore, I have concerns about that, especially since it may also have

implications for how you factor comorbidity of psychiatric disorders into admissions for other medical conditions, which may seriously affect readmission rates.

The second question is about, I believe you had a 12-month continuous eligibility clause built in prior to the index issue. While I understand that, because of claims issues, I am wondering how you factored in the potential differential impact of excluding those individuals, given that there are many factors which could be acuity linked that would account for discontinuous eligibility.

MS. DRYE: On the psych patients, it is a relatively small number of patients coming out of the measure. I am going to have to look it up. I apologize that I don't have that number off the top of my head.

But, really, is it just patients who are admitted for an acute psych diagnosis? There are other psych patients in the measure

who come in for medical problems or who come in for surgeries.

And it was just a small number, really not evenly distributed across hospitals, that was the concern. Like the smallest hospitals would code them in different ways. So, I mean, it would go to a rehab part or a psych part of the hospital and not show up in our data.

So, I think just given the Medicare fee-for-service data, it didn't seem like it would be accounting for it uniformly across hospitals to include them.

I can look for that number for you in a minute, but I don't want to hold up the group.

MEMBER GHINASSI: And I just wanted to add, I don't think the statement that they are usually treated in rehab facilities or specialized psych hospitals is accurate.

MS. DRYE: Right. Let me pull the

information up. Whatever is easiest, I can probably get it for you at the break.

But, again, they are coming in for primary psychiatric care. If they come in for something else, then they are in there, and those are the ones that are usually not at least spending very much time in acute care hospitals, if going there at all.

And I forget your second question. Do you mind?

MEMBER GHINASSI: You chose a 12-month --

MS. DRYE: Oh, right.

MEMBER GHINASSI: -- eligibility clause prior to the indexed admission.

MS. DRYE: Yes.

MEMBER GHINASSI: And I wondered if you had accounted for potential differential in their discontinuous eligibility on acuity.

MS. DRYE: I don't think we looked at that specifically for this measure. We

looked at it in other measures, but I don't think we looked specifically at who was falling out. You know, we had the proportion of patients falling out, but not their diagnosis.

We could go back and look at that, if you are interested.

MEMBER GHINASSI: I am.

CO-CHAIR LAZAR: Okay. Tanya?

MEMBER ALTERAS: Just another quick question. In the surgery/gynecology cluster, does that include obstetrics?

MS. DRYE: No, we did not include obstetrics in the measure because we built the measure on patients 65 and older. But when we tested it on all-payer data, we still didn't build an obstetrics in that group. In all-payer data, we used California's all-payer dataset. But we could. It is just that, given our development data, it really wasn't applicable.

MEMBER ALTERAS: When you say that

you could have something that could happen, you know, is it feasible to include obstetrics data?

MS. DRYE: I am just saying that, as measure developers, we are open to including that in an all-payer-specified measure, but we really haven't had time to explore it at all.

CO-CHAIR LAZAR: Brent?

MEMBER ASPLIN: I wonder if you could speak to the use of the specialty cohorts in general. Just a broad question about that. The rationale cited in the document was that conditions, "cared for by the same team of clinicians are expected to experience similar, added, or reduced low levels of readmission risk".

I understand that the categories from a claims perspective are mutually-exclusive, but my guess is that the care teams, it is going to be much grayer as far as those different categories when you look at

hospitals across the country. And I just wonder if you can speak to that.

MS. DRYE: Yes. When we started, and we spent a lot of time thinking about it looking at data about it, what should we do? One, a model that puts everybody in one cohort or 150 models? You can kind of find examples of everything out there.

And we were wondering whether we needed to break up the cohorts quite a bit to adequately account for the amount of variation and service mix across hospitals. We can risk-adjust, but we just didn't know, you know, like there are specialty surgery hospitals, all these different types of hospitals. Do we really need to break the cohorts up so it would be fairly assessing hospitals?

And actually, we are balancing, also, consideration of sample size. One thing great about the hospital-wide measure is you get a lot of sample size. So, you only need

one year of data, and the shrinkage concerns are not very big. I mean, it is great to have fewer cohorts to get more sample size. So, we could share more with you.

But, basically, when we looked at the statistical issues, and it looked like we could get really good risk adjustment just with our five cohorts. In the end, we felt like we have slightly better model performance if we keep it in five cohorts. Particularly surgery has a much lower readmission rate, and we didn't want to blend that in.

But the arguments weren't mostly statistical in the end. I mean, we could have forced everybody into one cohort. It is just that we felt that the measure would be more useful, even though it doesn't approximate care teams exactly, that's for sure, it would be more useful to hospitals, and it performs slightly better statistically to keep it at a level of five cohorts.

CO-CHAIR LAZAR: Okay. Richard?

MEMBER BANKOWITZ: Yes, I want to ask a question about the decisions regarding socioeconomic status. You presented in your technical paper some data about the percentage of Medicaid patients in the hospital, and it looked like there were pretty substantial differences between the low Medicaid hospitals and those with greater than 30 percent Medicaid. So, given those differences, why did you decide not to make that part of the model?

MS. DRYE: So, this is a really tough issue. I don't think we know how much socioeconomic status, you know, what the sort of cause/effect sequence is between that and readmission rates. And we don't want to adjust away differences in socioeconomic status if it is fair not to, because we want to elevate disparities. We want this measure to highlight disparities across hospitals.

If hospital surveying lowers the SES patients to really lower quality, we want

to highlight that. If we adjust for it, we won't see that difference.

So, I think both consistent with NQF guidelines and with our own philosophy, we don't want to adjust for SES unless we have to because it makes that difference invisible.

And so, then, when we looked at these performance differences for SES, we do see some hospitals with a high proportion of Medicaid patients doing very, very well in the measure. And so, we think it is fair to hold them to a benchmark as a group of sort of an average hospital. And that was our rationale.

MEMBER BANKOWITZ: So, what do you think accounts for that, the difference you demonstrated? You did demonstrate quite a difference? So, what accounts for that? Is it, do you believe that those hospitals are systematically biased against Medicaid patients?

MS. DRYE: I personally think it is a mix of factors, and some of it may be --

you know, the part we don't want to hide is quality, but some of it may be resource availability, both at the hospital and then the patient's own support as they transition from the hospital into the community. But trying to figure out how much is what, it is not really -- I mean, we are people who are doing research; we are thinking about, but we can't pinpoint it.

So, really, again, at this stage we didn't want to hide those differences. I think we feel like, if there are hospitals that are having a hard time doing well on the measure, the policy answer is not to adjust away their differences, but it would be to -- and I am getting outside the measure developer realm -- but it would be to respond to those hospitals with the support they need to do better, not sort of give them a different measure result.

MEMBER BANKOWITZ: Yes, I guess it really depends on the cause of that

difference, which I think at this point is pretty much unknown.

CO-CHAIR LAZAR: We just want to give everybody a little bit of a process check in terms of the time. And also, just remember that the questions for the developer are really on the technical, should be really more on the technical nature. We are starting to edge toward the issue of debate around the measures. And obviously, we have got a specific format for that.

So, we notice a number of hands are up. We would just ask you to please keep those specific to technical questions that the developer needs to answer. We will have plenty of time to express opinion on the measure.

So, I think, Mark, you were next, and then we will just try to close out the rest of them and then move on.

MEMBER SCHUSTER: Thanks.

So, I was wondering, I couldn't

find in there how you picked the five service lines that you focused on. But, then, more importantly, it seemed like you were saying in the materials that the risk-adjustment variables were forced to be identical for all five. And I realize that makes it easier to present to the outside world, but I could see ways in which the risk-adjustment variables might be appropriately different across the five.

And as a part of that, some adjustment variables that sort of acted one way once and one way another time were just dropped, when that may be exactly the kind of variable you want to include if it is different in the way it acts when different service lines.

So, if you could help me understand better the risk-adjustment process, that would be great.

MS. DRYE: Okay. Sure.

So, we used AHRQ's condition

categories to group patients into condition and procedure categories, into condition and procedures. And we started by doing a giant spreadsheet where we looked at how all the risk variables that we typically used -- we used a different grouper to group ICD-9 codes into risk adjusters.

But, anyway, we looked at how those typical variables, when used in other outcome measures, related, just like you are saying. What we saw was a lot of uniformity.

We are trying to balance -- I mean, there were some variables that just looked really, that didn't seem to be heading in a way that we would think made clinical sense at all. And some of them we just took out as potential risk adjusters. We had so many risk adjusters that we had a lot of stronger, consistent risk adjusters to choose from.

And we were trying to balance, again, for usability. It wasn't so much more

for complexity. But if you are going to do this measure, you need to collect data on the risk adjusters, and maybe eventually out of electronic records. We just sort of wanted to get down to some reasonable set.

When we looked at the ones that we ended up with across the five cohorts, surgical, medical, et cetera, they behaved similarly. I mean, sorry, let me back up.

They all were effective as a group in risk-adjusting. Because we fit the models individually, so we take those risk adjusters, but we allow them to go in different directions, depending on what the outcome is, because we fit five different models. So, we used the same comorbidities, but the value of the beta-coefficient and the risk adjustment varies depending on its relationship to the outcome in that cohort.

CO-CHAIR LAZAR: Okay. Jeff, we will make this the last one. I know there are some others that are up, but we are starting

to run tight on time. We suspect a lot of this will come forward during the subsequent discussion.

So, Jeff, if we can make it a quick question, then we will move on after that.

MEMBER GREENWALD: I hope this is quick, at the risk of another slight Pandora's box.

I was very pleased to see an attempt to remove planned readmissions from your measure. That was very nice.

The methodology for it, however, was less transparent to me. And I wondered if you could comment more on the methodology for how you determined and how valid that has been shown to be, because I couldn't find that, in terms of identifying a readmission that was planned?

MS. DRYE: Okay. So, we started with conceptually just that, essentially, procedures and routine chemotherapy and rehab

are planned and almost nothing else is. And so, that was conceptually one leap that we made.

And the other is admissions -- in this case, we are thinking about readmissions -- for any acute diagnosis, like an actual MI or sepsis, those things are unplanned. Then, we used those two considerations and we looked at, given our data, what are people being readmitted for? What procedures are they being readmitted for? And we pulled those out and we looked at the conditions associated with those procedures.

So, this is a multi-step process. It is a little complicated, but we tried to write it out very carefully in our technical report.

So, we took the common procedures -- I think there are about 30 or more -- and then we looked at the acute conditions, and if those acute conditions -- from the acute condition, or, sorry, the diagnosis associated

with the procedures, we decided whether that was an acute or non-acute. Let me just give you an example.

If you are getting a gallbladder removal and you are admitted for sepsis, we are not going to call that planned. But if you are getting a gallbladder removal and your diagnosis is cholestasis or something, that is planned. If you are getting a PCI and you have coronary artery disease, that is planned. But if you have an acute arrhythmia, that may not be planned, if you are having an MI, that is not planned.

So, the framework we came up with, which we try to be very transparent about it, pairs procedures with diagnoses. And the ones that we addressed were the ones that appeared in the data.

I would just add that we did put the measure out for public comment briefly. We got some comments and revised it, and we are actually in a process now of just

consulting one more time with surgeons from specialty societies on these calls, because we want them to be as sort of well-vetted and transparent as possible.

MEMBER GREENWALD: Yes, I think this has a lot of gut validity and makes sense from the description that you provide. I just wonder, have you done any analysis to see, if you eliminate that concept, how that affects the model of planned? Because there is, obviously, some squishiness in these gut valid processes that are sort of judgment-based like that.

MS. DRYE: Maybe I can come back to the Committee with a little more detail on this, if you need it, because I am not sure I can think of analysis that is addressing the squishiness factor per se.

So, we looked at who falls out and what percentage, and do those percentages make sense. But I am not sure that gets totally at your question.

CO-CHAIR LAZAR: Okay. Good. We will move on now to reliability and validity.

Taroon, you will take us through, and then Sherrie is going to make some comments as well.

CO-CHAIR KAPLAN: I will go first because, as a psychometrician, forgive me if I am insulting anybody in the room, but people often confound and confuse reliability and validity.

So, reliability is precision. It is the replication over and over again of this; if you did it a bunch of different ways or if you did it over time with the same measure, would you get, more or less, the same answer?

I love my bathroom scale because my bathroom scale tells me the same, exact answer every day. It is completely wrong, but it tells me the same, exact answer.

(Laughter.)

So, that is reliability.

Validity is accuracy. Are you measuring what you think you are measuring? So, just remember that reproducibility is reliability, and accuracy is validity.

MR. AMIN: That was a great introduction, Sherrie.

I am just going to talk through the various components of reliability and validity that we are asking the Committee to consider.

And procedurally, one of the things that I would recommend to the Committee, this would be the time to sort of focus the discussion internally. If there are further questions of the developers, those are appropriate. But this is really the time for you, as the Committee, to come to some consensus on the information that was presented in the measure. And, really, the evaluation of the measure is on what was presented to you in the measure packets.

So, as we look at reliability and

validity, we are really looking to assess the precise specifications, how precise the specifications are, and the extent of the reliability testing, either at the data-element or the measure-score level. And as we move to voting, where you really would evaluate a measure as high is whether they actually evaluated the data elements and the measure score.

Looking at the validity, we are really looking to see if the specifications are consistent with the evidence, the extent of the validity testing at the data-element level or at the measure-score level -- again, if the developer has actually demonstrated both, that would be rated high -- the justification for the exclusions, the risk adjustment, and whether it is identifying differences in performance.

Some things to keep in mind as we are evaluating the testing is -- on the next slide -- was the appropriate method used to

consider the level, the data source, the conceptual relationships, and the scope of the testing, whether it was accurate -- or adequate? And are the results within the accepted norms?

Actually, if we could go back, that would help sort of get the discussion started.

So, I will turn it back to the Committee Chairs on the preliminary evaluations and anywhere else you would like to take the group.

CO-CHAIR LAZAR: Okay.

MR. AMIN: To the process check, just quickly, we were set to end this session at around 12:00, but feel free to take a little bit more time, if you would like to do that.

CO-CHAIR LAZAR: Okay, let's open the floor for discussion on scientific acceptability.

CO-CHAIR KAPLAN: Why don't we

march through with reliability first and then go to --

MR. AMIN: Yes, and they will be voted on separately.

CO-CHAIR KAPLAN: Yes. So, do you want to start? Let's start with reliability. So, precision, replication, reproducibility.

MEMBER JHA: I think it is excellent. I have no concerns based on any of the stuff that has been submitted that there are any issues with reliability of the data, of the measure.

CO-CHAIR LAZAR: Bruce?

MEMBER HALL: I think the measure as specified is very reproducible; it is impressively so. But I am not sure that that satisfies my concerns about whether we are reliably learning what we think we are learning from this specification. But I think reproducibility is amazingly high.

CO-CHAIR LAZAR: Okay. Does anybody feel differently?

(No response.)

Okay. So, should we move to the topic of validity? Who would like to open the discussion?

MEMBER HALL: Well, I will. This is a major concern. I know, in my mind, validity really derives from the specification, from of the methods.

And this group does impeccable work of the highest quality. As a measure developer, wearing a measure developer hat, I look at every decision and I say that was rational and that is the same decision I would have made.

But, sitting here as a patient advocate, with that hat on, again, my overall, big-picture concern is, are we really learning what we think we are learning here? I have a long list of specific questions that I feel affect the validity of whether we are learning what we are learning.

One is I am still not sure of what

actual time horizon is being specified.

Two, the discussion we just had about whether the planned algorithm is working as proposed is a concern.

Three, technically, the approach to calculating a geometric mean on the individual model results has some advantages and disadvantages to it.

Fourth, the definition of cohorts, again, as a measure developer, I say that is exactly the same decision I would make. But when I look at the cohorts, I see several conditions that are listed as both medical and surgical. Looking at the medical cohort, I see fractures. Looking at the surgical cohort, I see things that I, as a surgeon, don't consider surgical disease.

And I am amazed that the cohorts functioned statistically as well as they did. I mean, as a closet statistician/economist, I am thrilled that they did, but I am amazed that they did. Because when I look at the

condition lists, I can't believe that that fell out, that those advantages fell out.

So, again, to back up to the big picture, I think the work of this group is amazing. I think I would have made all of the same decisions. But, as a patient advocate, I am concerned that several of these issues are not resolved.

CO-CHAIR LAZAR: Okay. Other comments?

MEMBER JHA: So, I guess I made some of my points a bit earlier. But I will just say I want to echo a lot of what Bruce said about the quality of the thinking that went into the decisions I think that the measure developer made.

On almost all of the specific aspects of validity, approaching risk adjustment, whom to exclude, validity testing, I think it is actually pretty terrific, and they have done really an outstanding job.

To me, the fundamental issue

is -- I shouldn't say "the fundamental issue"; that overblows it -- but the place where I get stuck is at the hierarchical model and the way it deals with a large number of hospitals. It essentially gives us a number for that hospital that, in my mind, is not a true representation of what is happening in that hospital.

And so, you can have a small hospital with a couple hundred of patients and the model will predict out that their readmission rate is 12 percent. But much of that data comes from the national average and not what that hospital actually did.

And so, from a validity point of view, I think it raises real concerns. That is the place where I think it leads us awry in terms of both as a quality improvement tool and, much more importantly, as a tool that consumers can actually make sense of.

CO-CHAIR KAPLAN: Can I follow up with you, Ashish? How would you know what the

hospital really did? The problem with validity is, if you don't have criterion validity, you don't have something out there that is the true answer, then you are kind of in trouble. You get construct validity, which that it overlaps with some other characteristic of the hospital like volume or something else. And then, you are confounding; you are in kind of a continuously confounding of the problem. So, how do you know what the hospital actually did?

MEMBER JHA: So, that is a great question, to which I obviously don't have an answer that is easy.

But I guess, if you are stuck trying to make choices around this without a clear gold standard, it seems to me that this is why we have estimates based on the data with confidence intervals. So, you can say for this hospital that was small their readmission rate was 20 percent, and our 95 percent confidence interval is 10 to 40, big,

but that is all the precision the data give you.

And you are now letting the data completely drive. So, that really is what the hospital did. There really was a 20 percent readmission rate.

Based on the risk-standardized hierarchical model, as I understand it, a hospital that gets a number of 20 percent didn't actually have 2 out of every 10 patients get readmitted. That is what the model predicts, based on the choices that have gone into the modeling.

And so, if you are stuck on this issue, my inclination is just be as transparent as possible with the data, and be as circumspect as possible about the precision of that.

CO-CHAIR KAPLAN: Okay. I think we just, again, are moving precision and validity back and forth over some kind of transom. So, you want to make sure -- I

understand heterogeneity and confidence intervals, but, again, validity is about something that we have to demonstrate that is related to something else that is truly related or some criterion that we know for sure measures the overall hospital quality with respect to readmissions.

So, I have one question to the group. How many of you have actually done risk specification modeling?

(Show of hands.)

Okay. How many of you have no clue what we have been talking about for a lot of these shrinkage estimators and risk specification or at least they know the principles, but truly how many of you would like a little more English version of this?

(Show of hands.)

Okay. Brave souls.

So, I would just like to ask us to now, are there any other validity issues that you all have concerns about? Because risk

stratification, risk adjustment, and that sort of thing strikes me as something that came up in the last round of discussions a lot. Are there any other burning questions of the other six criteria other than risk specification that people are concerned about?

MEMBER LANGBERG: Since I raised my hand as a brave soul, I am not sure this applies here. But in looking at the submission, under validity, the submission included, I think, three assumed-to-be-related measures of quality or performance that validate, I think, or attempt to validate, this question we have just been discussing. There are the HCAPS scores, the Thomson Reuters top 100 hospitals, and the Joint Commission's top performers, and key quality measures programs. Those are offered, I think, as validity of the outcome of the work if it is connected to those in some way.

I am just a little mystified myself since some of these things are black

boxes in terms of how they are created. And there are hospitals, I would expect, not to be on there that are on there, hospitals that are not on there that should be on there. And it was mystifying to me that those were identified as the standard by which the validity was being compared, if I understood that correct.

CO-CHAIR KAPLAN: Is that a question? Well, what else would you have offered?

MEMBER LANGBERG: I think that is a really good question. I am sure we can all get that question. Virtually, everything is being asked. I haven't thought about that because I didn't get at the validity of the whole metric anyway. And that is not my expertise.

So, I wish I could offer a constructive comparison. I don't know that those three are. That I can't offer a suggestion by itself does not validate or

invalidate the observation.

CO-CHAIR KAPLAN: Sure. Absolutely.

Does anybody else have some criterion that they could propose as possible validity variables, other than those in HCAPs and Reuters and the other one?

MEMBER HALL: I mean, there is a quandary here because what is being put forward is better than anything we have. So, to compare ourselves to other things that we don't think can do what we are proposing to do as well as we do, it is, you know, we are chasing tails. So, I don't know that there is an answer. I don't know that there is a resolution to trying to find some gold standard when there isn't one.

MS. PACE: I will just offer a comment in regards to our criteria about testing and validity testing. Actually, this is more validity testing than we often see with measures submitted to NQF because we

still allow some face validity as a consideration for NQF measures.

But, in general, validity testing, there is usually not one definitive validity test. Validity testing is usually very conceptual and is built over time, and you look for consistency of results.

And so, ideally, you would like to have a gold standard that you could say, yes, we have another valid measure of quality that we know is a measure of quality, and we can compare this new measure to it. We just don't typically have those in this field of quality performance measures.

So, I think part of this is you, as experts in the area, to understand these conceptual relationships, and I think it is definitely a good question that we don't necessarily know how those other measures were built. But I think it is looking at this not as there's going to be any one definitive answer for the validity.

And, Sherrie, you may want to add to that.

MEMBER GHINASSI: So, I want to be clear. For this part of the discussion, were you excluding as part of the validity the risk adjustment or is that still on the table?

CO-CHAIR KAPLAN: No, I was proposing that we come back, that we focus on risk adjustment, but if anybody has any other issues in the one, two, three, five, and six categories, we should discuss those as well, and finish off with maybe a little bit, if possible, a little bit longer discussion, a more robust discussion of risk adjustment.

MEMBER ALTERAS: I just wanted to ask if, when we are talking about risk adjustment, we group it with the identification differences in performance. I think in this case they are very closely-linked.

CO-CHAIR KAPLAN: Right. That is a good suggestion.

So, does anybody else have issues with the validity testing, and whatever?

MEMBER KELLY HALL: I just had a question with regard to justification of the exclusions. Does that relate to any time of a patient preference or is that included as well? For instance, advance directives where all extraordinary measures would be taken, and the patient might be readmitted based on that preference from a palliative care setting or another setting?

CO-CHAIR LAZAR: Any other comments? Mark?

MEMBER SCHUSTER: Yes, I mean, just because the CAPS example was brought up, when I read that, I didn't know how to interpret it because a lot of those people I think would have been filling out the CAPS survey after the readmission, based on the 30-day readmission time point and the range of time when CAPS surveys were filled out.

So, it seems very predictable

that, if you are filling out a survey on an admission and, then, in the interim you have had yet another admission, your experience ratings are going to go down. So, I wasn't sure what methodology was used and if you excluded people who filled out CAPS surveys after that readmission, but it seemed like almost an intentional setup to get the desired result.

MEMBER HALL: I guess, with respect to your question about what other issues are on the table, I do think I would like some clarity on the time horizon, the implications of using a geometric mean calculation with potential differential variance of each of the components, and some clarity about the definition of cohorts, because many of the conditions seem to be represented repeatedly.

CO-CHAIR LAZAR: Okay. Other comments on either reliability or validity?

CO-CHAIR KAPLAN: Apart from risk-

based.

CO-CHAIR LAZAR: Right, other than risk?

(No response.)

Okay. Taroon, hearing nothing further, is this the time to vote on scientific acceptability?

MR. AMIN: So, what we will do, actually, because the further discussion around risk adjustment is yet to be had --

CO-CHAIR LAZAR: Okay.

MR. AMIN: So, what we could do is, if the Chairs are okay with this, we can go to public and member comment, then break, and come back at 12:30 for the risk-adjustment discussion, and finish up with usability and feasibility before the lunch break.

CO-CHAIR LAZAR: Okay. Everybody agree? Okay. Terrific.

MR. AMIN: So, Nicole, if you can open up the lines for any public or member comments?

OPERATOR: Certainly. That is *1 for any public comment at this time.

(No response.)

MR. AMIN: Any comments from members?

(No response.)

CO-CHAIR LAZAR: Comments from anybody else in the room?

(No response.)

CO-CHAIR KAPLAN: Okay. So, we have a 15-minute break, and then we are going to come back and discuss risk adjustment and then usefulness and feasibility.

(Whereupon, the above-entitled matter went off the record at 12:01 p.m. and resumed at 12:16 p.m.)

MEMBER GHINASSI: Yes, just a couple and I'll be brief. I didn't see it. It might be there. But some of the issues that I think might effect the level playing field because I think at the end of the day, that's probably the most --

CO-CHAIR KAPLAN: Right.

MEMBER GHINASSI: -- one of the most important variables here because if quality of care is going to be improved, it has to be not only timely but it has to be valid in that sense.

But I don't see any way to, at this point, account for differences in locally, regionally, by state, or nationally, in differential ability of that region to provide timely and quality follow-up care. So access issues following inpatient care, how long the wait is for the next level of treatment, whether or not that's a regional variable that is differentially impacting people, I think it's a fact. I don't see that built into the model at all.

So if you're looking at rural North Dakota how does that compare to inner city D.C.? How does that compare to Memphis? And since they're not using ambulatory care claims, I don't see any way to adjust for

that.

Second, I don't see -- it's been raised by several but I don't see any adjustments currently for patient variables that are not diagnostic codes. So issues around homelessness, domestic violence, issues that are going to effect the ability to follow through, childcare issues that are going to effect transfers from people getting from home to an ambulatory care treatment when they've got two kids to worry about and have to do a bus transfer.

These are intangibles but I've been in hospital systems for almost 30 years and know that these drive ability for people to follow up for care. And not followed up care often effects inpatient.

The other thing I'm not sure how they're adjusting for was specified in one of the other measures that we'll get to, I didn't see it in this one -- transfers between hospitals occur fairly regularly. It is rare

that you see a transfer go from a -- how can I put this -- from a hospital that's likely to be dealing with more high acuity patients to hospitals where there was more acuity.

It appears, from my experience, it is always in the other direction. That lower acuity, community-based sometimes facilities, will transfer cases mid-treatment that are going wrong to facilities that are more known for specialty care.

And yet I believe the readmission model only accounts for the receiving hospitals' performance, not the sendings'. And I don't see that built in anywhere.

I'll stop at that.

CO-CHAIR KAPLAN: So I would actually like Elizabeth and then Karen to talk about adjusting a way versus, you know, making a level playing field -- adjusting the way the variants you're trying to explain in terms of some of these issues like, for example, the safety net hospitals are for tertiary care

facilities versus assuming that what you get is what you treat. And you should be doing a good job.

MS. PACE: I'll talk a little bit about, in general, the NQF criteria regarding risk factors. And we really do not recommend that factors that are generally associated with disparities in care, such as race, ethnicity, and socioeconomic status or Medicaid as a marker for socioeconomic status, be included in risk models. That's our general approach.

If they are included, they have to be strongly supported in terms of analysis to be included. And the general philosophy about that is that if you include them in the risk model, you're kind of accepting the assumption that outcomes should be different based on those factors.

And so what we're trying to do with risk adjustment is to account for variability in the patients' conditions as

they present to the provider that's giving care, whether it's a hospital or physician, or health plan.

Some of the other things you mentioned about homelessness and, you know, you could add patient education and those kinds of things, certainly they would be fair game for risk models. The problem that we often run into is the balancing the burden of data collection with risk models.

And so a lot of times we don't have that information in a systematic way available for these various types of risk models. So that's always kind of, you know, a waiting thing in terms of balancing what data you have with what you are able to do with risk models.

The other thing that I'll just mention that Elizabeth talked about earlier is that when you are doing risk adjustment or risk models, you're not trying to -- it's not like a model where you are trying to identify

every factor that influences the outcome because we're actually trying to separate out the things that the patient comes with hopefully then differences in quality of care.

So, you know, we could include a lot of other things in the model, which you would have a better predictive model. But that's not what we're doing with risk adjustment in these various risk models.

So, you know, risk adjustment is not a perfect science. I don't know that it ever will be. But, you know, these are all things that need to be weighed. And it's not just statistical questions. A lot of it is, you know, conceptual relationships and what we're trying to make distinctions about.

In terms of the question about area resources, I'll make a comment. And then you guys certainly chime in and make any other comments here. But so if you're -- you were mentioning resources in different geographic areas, if people are going to be, you know,

looking at Hospital Compare, for example, most of the times they are going to be doing a comparison of hospitals within their geographic areas.

So even though you may have comparison to a national standard or a state average, in terms of using this, most of the time it would be done on a geographic area. So some of that will play in a geographic area.

But the other this is about -- as I mentioned earlier about these readmission measures, they really are seen as driving more coordination and collaboration of care. And so they are really purposefully not trying to isolate only the hospital because part of what they are trying to do is drive hospitals to help with transitions.

And if those things aren't at play in a particular community, then those kinds of collaborations need to be developed in order to really make an impact on readmission. So,

you know, there's a lot of different things that play here in terms of what the measures are trying to drive and the realities of data and what we're trying to do with risk adjustment.

But that's probably more than you wanted. But I'll stop and see if you guys want to add anything else.

CO-CHAIR KAPLAN: Let me just interject and ask Elizabeth if there's -- I mean it is my understanding that you are adjusting for disparity, the index condition, and also for comorbidities. So to the extent that those absorb a lot of this other variability, they're already in your risk model.

Is there anything else you want to add to that?

MS. HORWITZ: No, that's actually what we would have said. They are precisely the reasons that we elected not to adjust for either of those conditions.

MEMBER GHINASSI: Just one follow-up statement. I just want to reiterate that I believe that while comorbidity is essential and that an adjustment for that is clearly a great idea, I think that the amount of variance accounted for by factors that are not included in the comorbidity lists may drive far more of the variance that we'd like to acknowledge in this case.

And to leave it off the table because it is a burden is tempting. And as a provider, trust me, I understand about the burdens of new mandates for measurement, at the same time, I just want to caution all of us to remember that leaving certain things off the table leave us with an easier-to-measure product of questionable validity, much like your bathroom scale.

MS. PACE: I totally agree. And remember that every variable you include in the risk adjustment level also has to be subjected to reliability and validity evidence

in order to make the case for including it. And a lot of times those variables have really incomplete evidence to put it generously.

MEMBER BANKOWITZ: I wanted to agree with Frank on this issue of the fact that complex suicidal factors can really impact the readmission rate. And I agree with the aspiration to have equity. I also agree we need to push hospitals to have better coordinated care. I don't disagree with that.

But I don't think it is quite as simple as saying well, we aspire to equity so therefore we won't look at socioeconomic status because we know that factors like healthcare literacy, access to care, the complicated lives of the dual-eligibles, also even things like access to healthy food and access to healthy playgrounds, these are all important.

And I don't think we can simply say well they are important. They might, in fact, influence the readmission rate but we

were not going to look at them because we want to highlight disparities.

This is not a measure of disparity. This is a measure of quality. At least that's what we said at the beginning. And the stakes are quite high here. We are going to direct consumers, based upon public reporting, and we're going to redirect resources, based on payment, to these hospitals.

And I don't think we should start off with the assumption that well, there are differences between the hospitals who have high Medicaid populations. It must be a quality issue.

No, I think we should say look, there are complex issues at play. If we're going redirect resources, we'd better get it right. Why do we want to withhold resources from those hospitals that may need the most?

So I think we need to pay some attention to this assumption about being sort

of blind with regard to socioeconomic status in this particular issue, which is a very complicated social issue as well as a health issue.

MEMBER GLANCE: So I generally agree with all of the comments that are being right now. It's not at all clear to me how much of readmission is being driven by admission diagnoses versus comorbidities versus some of the SES factors that we're talking about.

And although I generally agree with Karen's comment that in many of the risk-adjusted quality metrics that we use, SES should not be incorporated into those models. For example, for mortality models, you know, we can collect some pretty good data on comorbidities. And comorbidities and admission diagnosis and quality is what drives mortality.

Now for readmission, it is a little bit different. For readmission, yes,

comorbidity and submission diagnoses drive readmission. But also SES, what happens in the community access things.

And what I haven't really seen here is a discussion of the relative importance of those various factors driving readmission. And if, in fact, access and SES are incredibly important factors, maybe they do need to be included in these models despite the data collection burden that we're discussing.

MEMBER FOLTZ: I'm going to join the bandwagon here. Harborview is a safety net hospital and a tertiary center. And one thing, by excluding the socioeconomic factors, I think we may see unintended consequence of cherry picking.

I mean we've certainly seen it in the past where hospitals refer patients or transfer them to public hospitals or tertiary hospitals. And they're not the commercial payers I'll tell you that.

CO-CHAIR LAZAR: (Speaking from an unmiked location) Let me just see is we can sum this discussion of SES up. And I think, Karen -- with you before we have further consideration.

Is there anybody who feels -- oh, I'm sorry -- is there anybody who feels strongly that SES should not be included in a proposed all-costs readmission measure?

Ashish?

MEMBER JHA: So we've done a lot of empirical work. My research group has done a lot of empirical work on this. And the empirical work is overwhelming that safety net hospitals have much higher readmission rates, that African Americans have higher readmission rates than whites, and that white patients who are in predominantly minority-serving institutions have high readmission rates than black patients in non-minority serving hospitals.

So that seems to be a funny

prologue to saying why we should not include this. But let me kind of make the point --

PARTICIPANT: I'm on the edge of my seat.

MEMBER JHA: -- or I'm about to give you something completely illogical. But it sort of goes back to whether you think this is a good quality measure or not.

If you fundamentally believe that readmissions are a good way to measure quality, they're a good way to measure an important health outcome, then I think there is no way we can justify including SES because basically what you're saying is it is okay you're lousy. You just take care of poor people. And that's not, I think, where we want to go.

I think the tension here and the reason why a lot of us are struggling with this is because we all believe -- I shouldn't say this -- some of us might believe that at the end of the day this is not a great measure

of health outcome or quality for the hospital.

And that factors outside of the hospitals' control are driving so much of why people get readmitted. And the socioeconomic status of your patient population is such a major driver that it sort of fundamentally gets back to do you really believe this is an important quality measure.

But I don't want to re-litigate that because we've already concluded and voted, I think, 18 to 1 or 19 to 1 that it is. So I think once we've made that decision, it's hard, in my mind, to justify then including socioeconomic status if at all.

MS. DRYE: Hi. I just wanted to make one more technical point that -- on Karen's construct, which is the alternative. There's an alternative.

I'm not recommending it at all because we do see a whole range of performance across hospitals with a high proportion of low SES patients. But an alternative to risk

adjusting is to stratify your population into hospitals with a lot of low SES and hospitals that don't have a lot of low SES patients, for example. There's no clear-out put.

Imagine that you could make one. So then you really would exclusively be holding those hospitals with more low SES patients to a different standard potentially. And -- but you wouldn't be burying their difference in a risk adjustment variable.

And I just wanted to say that is the NQF guidance, that if there's really differences for whatever policy, fairness, philosophical reasons, you note them and you don't think that it is fair to report without adjusting them, it's not that you adjust -- you would stratify the measure.

CO-CHAIR LAZAR: Okay. We have a number of other comments. So we're back to front.

Brent?

MEMBER ASPLIN: Well, I just would

ask -- didn't you answer you own question though about why there would be an exception because I think you laid it out very nicely. And why you would normally not want to include this.

But I thought you answered -- you gave the justification for why we would include it this time because it is -- we're saying yes, it is a quality measure. It is a quality measure of the system that happens to be focused on one aspect of the system, which is the hospital.

And there's the crux of the matter. And that's why we would normally not include. There is a reason why we need to include it this time.

So I don't -- I mean I don't really have anything else to say beyond what I've said. As you can, I've struggled with this. And I continue to struggle with this. I guess sort of just for intellectual consistency, I sort of believe that if we

decide this is really a quality measure for hospitals, then I fall back to we should not include SES.

But -- and I've already sort of said how I feel about that in general. So -- but I clearly am not very clear in my own head about having this.

CO-CHAIR LAZAR: Okay.

David?

MEMBER POLAKOFF: I know we, with the introduction, simply represent ourselves, not organizations. But I am the Chief Medical Executive of a Medicaid Agency. So the population we're debating right now are sort of our exclusive concern.

And I, too, like Ashish, I have concerns about including SES for the same reasons. I'm really concerned that it could adjust away very real differences that we would like to see.

I'd also like to just add one little bit of empiric evidence to the

discussion, which is that we recently went through an exercise in Massachusetts Medicaid of evaluating all the hospitals that we contract with, that we pay in the state. There about 70 hospitals with one readmission measure. And then are rating them by rates.

And interestingly, it didn't track very well with their level of participation in Medicaid at all, with their volume of Medicaid patients. There were high Medicaid hospitals that had low readmission rates. And vice versa. There is a bit of trend but there wasn't -- there really was not a consistent trend at all.

So I'm concerned about losing differences that would be very, very meaningful.

CO-CHAIR LAZAR: Okay.

Tanya?

MEMBER ALTERAS: I agree. So that was a great segue for that. I think the -- going more towards stratification is the way

I would feel much more comfortable.

And I just wanted to make a comment. I'm kind of laughing inside even though this isn't funny at all. But three years ago I was on the NQF Hospital Outcomes and Efficiencies Steering Committee. And we had many of the same conversations about SES and also about how much of the readmission is really due to the hospital care versus what happens after the hospital -- after the discharge.

And, you know, I think that -- I think, Frank, you said, you know, there are a lot of issues about care coordination, what happens after the hospital discharge, you know all the statistics we know about, you know, the high percentage of patients who are readmitted having no family or other care giver support locally, you know, and how that plays into a readmission.

But that doesn't take away the hospitals' role in this. So I just wanted to

bring that back up to the surface.

I know we all know that since we're all here. But I just -- and, you know, in an ideal world, we would have a measure that included lots of ambulatory care data, data on care coordination. We're not in the world yet. And so, you know, we do want to focus on the readmission and what the hospital, you know, what the hospital data says about that.

CO-CHAIR LAZAR: So just to sort of, you know, bring everybody back to the task at hand, our charge for this particular section is to simply vote on the scientific acceptability of, you know, the measure the processes, vis-a-vis reliability, validity. And obviously risk adjustment being a part of it.

I just want to make sure that we don't, you know, try and create the perfect measure. We're really here to simply make some kind of a, you know, recommendation. And

obviously, you know, some kind of a assessment of the particular characteristics of this measure.

So I know there are a lot of cards up. I anybody has a comment that is critically important to making that decision, please keep it up. If not, I'd like -- I think Sherrie has one last comment. And then we need to vote on this section and then move on if that's okay.

It sounds like a couple of cards are still up so --

CO-CHAIR KAPLAN: Well, wait a minute, if you don't mind, let me just interject something here, which may or may not help.

But a recent, a very recent systematic review of risk stratification models by Ken Sagar et al., and I forget what the Journal citation is, say most of our risk stratification for readmissions don't work. Mortality, these risk stratification models

work. There are some good ones out there, different approaches that work.

But we are nowhere near scientifically sound risk stratification models. And if you think SES is an easy -- we are all glibly tossing it around, you should sit in on some of social scientist-kind of meetings about socioeconomic status and how the complexities in measuring that.

So in the database, it is not consistently agreed that there any variables remotely touch on the issues they raise, including race and ethnicity.

So just to kind of give you a little guidance on is there something in everybody hip pocket, absolutely not. There's not anything out there right now that is the appropriate approach to risk stratification.

CO-CHAIR LAZAR: Okay. So critical comments now that are going to be essential for us to cast a vote on the scientific acceptability. If you don't have

one, please drop the card. If you do have one, then we will get around to you.

Laurent?

MEMBER GLANCE: So I think as we were talking about the validity of the risk assessment, because I think it is the question at hand, one of the things that I did not hear discussed at all were statistical measures of performance.

So we've look at the -- the measure developers do talk about discrimination and give us a little bit of information on that. If I recall correctly, I think a C statistic was in a .65 range, which seems to be pretty good for this kind of a model, not so good when you're talking about risk adjustment models in general. And that doesn't really add a lot of information to this discussion.

But the point that I do want to ask about is model calibration because that really is critically important when you are

going to use risk adjustment models for bench marking. And I don't recall seeing anything about model calibration in the technical manual. And I was wondering if we could find out about that before we vote on this.

CO-CHAIR KAPLAN: Elizabeth or Jeph or anybody? Is it in the technical report?

MS. DRYE: It is in the technical report.

CO-CHAIR KAPLAN: Can you tell us a page number?

MS. DRYE: It's 53 of the technical report. That's the model specifications, including calibration and discrimination for each of the five submodel group measurement. That goes through to page 55 -- 52 to 55.

And did you want us to speak to those results?

CO-CHAIR LAZAR: Briefly.

MS. DRYE: Can you guys bring it

up on the screen? So do you want to --

MR. HERRIN: So there, you know, you've got the technical report there. A lot of results here because we had five models. But the full results -- we do have the calibration results here. They are -- typically one calibration would be the two calibration numbers would be close to zero, close to one.

And pretty much across the board we found that we couldn't find any central calibration results for any of our models. You know it is a lot of detail. I don't want to go through all the numbers.

But I think if someone wants to look through these and has specific questions about how I calibrated --

CO-CHAIR KAPLAN: So I think one of the things that we often ask as far as like risk deciles, can you just describe a little bit more -- I'm not sure what the calibration numbers in here, the zero, one -- I'm not sure

what --

MS. HORWITZ: So we provide a little detail on the bottom where the gamma zero we expect to be as close to zero as possible and one -- as close to one as possible. And it is set that way for the development sample. And then we compare the validation -- the two validation samples to them. And as you can see, that's very close.

The risk deciles we present in the next row here, so we're looking to see what is the predicted risk for patients at the lowest decile of risk versus what's the predicted risk for patients with the highest decile of risk.

And you like to see, again, as wide a spread as possible because you want to show that your model is able to discriminate levels of risk.

And so as you can see, these range from a very low risk of readmission of 9 percent in the lowest decile to the highest

risk or readmission being around 33, or 35 percent in the highest decile. And that's pretty consistent across all of our measures four to 27, five to 31. So it's on that range. And that's a pretty reasonable discrimination of risk for our models.

CO-CHAIR LAZAR: Good. Can we move on?

Okay, Brian, do you have a comment?

MEMBER ASPLIN: Real quickly.

CO-CHAIR LAZAR: Four more and then we're going to move on.

MEMBER ASPLIN: Just need to understand better the validity threat that was mentioned around how volume is treated in low-volume hospitals in this model. So I conceptually get that basically if you have a very low end, it's the national average that's somehow -- almost -- not necessarily being substituted. But if somebody could just explain that link a little bit better, that

would help me.

CO-CHAIR KAPLAN: Jeph, you want to respond to that for your measure?

MR. HERRIN: Yes. First of all I'd like to point that since this is all cause readmission, we don't have any hospitals that have very small volumes. Certainly for the condition-specific measures, we do often have hospitals that are publicly reported that have 25 admissions. In our case, we don't have -- every hospital has at least several hundred. So we're not talking about the same kinds of small volume.

The idea is that if you don't -- there's always a tradeoff when you're trying to measure things. There's always lots of tradeoffs. And in this case, the tradeoff, I think, is between precision and accuracy.

And it may be that if you took, you know, a small hospital and you just took the number of readmissions, you would get a nice number which everyone agrees that's their

readmission rate. But certainly everybody should know that it would be very wide.

As Ashish was saying, you know, it could be -- there's a very wide confidence interval. And how helpful that is to someone, I don't know. You know it is between 10 percent and 40 percent. Maybe it is not very useful.

One way to improve the precision of the estimate is to combine other information yo9u have. This hospital is not operating, you know, in a vacuum. We know that, you know, our best guess without measuring it is not zero percent. It's not 100 percent. Our best guess before we measure it is the national average.

And so the logic is to factor in the information and the degree to which it is factored in is different than the byproduct though. I think that there -- well, I'll stop there.

CO-CHAIR KAPLAN: You could --

excuse me, you could add that if the Committee felt strongly that hospital volume turns out to be one of those issues that we'd like you to address, it could be added is my understanding to your model as it exists now?

MS. HORWITZ: Do you mean an indicator for just sort of the decile of volume of the hospital?

CO-CHAIR KAPLAN: If you -- at the hospital level, in the model if you wanted to include hospital volume, it could be included if the Committee felt strongly that that was some kind of marker that we wanted to take into account?

MS. DRYE: Yes, I mean technically you could add it, if, again, as we talked about before, it is a policy in measurement decision about whether you should add it. So if you want the technical answer, yes.

MS. HORWITZ: I do just want to reiterate the point that, you know, a lot of this discussion, I think, derives from the

publicly-reported measures we have now that are, you know, that are very volume limited because of the open issues. We just do not have that problem in this measure. They are not very small volume hospitals.

There will be much less of this kind of averaging to the mean for these hospitals because we have real data and real census outcomes for these hospitals.

MEMBER JHA: (Speaking from an unmiked location.) -- so I agree that this will be less of a problem with the current measure as it is thought through. I mean I guess first of all one could look at this empirically. And the way to look at it empirically is look at just the small hospitals and see how much -- how different the average is based on your hierarchical model versus a non-hierarchical, straight out logistic regression that handles clustering.

And that would actually tell us are we biasing it by two percent, are we

biasing it by six percent. And we'd actually have some sense of how off it is.

Because the work that we've done with looking at AMI, CHF, and pneumonia, especially for AMI, but for heart failure and pneumonia as well, says that for small hospitals, the predicted readmission rate based on the hierarchical model way underestimates these hospitals readmission rates because volume is actually related to readmission rate.

High volume hospitals on average actually have lower readmission rates. There are more practice, volume, outcome relationships are well established. So the fact that the model remains agnostic to volume is, in my mind, not a justifiable position to take, given what we know about hospital performance.

MEMBER GHINASSI: Michael, and then Frank, and then Richard. And then we've got to move on.

MEMBER LANGBERG: My comment, unrelated to any of this, has to do with validity. And when I think about validity non-statistically, I think about variables in the measure. And if those variables are not included, then it makes me wonder whether or not the measure is valid.

So the variable I'm thinking about -- I've been thinking about it for the last 10, 15 minutes, is mortality rate. So we know that mortality rates do vary. CMS publishes that, at least the three conditions, that can statistically be variable.

And when I think about a variable like mortality, I go to the extreme. And I realize we can bring readmission rates to zero if everyone dies in 30 days. Since everyone dying in 30 days is not necessarily a preferred outcome I think the public would be happy with and perhaps others, then I have to wonder how it is that mortality is not considered as potentially influencing the

validity of the measure.

So is -- are differences, statistically differences in mortality rates, considered in this? If not, does that or does not influence the validity of the metric. We do know that based upon hospital-prepared website that has a statistic on this, I was involved -- actually the Chair was involved in a six-hospital study that looked at difference in mortality through a process of kind of severity adjustment that showed up to a 50 percent difference potentially in mortality for heart failure patients. And also a related association between heart failure mortality and readmission rates. The higher your mortality, the less likely it is you have a population of people being readmitted.

So I was wondering whether or not that is consider. And if not --

CO-CHAIR LAZAR: Do the developers want to make a brief comment about that? Very brief.

MS. HORWITZ: Yes, we think this is a really important issue, too. And wherever possible, we think it is good to pair the admission measures with mortality measures. And so whenever somebody comes up with an all-condition mortality measure, that would be a useful thing to pair with.

When we've looked internally at our data, comparing the hospital performance on readmission with hospital performance on mortality, we do not find a consistent relationship. So it is not true that if you are a hospital with high mortality, you are, therefore, a hospital with low readmission or vice versa.

But that said, you know, it is always important to think about that. And for that reason, that's why we omitted the cancer patients from our model because they do have such high competing risk, we thought it was unreasonable to include them.

And so whenever possible, we did -

- we just sort of tried to consider that for the situations that were dramatic. But again for the existing publicly-reported measures, there really is not a consistent relationship between readmission and mortality. So it doesn't necessarily follow that way.

CO-CHAIR LAZAR: Okay. Frank, and then Richard, and then we'll close this discussion out.

MEMBER GHINASSI: Just briefly, given that the fundamental purpose of this is to improve the quality of care across institutions, I just want to be clear that we're judging the validity of this based on the current set of variables that have been included in the model.

CO-CHAIR LAZAR: That's correct.

MEMBER GHINASSI: I want to also be clear that there have been a number of variables raised in the room that are not included in the model, decided against, some of which may have burden issues, some of which

don't, many of which in the model itself have been pointed to as possibly effecting potential outcome.

So I guess part of it is given that this is an expedited review, we're attempting to assess the current validity of this model based on the variables they are adjusting for, not are we looking at is this the best we've got given what we've got to do.

We're looking at the validity of the model as it stands. I just want to be clear.

CO-CHAIR LAZAR: That's correct, I think.

MEMBER GHINASSI: Okay.

CO-CHAIR LAZAR: Richard?

MEMBER BANKOWITZ: I guess on that note, I think there is no perfect measure. And we can't let the perfect be the enemy of the good.

But in this case, we are presented with evidence in the technical document itself

that as Medicaid increases so does the readmission rate. Now you could argue that the dispersion of the data is there. And there is. But the central tendency of the data shows Medicaid goes up, readmissions go up.

So I think the burden, the scientific burden is to explain why that is excluded on the basis of a scientific reason, not an aspirational reason. And I do think we need to be careful we're measuring the performance of the hospital, not the population whom it serves.

And if you do a little thought experiment to think about taking a high performing hospital and just changing its population overnight to a very complex, chronic, Medicaid population, do you think the readmissions rate will change.

If you think it will change, then you've got to figure out do you want to measure the hospital? Or do you want to

measure the population whom the hospital serves?

CO-CHAIR LAZAR: Okay. Time to vote.

Taroon?

MR. AMIN: Can you just read off the vote? Can you do that, Adeela?

MS. ADEELA KHAN: So we're going to vote on scientific acceptability of measure properties. So we want to consider are both reliability and validity rated moderate or high.

So looking at subsection 2a, reliability, including 2(a)(1). This is looking back at your measure submission form.

So 2(a)(1), we have precise specifications; and

2(a)(2), testing are appropriate methods and the scope with adequate results.

And then looking at 2b, the validity, including:

2(b)(1), specifications consistent

with evidence;

2(b)(2), testing appropriate method and scope with adequate results, and threat stability adequately addressed;

2(b)(3), exclusions;

2(b)(4), risk adjustment and stratification;

2(b)(5), meaningful differences; and

2(b)(6), comparability data sources.

So if we want to vote, was the criterion scientifically acceptable -- was the criterion scientific acceptability of measure properties met.

So we can start voting now.

CO-CHAIR LAZAR: Okay.

MS. ADEELA KHAN: Yes, could we have Patricia as well?

CO-CHAIR LAZAR: Patricia, are you on the line?

MEMBER McDERMOTT: My line is

disconnected.

MS. ADEELA KHAN: Oh, okay.

(Laughter.)

CO-CHAIR LAZAR: Okay. So just from a process point of view to refresh everybody, we've got to meet sort of the four broad criteria. And if the answer is no for any one, the discussion essentially gets truncated and the measure cannot be advanced.

Obviously here we're tied. And, you know, we'll take it forward and see what happens when we get to usability. But had we, you know, had this been a more a -- or had this been a no, then the discussion probably would have stopped right there.

So why don't we move on to usability? We'll get some brief --

MS. FORMAN MORGAN: So for --

CO-CHAIR LAZAR: Okay.

MS. FORMAN MORGAN: Oh, go ahead.

CO-CHAIR LAZAR: Yes, I was just going to say Alexis will give us some brief

comments. And then we'll essentially have another discussion.

The only thing we would ask is that we try and keep the comments as tight as we can. And we try not to repeat them.

The point here, of course, is to simply make a judgment on what is before us rather than trying to create, you know, perhaps a better model.

Mike?

MEMBER LANGBERG: I just have a procedural question. This is clearly an unexpected and probably uncommon experience. Maybe you have had this experience before, I have no idea.

I was just wondering what the criterion was for moving forward. Is the question is yes we move forward? Or if no, we don't move forward? Do you have to achieve a threshold yes to move forward? Or do you have to achieve a threshold no not to move forward?

CO-CHAIR LAZAR: We'll let Taroon

answer. But my understanding -- my picture of the understanding is that a yes, we continue and no we get stopped.

MR. AMIN: If it does not pass scientific acceptability, it does not move forward. So since it is inconclusive, we would continue the discussion. So it's if no, we don't move on to answer the question in particular. And it is by simple majority.

CO-CHAIR LAZAR: Okay.

Alexis?

MS. FORMAN MORGAN: Okay. Looking at our third criterion, usability, for this criterion we were looking at can the audience, the intended audience, whether it is consumers, providers, health plan, can they understand the results of the data? And also is this measure useful and meaningful for public reporting, quality improvement, and accountability?

All right. So the issue is is this measure meaningful, understandable, and

useful for public reporting? And is it meaningful, understandable, and useful for quality improvement, and accountability, which means that it has got a dual purpose in this case, which some of us would worry about.

But for the discussions on the interpretability, the extent to which audiences, the intended audience -- and it is a broad one, consumers, purchasers, providers, and policymakers can understand the results and are likely to find them useful for decisionmaking?

MEMBER JHA: So I think there are some real challenges here to the usability. And they are a couple-fold.

The primary issue here, in my mind, is that -- well, there are two different issues. One is at least on what we've seen for the AMI, CHF, and pneumonia measures, most of the hospitals fall within expected range or don't fall out of the expected range.

And so as a quality improvement

tool, that's a real challenge. If we're all pretty good, it's really not clear that it is a motivating force.

It would be interesting to see, and I assume that the measure developers probably know this, what proportion of hospitals fall out of expected range based on the hierarchal model.

The second issue which kind of is related is, again, how -- what happens to small hospitals. And here if the issue is comparison, it is really easy to see a scenario where you have a large hospital with a readmission -- let's say the national average is ten percent -- you have a large hospital with 14 percent and a small hospital whose actual underlying rate was 20 percent. But the way the shrinkage model handles it, their small hospital gets back to 11 percent because it gets averaged out to the mean.

So what you end up doing is giving information to consumers saying the small

hospital is better than the big one when, in fact, you don't really have, I think, adequate evidence to make that. And so you get numbers that are not meaningful, not understandable, and, in my opinion, not useful for public reporting.

So that's the challenge with restandardized rates with hierarchical modeling is for a lot of small hospitals, it's not the hospital's actual data. It is a series of judgment calls about what is the best amount of information we have about those hospitals.

And we can quibble about whether that is, in fact, the best set of information we have about the hospitals. But ultimately it is not the hospital's own performance.

CO-CHAIR LAZAR: Okay.

Bruce?

MEMBER HALL: Thank you.

I just have a question maybe for the developers. Whether there is an actual

specification about the degree of acceptable certainty. So in my mind, trying to decide whether a consumer can make sense of this, I need to know when are we going to tell them that that performance was good enough? And when are we not going to tell them that?

And what that means is despite whatever method we choose to reach our point estimate what are we telling the consumers about the uncertainty of that estimate? And when that uncertainty says this is acceptable performance or this is not? Because I think the consumers need that to make the interpretation.

And then irregardless of how I may or may not have voted on the prior issue, I would raise the general question if we have 18 experts in this room who couldn't decide whether the specifications meet criterion then how do we expect the consumers to make use of the information?

CO-CHAIR LAZAR: Leslie and then

Mark.

MEMBER KELLY HALL: Just following up on Bruce's comment, to make it useful to the consumer, it has to be easily understood. And I think that the idea of readmission or not is an easily-understood concept until you get into the detail.

My concern is that does the public always equate a readmission with bad versus good. And the unintended consequences, if not explained well, could be patients self-selecting two inappropriate care settings.

And so my concern is really about making this understandable and meaningful requires more education and not necessarily always the idea that something is bad because it is being reported. Just a general comment.

CO-CHAIR LAZAR: Mark?

MEMBER SCHUSTER: I wanted to go back to the shrinkage issue. And I'm assuming you are a consumer expert. But I would think that a consumer, a potential patient, would do

better with seeing 10-40 than an average number with a little asterisk that may or not be in the report but maybe it will be in the newspaper someday that say that average isn't really accurate.

I think it is more usable if the information is more accurate and complete. But in terms of consumer testing, and I may have just missed this in the materials, but it looked like -- well, first of all I couldn't tell what the consumer testing was but it looked like what's been studied is disease-specific measures. And I wasn't sure if we know how people use all-cause measures. If that had been studied, I couldn't find that it had. And that would be useful.

CO-CHAIR LAZAR: Do any of the developers have any comment on that?

MS. HORWITZ: Well, to address these points so far, to me it seems to do the consumer a disservice, to give them a number that's highly contingent on random

variability. And I don't think it is useful to tell a consumer that a hospital's readmission rate is 20 percent when really it could be anything from ten to 40.

And I think actually consumers are not that great at understanding reporting. That sort of makes it difficult to understand.

I think it's much more -- it's a much more accurate representation of reality. It's not a bias. It's a genuine representation of reality to use this measure to determine the most reliable rate for a small hospital, which in many cases will be the median.

Now second, people had asked what is the rate at the point in which we call something better than or worse than average. And that's a policy decision. So currently CMS sets that at the 95 percent confidence interval. And it's a tool that could be set at other confidence intervals. That's not intrinsic to this measure. That's a policy

decision about reporting.

So one would take this measure and report it at the 80 confidence interval and have more outliers. That's, again, not intrinsic to the way the measure is developed.

The measure is developed to be as fair and as unbiased possible with the actual results. And then the way in which you chose to interpret that for the consumer is a public reporting issue and not a measure issue.

And finally you had asked have consumers been tested for this particular measure or for the public report measures. We have not tested this measure with consumers for this measure. And I'm actually not aware of what testing has been done for the public.

MS. DRYE: Yes, CMS did consumer testing to evaluate the understandability of the data on Hospital Compare and make revisions. And so, for example, a revision that was prompted was in the first year reporting, hospitals -- all hospitals that

were fewer than 25 cases, almost all of them ended up in the middle bucket because they were not better or worse by a criterion.

Now those hospitals that are smaller report as a fourth column, which is just too little information to assess quality, which is more accurate reporting for the consumer. I think it is much clearer for people.

But we haven't looked at anything with respect to this hospital-wide measure at this point.

CO-CHAIR LAZAR: Okay.

Tanya?

MEMBER ALTERAS: I would agree, you know, having something that says the readmission rate is between ten percent and 40 percent isn't useful. But I just think that this speaks to the usability of this measure as a whole.

And whether, you know -- and what you're saying about the consumer testing, I

think that the previous way of displaying this data and the current way of showing, you know, not enough cases, those aren't usable either.

And, you know, this all comes down to my concerns about the hierarchical risk modeling in general. And I note these are two different issues. There is the measurement issue and then there is the data display issue.

There's -- you know confidence intervals have nothing to do with what the measure is but I just -- you know, my concern is about this measure and how it is specified. And what the results are going to be in terms of the data display just based on the history that we have with Hospital Compare.

CO-CHAIR KAPLAN: Let me just -- I know there are a lot of people to get to but let me just raise the issue of who is the consumer because it is kind of an insult to consumers to assume that they are a homogenous entity as well. I mean they come in all

stripes and all flavors.

And also the proportion variation, how often a patient chooses the hospital versus chooses the doctor who chooses the hospital is maybe an issue that's worth entertaining here when we're talking about who is reading this data.

CO-CHAIR LAZAR: Mike? Do you want to make a comment?

MEMBER LANGBERG: Thank you.

I'll focus for a moment on the usability of the information from a hospital point of view. And then maybe we can talk about the consumer.

For the purpose of performance improvement, data that is 12- to 18-months old is largely useless. That doesn't mean that we can't learn something from it. But if the purpose of the information is to allow a hospital to apply improvement strategies and retest to see whether the strategies have had an effect, if the cycle for doing that in a

dataset that is so old is so slow that the ability to make real improvement would be virtually impossible on that data.

So secondly, the ability for a hospital to understand the full extent of the opportunities for improvement when evidence currently -- in the current version of this -- there is shielded from the hospital information about other hospitals the patients may have been admitted to is again a deficit with regard to usability of the information for hospital improvement.

Then as far as the -- if I can say about both the hospital and the consumers, the biggest problem I have with the usability is the title. It's about hospital readmission rates. It's about a hospital quality measure.

For those of us who live in that world, there are variables well beyond the license of the hospital and even the function of its medical staff that we know, and I can share with you if you want to, that have

impact upon readmission rates.

So it's less, to me, a measure uniquely of hospital quality and more about community health quality. And I think the confusion for the consumer and certain the frustrations for a hospital would be that tagged to hospital, whether it was socioeconomic status, whether it's resources available in a community such as effective nursing homes, or whether it is distance to the hospital.

A significant number -- proportion of patients we admit into my hospital live 40 to 120 miles away. What's the reach of the hospital to be able to manage patients once they have been discharged. What is the expectation that the hospital should be accountable? And from the point of the CMS, financially penalized if they can't manage these things?

Those, to me, raise questions as to the usability from the hospital

accountability perspective.

CO-CHAIR LAZAR: Thank you.

Karen?

MS. PACE: Yes, I just want to make a couple clarifications about our criteria on usability. And NQF really does make a distinction between endorsing a measure versus how the measure ultimately is implemented and displayed because measures can be used by a variety of implementers and how they display the information at this point in time is not part of the NQF endorsement.

So I just wanted to mention that. And actually NQF has a current task force looking at the usability criterion. And, you know, if they're -- and I think this has been alluded to already but how measures are explained is something that is easily correctable by experts in how to communicate performance data to consumers if that's an issue.

So we do not require that they've

done some testing. If they have done some testing with consumers, that's fine to provide. But we do make a distinction of endorsing the measure and ultimately how it might get displayed on the website or in the report.

The other thing about quality improvement that I'll just mention is that the criteria is about facilitating quality improvement. And so the -- and the specific issue came up when the CSAC and task forces were looking at our criteria and specifically about outcome measures because it often is questioned about how is an outcome measure helpful for quality improvement.

Basically we talk about facilitating quality improvement, which means that if you have data on performance on outcome measures, whether it is this measure or some other outcome measure, and have information that your hospital is doing more poorly than another hospital, that in and of

itself, that information identifies that it is an area for your hospital to pursue to look at what are the issues in your hospital.

An outcome measure will never tell you exactly what you need to do. But it should tell you where you have areas to pursue opportunities for improvement.

So I just want to make those clarifications based on what our criteria actually are at this point.

CO-CHAIR LAZAR: Richard?

MEMBER BANKOWITZ: I echo Michael's comments. And I do want to say this issue of timing is very important because hospitals are not in the position where they can understand these rates across all readmissions. They can only see their own. And it makes it very, very difficult to do a rapid cycle test to change when the data are two years old.

And the second point -- and I don't know where this fits but I would like to

see somehow us think through unintended consequences and perhaps as part of usability, try to elicit whether or not those are happening like if all we do by decreasing readmissions is increase the number of observation stays of 23 hours and returns to the ED through a revolving door, that's not very patient-centered care.

So I'd love to somehow work that in. I don't know how we can do that.

CO-CHAIR LAZAR: Perhaps not.

Laurent?

MEMBER GLANCE: I think there has been a lot of discussion about hierarchical modeling today. And Ashish you made a point, again, about usability with regards to hierarchical versus non-hierarchical.

And I just wanted to say that I think because this is an all-cause readmission where you are kind of folding in all the patient populations into one model versus looking specifically at specific cohorts like

AMI, pneumonia, I don't think that the shrinkage is going to give you that much of a bias here.

I mean certainly we could ask the developer group if they've looked at that in terms of comparing shrinkage versus non-hierarchical modeling. But my gut feeling is that may not be much of an issue here. And so it probably wouldn't effect usability as much as it would if we were looking say at AMI alone or CHF readmissions alone.

CO-CHAIR LAZAR: Okay.

Frank and then Jeph and then I think we're going to call the question.

MEMBER GHINASSI: Yes, just addressing the idea of usability, I was trying to put myself in the position of using this as a hospital administrator, which is one of the roles I have now. And I agree with Michael, I think, about the time lag on this makes this very difficult. So I question the usability around that.

I was trying to imagine myself in the role of a consumer of services, which I am, too, and I would find it confusing I think. And in fairness of advertising, I would presume that if we endorse this and it moves forward, it would forward, I have to presume this, with the list of caveats about what the measure does not measure.

So I would have to assume that the consumers would have to be told we didn't adjust for this, we didn't adjust for that, we didn't adjust for -- these were things that we didn't do because of either burden or complexity.

And I have to also assume that consumers would have to be told in fairness in advertising -- and I agree with Karen's point completely -- that what we're doing is we're assessing a continuum of care, which includes hospitals, it includes the bifrost between hospitals and ambulatory and also includes ambulatory.

And we'd have to tell people consuming this that although we're talking about that continuum, we're only measuring one part of it. And I'm just trying to imagine at the end of that list of caveats, what would I make of it as a measure of the facility itself.

I personally would probably be confused by the time I got to the end of the list. So in terms of usability, I'm just trying to imagine are we going to include all of those caveats? And what does that do to the audience? That's al.

CO-CHAIR KAPLAN: And Taroon, you want to comment on that? It's my understanding that that's outside of the scope of this committee.

MR. AMIN: Well, and I think the -- Karen can probably speak to this as part of usability task force -- but I think there was strong sentiment that no matter how complex some of these measures, that they can be

constructed and reported by people who are experts in this area.

But, Karen, if you have anything else to add?

MS. PACE: Well, that's true. And Tanya can probably speak to it more than any of us, but the other thing to keep in mind is that we really are, you know, public reporting just one aspect of accountability. And we're in the process of broadening our thoughts about usability. And it's really public reporting and other ways of accountability.

Public reporting is not just directed at consumers. Probably those of you who are providers have looked at public reports. And those of you who are referring to other providers have probably looked at public reports as well.

And just in terms of performance improvement, public reporting not only informs consumers or people who are choosing but also provides some external motivation to improve

in activities.

So there's, you know, the public reporting piece is complex. NQF really recognizes more than just public reporting in the accountability realm. And public reporting can be targeted at more than just consumers.

CO-CHAIR LAZAR: Jeff and then we'll turn it over to Adeela.

MEMBER GREENWALD: So, you know, as I think about the kinds of usability here, it's really -- at least one element of that is the impact of the use of this measure. And, you know, just like so many other measures that we see in public reporting, it certainly can be used. The question is to what end. And I'm struggling a little bit with that A, because in an all-cause readmissions is extraordinarily blunt.

And as a hospital administrator, if I were putting on that, you know, given that it is so lumbers, not splitters in

approach, it is very difficult to know what I would use that information for. And two is -- and that's not a reflection on this groups measure but the concept in general.

Number two is -- and I think if were using this, getting back to the whole SES discussion, I might in the precarious role of saying, as a hospital administrator, well, my patients are different. I have a large Medicare population. I have a large research population. I have a population in the middle of North Dakota that doesn't have primary care doctors or all the other issues that we've begun to address.

So I'm struggling a little bit with those sort of -- those paradoxes. And we discussed this. From a patient perspective, again, because it is so blunt as an all-cause readmission, again not reflecting the specific measure but the concept, the usefulness, again, becomes somewhat more limited.

And it becomes more of an elevator

speech in some ways at the consumer level or something that gets advertised rather than something that is truly useful. So I'm struggling -- I starting to struggle with this. I'm not sure how much of that is a reflection on the metric proposed as a conceptual model of all-cause readmissions.

MR. AMIN: Can I just add just a procedural thought as we sort of move into voting for usability and feasibility and really thinking through the fact that this question of -- the criteria is really asking about this measure in particular, about whether or not it is usable and understandable.

And starting with the assumption in the first criteria of importance to measure as the measure was constructed about all-cause hospital readmission, we want to keep in mind what's at hand, which is this measure as specified.

So not to necessarily bring all

those larger questions that we spent a lot of considerable time thinking about at the start of measure -- discussion. So -- because that would have the implications for all the measures going forward.

MS. ADEELA KHAN: So we're for measure usability, we're looking at:

Subcriteria 3(a), is it meaningful, understandable, and useful for public reporting and accountability; and

3(b), meaningful, understandable, and useful for quality improvement.

So to what extent was the criteria usability met? And you guys can start.

CO-CHAIR LAZAR: Patricia, are you on the phone still?

(No response.)

MS. ADEELA KHAN: I think we're waiting on one more person if you want to put in your vote again. There we go. And we have seven to 11 low.

MR. AMIN: We have seven moderate

and 11 low.

MS. ADEELA KHAN: Okay. Moving on to our fourth criterion, feasibility. And that's the required data is readily available or could be captured without undue burden.

So looking at it for clinical measures, the clinical data is generated during care process, which could be the blood pressure or lab samples.

All data elements are in electronic claims. And if they aren't, a plan for the data elements to get to electronic collection.

Susceptibility to inaccuracies and unintended consequences identified.

And the ability to audit the data to capture any of these issues.

And the last one, data collection strategy can be implemented.

CO-CHAIR LAZAR: Okay. So do we expect much discussion on this one? Does anybody want to make any comment on this one?

It's pretty clear where the data comes from, I'm guessing by the expressions around the table.

MS. PACE: Just the one follow-up comment I think that Richard made earlier. If the data is collected at that secondary hospital and it is not available to the first hospital, there is a void immediately to know whether you have a problem or not to be addressed.

So from a consumer point of view then is that measure effective and understandable? And can it be actionable?

If it is not actionable by the prior -- the first hospital because they don't even know it exists, how is the data available?

CO-CHAIR LAZAR: Yes, so I'll try to answer that but I think the issue is that measure as it is specified now either would or would not pose difficulties with feasibility, whether that information gets to wherever one

thinks it should go, to another hospital, to the consumer, to the, you know, the practitioners, the hospital leadership, I think that was really all wrapped up in the discussion of usability.

So from a pure feasibility standpoint, the issues of inter-hospital transfer of information are not really -- at least if I understand it, are not really germane. It's essentially billing data that is essentially transmitted. Frankly, CMS has it at present.

Any other comments?

MS. HORWITZ: Can we just correct a factual issue? So as currently done with public reporting, you, as a hospital, get a full list of readmission no matter what hospital they go to. So you are given that information no matter what hospital the patient goes in.

CO-CHAIR LAZAR: But what you don't get --

MS. DRYE: You get whether your patient was readmitted -- sorry -- but not to what hospital.

MS. HORWITZ: But you know how many of our patients get admitted to other hospitals.

MS. DRYE: And that's a CMS decision. And as we mentioned before, you know, CMS could give you the hospitals to which they are admitted if there is a clamoring for the data.

CO-CHAIR LAZAR: Okay. Other -- yes, Karen?

MS. PACE: Yes, and I would just again clarify that that's not a function of the measure. That's a function of data availability.

So the idea of creating a measure that wouldn't include readmission to another hospital would be the alternative. And I'm not sure that that would really serve the purpose.

So as has been stated, that data exists. And certainly hospitals are given what patients are readmitted. And, you know, definitely can push for the additional information of identifying the specific hospital. But it's not necessary a measure property that's how it is implemented.

CO-CHAIR LAZAR: Right. Okay. I think, Adeela, we can move directly to vote on feasibility unless anybody has something very pressing to offer.

(No response.)

CO-CHAIR LAZAR: Okay.

MS. ADEELA KHAN: Okay. So again we're looking at:

4(a), data done during care;

4(b), electronic sources;

4(c), susceptibility to inaccuracies identified, unintended consequences identified; and

4(d), data collection can be implemented.

So to what extent was the criterion feasibility met? You have high, medium -- high, moderate, low, and insufficient.

You can start your vote.

And we're -- is Patricia online?

(No response.)

MS. ADEELA KHAN: Well, we're short one person. Can you all enter your vote in again please?

So 11 for highs, six for moderate, one for low.

CO-CHAIR LAZAR: Okay. All right.

We now have to move to an overall vote on the measure itself. And I'll just to reiterate what I think a number of folks said earlier. And that is the topic for the two days is all-cause readmissions.

And what we're looking at now is an evaluation of the specifics of this particular measure. So as Karen said earlier, it is less about your -- or any of our

philosophic views on whether all-cause readmissions are a good thing or a bad thing but really on the particular characteristics of this measure. And obviously tomorrow we'll have some comparative discussion.

So I think we've discussed this, you know, in great depth and, you know, repeatedly on some of the issues. Certainly we can open the floor briefly if anybody has got something new to add to the discussion that we have not considered before. But I think we'd all probably want to try and avoid, you know, rehashing the same points and same arguments that have been so cogently made earlier.

So let's open the floor.

CO-CHAIR KAPLAN: Well, let me just add here at the risk of reiterating myself, as a measurement scientist, you always have a better mousetrap. You know people in my biz have been estimating intelligence for over a hundred years. Even the SAT scores are

now new.

So in sort of where we are, is this a good representation of what could be a terrible concept but at least it is trying to estimate all-cause readmissions. Whatever you think, as you've just said, whatever you think of the concept itself, if we're trying to estimate this construct, whatever it is works, all-cause readmission, is this measure a good measure.

MEMBER JHA: Excuse me. So I'm not interested in going -- reiterating any of the arguments I made earlier. It seems to me that a bunch of the issues that were raised are essentially empirical questions that if we had better data we could answer.

And I'll just give a few -- and these are not even things that we need six years of data collection. These are things that potentially the measure developer could give us relatively quickly.

So one thing certainly is how much

does a hierarchical model -- so Laurent, to your question, I agree with you. This may not be as much of an issue. And actually the data will tell us that. So there are keys that could be done.

The second one about whether consumers are more likely to understand on the usability issue confidence intervals or if they are more likely to understand hierarchal models with predicted rates based on a series of assumptions is an empirical issue. I mean you can decide which is more usable based on surveying people who would use these data and saying which is more transparent and more understandable to you.

It seems to me that my vote on the overall might very well change based on the answers to those questions. And I suspect that there may be other people who might feel differently if the model -- if the hierarchal model really effects small hospitals in a big way or if it is trivial and doesn't have much

of an impact at all.

So I understand we're voting. I'll vote. But I guess it is almost a process question. Is there a way for the measure developers to come back with more information to clarify these issues? Or is that really out of the scope of what we could ask?

MR. KRUMHOLZ: Can I ask Taroon and Karen to ask the question of do we -- if we do something now, how long are we approving this measure for? And are we going to revisit some of these issues and ask these questions at some point as part of our constant improvement process on your end for these measures?

MR. AMIN: Well, I'll start and Karen, feel free to jump in.

So measures that are evaluated through this process will be reviewed under maintenance in a three-year cycle to evaluate as one of the measures that you'll consider today by United Healthcare. So there will an

opportunity to revisit the measure and some comments from the field on the implementation of it.

The question on hand of whether or not we can ask the measure developers for further information and then they'll reply, I think I'll sort of leave this to Karen and Alexis. In some ways, this is the first time we're going through an expedited review.

As part of the expedited review, there would have been -- we would have liked to have more back and forth with the developers not only for this measure but for others. But for the sake of time we just didn't have that flexibility.

I'm not sure that we -- I'm pretty confident that we will not have the flexibility to have the measure developers respond to comments back to the committee because we just don't have that much time in the cycle.

There was another comment that I

wanted to make but I think I'll leave it with that. If you have anything to add Karen?

MS. PACE: Well, I mean, I think we're going to have to see where we're at at the end of today. And see what makes sense for moving forward before we, you know, some of these things could probably be answered in a relatively short turnaround. I think we'll have you vote on the measure as it is currently specified. We'll do that for all of the measures.

And, you know, we'll -- tomorrow maybe we'll end up identifying questions that we want more analysis about. And then come back and vote on the measures again. So, you know, we are on a pretty short time frame but, you know, we'll have to see where we're at.

The only other thing I just want to mention, just because it is about the distribution and the shrinkage, in the technical report there are some graphs. And I don't know if you can bring those up on page

58 -- and Elizabeth, if you want to refer us to another place -- but I think it does demonstrate that there is a spread of scores. It's not that everyone -- every hospital is looking average in this data.

MS. DRYE: And just quickly from the applications on the bottom of page 38 and the top of page 39, which is --

PARTICIPANT: What item number?

MS. DRYE: The number is 5(b) -- 2(b)(5.3) but the range is 12.5 or 12.6 to 22.8. And then it tents the 90th percentile, which I think is more helpful. It's 15.4 to 18.2. And so this is the range of the point estimates.

But I want to want to emphasize what others have emphasized, which is we have, you know, really a much higher volume for these cases so we haven't presented a confidence interval or an interval estimate around these. But we expect it to be much, much tighter than for other measures.

For AMI, there are over a thousands with one patient. That's why we see a lot of hospitals with too few cases to assess the underlying, you know, inherent quality of the hospital. But for this measure we don't have that problem at all.

And whether we could come back with some numbers on it, I mean I think we can. We'd just have to talk about the time range.

MS. PACE: But are we -- those are the distributions of the --

MS. DRYE: Yes, the left is the overall -- where the five cohorts are rolled up. And the other is for the individual cohorts. So, again, if those are the point estimates but the precision of the point estimates in this model versus condition-specific models is expected to be much higher.

So a narrower confidence interval and choice around when you take 95, 90, 80, etc. In other words, we think that it will be

easy to find a lot of outliers with this measure.

CO-CHAIR LAZAR: Okay. Bruce?

MEMBER HALL: I think Jo Ann was first.

MEMBER BROOKS: Oh, no, my question was answered. Thank you.

MEMBER HALL: I was just wondering technically if the NQF side could advise us on -- I realize, for instance, that reporting by stratification and SES or whatever is an implementation issue more than a measure issue to speak.

So do we as a committee have an ability to say -- you know whether we vote or yes or no -- to say if this were to move forward, we would strongly advise implementation with a consideration for stratification and so on? Do we have -- what's the limit of our ability on that?

MS. PACE: I think -- well, first of all, we do have a mechanism to recommend

measures on certain conditions. And, you know, a condition could be stratification. But then the measurer developer would need to respond to that whether -- what their response to that condition would be.

Our process would be we want you to vote on the measure as is. And then if someone wanted to propose a particular condition, we could certainly entertain that. And then we would ask the developer to respond to that.

So stratification can be an implementation issue. But it can also be a measure specification. And in, you know, if it were an implementation issue, it might be comparing like hospitals. If it were a measure specification issue, it might be for each hospital stratifying the results by a particular factor.

So, you know, doing this on the fly, you know, generally steering committees are not developing measures as they go along.

But it is something that you could request more information from the developer about possibilities in that regard.

But I think the first thing is to vote on the measure as it is specified. And then if there are conditions or questions, we can certainly ask the developer to respond to those. Not necessarily today unless they want to briefly. But again I think we're going to have take stock of where we're at at the end of the day and see how we want to move forward.

CO-CHAIR LAZAR: Sherrie?

CO-CHAIR KAPLAN: I just want to, as we go forward and vote, remember that it is usually a bad idea to try to measure weight with a rule, you know? And the sort of calibration issues, are they measuring all-cause readmissions? That's the validity question. You're not using a ruler to measure somebody's weight.

How it is calibrated and all that

other stuff is a precision issue. And it depends on the purpose of measurement.

Those of us in the measurement business get freaked out when you don't bring that into the discussion. But because NQF has decided to separate those two, my understanding -- correct me if I'm wrong, Karen, this is a measure for whatever purposes, for accountability and quality improvement, that's the purpose of measurement. And we're hoping we have a ruler to measure height. Right? Is that -- okay.

CO-CHAIR LAZAR: Mark and then Tanya.

MEMBER SCHUSTER: Yes, Elizabeth, I thought these graphs were really helpful and I'm wondering if you happen to know if it is the case that the small hospitals are all sort of in the middle and the outliers are the larger hospitals?

And to get to what Ashish and maybe someone else asked earlier, do we know

what percentage of hospitals actually do fall below the cutoff and our low end so if the whole shrinkage is even effecting them? Do you have -- like it is five percent? One percent? Do you happen to know roughly that number?

MS. DRYE: I don't think we have right in front of us this second the distribution of volume at these hospitals but we can come back with that easily. And, again, this is the distribution of the risk standardized rates. And so then the construct -- those can be used in themselves. And there's quite a bit of range. There's a ten-point range there.

But the construct CMS uses on Hospital Compares is to put -- you know, use a confidence interval around that to assign hospitals to better, worse, or no different.

And we're just saying with this measure, even with the 95 percent number, which is high and, you know, you could use 95,

90, 80, whatever. It's not really specified in the measure. We would expect to see many, many more outliers and many, many fewer hospitals in the too few category, if any.

You know I don't know if there are hospitals that don't have fewer than 25 cases but we can come back to you with those numbers. And I apologize we don't have that, you know, exactly for you at this second when you have to vote.

CO-CHAIR LAZAR: Tanya?

MR. KRUMHOLZ: Well, I just want to elaborate one second because --

CO-CHAIR LAZAR: Could you introduce yourself?

MR. KRUMHOLZ: Yes, I'm Harlan Krumholz. I'm one of the developers.

The -- I mean I think what you have to think about is I don't think you should be thinking so much about empirically what counts but what are the volumes. And what does this mean to address? And how does

this extend what we've been doing so far?

So if you are interested in a hospital signal and you know that a lot of readmission measures, they track each other. That's what the group has been able to show. And you know that by looking at the hospital-wide, you're going to increase your volumes for any specific period of time that you're looking at, then you can be confident that in that particular aspect, it is a much better situation than what you've currently got.

And you can look empirically and we can provide you the evidence about what the numbers are, the outliers. But more importantly, looking at the volume shows you that this allows the small-volume hospitals to be in the game because truly, I mean I would expect this conversation to be more relevant if we're talking about MI, heart failure, and pneumonia, where really what we've done is suck those hospitals into the middle and we've made it very hard for you to be an outlier,

high or low, in those conditions.

And that's a problem that we've struggled with for a long time. And it's because of the amount of information that is available. But in this particular case, I think you guys should -- I mean from our point of view, this is much of an advance in terms of being able to discriminate performance and be fair to the small hospitals by being able to draw on their experience. We just didn't have enough experience because so many hospitals see fewer than one a month in this nation for some of those particular conditions.

MEMBER ALTERAS: Is there any way we could hold off on voting until we get the information that Elizabeth was talking about? Just because -- I mean I've been critical of this measure during the discussion. But, you know, underlying all of that, I'm a big advocate for the all-condition -- all-cause readmission measure.

And so I would hate to then vote against this if it turns out that the whole issue that I've been concerned about isn't really an issue at all. And that's pretty important information for us to have I think.

MS. PACE: Well, we probably can. I'll see what the Chairs think about this. I think we have to be very specific about what question you have and whether that is something that can be provided in a short time.

But we may want to get a sense of the group whether people want to get that information before they vote. And I think then we can be flexible.

CO-CHAIR KAPLAN: In just chatting with Eliot briefly about this, it is my sense that we should take a vote now. And that in the comparison, because it may very well help the group make these decisions when we come back to compare the other two measures tomorrow, because if the consensus of opinion

at that time is this is not a perfect measure, this is not a perfect world, but it's doing the best we can with what we have now.

So I think that discussion is probably a better reflection of, you know, maybe we -- in NQF history and vote that nothing is appropriate for measuring readmissions tomorrow. But my sense is that that kind of relativeness will help us understand where we are and aren't. And is this an okay measure in this worst of all possible worlds?

CO-CHAIR LAZAR: So I think what you're hearing is a recommendation from us that we vote now with the option of sort of revisiting. We don't have to make it -- you know, we don't want people to feel that it is an absolute.

MEMBER SCHUSTER: Well, I guess that's just a clarification. If the vote were no, would we still compare tomorrow? I think that's important to understand.

MR. AMIN: So procedurally, as we set out the competing measure discussion for tomorrow, it would not go to tomorrow. Now if you feel that we really need to bring them all back for the head-to-head comparison, which we will have to do because most of these measures are competing, that's the will of the committee. Right.

I mean you would have to -- well, I guess it is a question. I mean the true procedural decision would be is if it doesn't move forward, then there is no competing discussion tomorrow.

However, if the committee felt otherwise and wanted to suspend, you know, that rule and then have that brought up for tomorrow, I mean that's up to the will of the Chairs and the committee.

But procedurally, strictly speaking, if it does not pass overall recommendation for endorsement today, it does not go to a competing measure discussion

tomorrow.

CO-CHAIR LAZAR: Okay. Let me make a suggestion and see if everybody agrees to it. Why don't we go through the three measures, you know, according to the protocol as it stands? And then depending upon what those results are, we can make a decision tomorrow about whether to revisit or do a competing analysis or what have you?

I think, you know, here sort of common sense and the desire for a good product ought to prevail. Is everybody comfortable with that? Okay.

CO-CHAIR KAPLAN: Let me propose a quick alternative, Eliot. I think maybe could we suspend the vote on the individual measures until the end of the day and then vote each one of them at the end of the day?

MS. PACE: So that kind of implies that you want to do some comparison?

CO-CHAIR KAPLAN: Why don't we call the committee on those two options

because the two different options. So if you'd rather vote now and then if you vote no understand that you don't have the option to compare tomorrow procedurally or --

MS. PACE: Let me just clarify. I think we have much more flexibility. I mean generally we, you know, if a measure doesn't pass, then it doesn't go into competing measures. We have kind of a unique situation here with having only three measures, all of them competing, and trying to address a Congressional mandate.

So obviously although it's not required, we would like to see a measure that is good enough to come out of this process. And that may take some back and forth with the developers.

So I don't think we want to present it as such an absolute thing. I think we should vote on the measure as it is. We'll do that for each measure.

And at the end of the day take

stock of where we're at. Regardless of how measures are voted up or down, we'll take stock, see if there are any issues that need to be revisited.

So we've not going to make this like -- this is not a final, not revisit vote at this point. And we want this vote to be based on, you know, given the measure and not think about either the other measures that are on the table or any kind of hypothetical measure in your mind because those don't exist.

So we're just looking at the measure as it is specified and with all of the analysis and data that were presented to you. Would it be suitable for endorsement? And this is just a preliminary, as I said, you know this can be revisited if more information would help you.

CO-CHAIR LAZAR: Is everybody comfortable with that?

Okay. So let's move to a vote.

Adeela?

MS. ADEELA KHAN: Okay, voting on the overall suitability for endorsement, does the measure meet all the NQF criteria for endorsement? And this is -- the final recommendation again will depend on the assessment of competing measures. So we're voting yes or no. And you may start.

We have one person left. You might want to try it again. And that's eight for yes and ten for no.

CO-CHAIR LAZAR: Okay. It's time for lunch break. Let's come back in 30 minutes and we'll start again.

(Whereupon, the above-entitled meeting went off the record at 1:49 p.m., and resumed at 2:22 p.m.)

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

(2:22 p.m.)

CO-CHAIR LAZAR: Okay. Everybody set? Terrific. So we're going to move the afternoon portion of the meeting to Introduction of Measure 0329, Risk-Adjusted 30-Day All-Cause Readmission. And the developers are here to give us a very brief overview.

MR. STETTLER: Thank you. So I'm Ron Stettler, from UnitedHealthcare. And to give you guys a little bit of a background, we submitted and got approved the original -- this original metric in 2008, and this was intended as an update.

And in general, what we're going to propose is a relatively major update. So over the last three or four years, we have determined that the metric can be improved, and what we're proposing is the improved metric.

And what we want to focus on, and

what we've been focusing on, using this metric internally, is for quality improvement. So we've had great success with this metric internally for quality improvement, both at individual facility levels, but generally across our markets and across our states.

And the other reason we wanted to make a change to this metric was, we were approached by NQF about making this capable of being applied to electronic data. So electronic health records, and so on. So in order to do that, we had to make a few modifications to the process.

The other thing I'd just like to point out is that within -- and just by using the historic metric, we have been able to increase and improve our results tremendously. About a 7 percent improvement in our Medicare readmission rate over the last three years, 3 percent in our commercial readmission rate. Again, mostly thanks to our partners at the hospitals, but in general based on a lot of

the quality ensurement programs we've implemented.

So our focus in the metric is to make it transparent and usable, as well as making it accurate enough to adjust out as much of the variation as possible that's due to case mix. So the idea on this metric is really a case mix-adjusted, weighted process that will eliminate a lot of the characteristics of the initial admission from the outcome, the outcome being the risk of readmittance.

And we tested it using kind of classic hold-out samples, and using our own internal data set, which -- about 20 million members, 4 to 5 million admissions. We used it to develop the actual factors themselves. And what we would propose is, we will make public our factors.

And those factors would, in a sense, replace the current factors that are out in the existing metric. So the existing

metric attempts to use the CMS DRG consumption weights. We propose replacing those with weights internally developed, specific to readmit. And we think that it improves -- in fact, we know it improves the accuracy of the model tremendously.

The categorization process we are going to employ is a -- is based on the AHRQ procedure and diagnostic groups. And there are two different sets of categories, one for 0-64 and one for 65+. And there's about 220 groups for the commercial population. We'll have a little less than 200 groups for the Medicare population.

And again, in our testing, what we found is that once you apply those weights, you can basically predict the ultimate results in the hold-out sample, or in the next year on a longitudinal basis, to a great degree. And we tried to make it simple. We tried to make it easy to implement.

It's implementable based on only

the admission that you're looking at individually. You do not need longitudinal data to create, or to calculate your risk at that point. So a facility can use it concurrently, without a lot of look-back.

Our implementations usually require a six-month run-out period. You need to have some run-out to let the lag for the readmit to actually occur.

But in general, we think that it really does give you an advantage in terms of the practical implementations of quality improvement. It doesn't rely on older and staler data.

In our implementation, we do share all the readmit information back to the facilities when requested. So if you have a -- you know, if you're a facility that does get a lot of tertiary care referrals, and so on, and transfers from others, you do get to see where those go and where they come from.

And I guess I'll stop there and

just open it up for questions. Or I'm not sure what the process is.

CO-CHAIR LAZAR: Okay. So are there specific questions for the developers? Any specific questions for the developers? Laurent?

MEMBER GLANCE: So if I understand correctly, your methodology is based on direct standardization, correct?

MR. STETTLER: Correct.

MEMBER GLANCE: And so for any individual, for any specific hospital, you have roughly these maybe 200 bins or so, and you determine the proportion of patients that were readmitted within each one of those bins, right? And then you apply those proportions to the standard population in order to obtain a standardized readmission rate.

And my question is, are you concerned that, within specific hospitals, that those cells may be very sparsely populated, and that that could lead to some

instability of the estimates that you derive based on those rates?

MR. STETTLER: So the weights are definitely developed across the entire population of data initially, and then they're applied at the individual facility, at the individual bucket level.

MEMBER GLANCE: So the rates are not based on the individual hospital rates.

MR. STETTLER: No, the rates are. The actual -- so, we do calculate an actual readmission rate for each individual facility, and we calculate an adjustment factor for each facility based on those admissions. So you get -- it's very similar to applying the CMS consumption weights, to get a risk-adjusted length of stay, or a risk-adjusted cost per admission or cost per case. The logic is really very similar to that process.

And depending on the case mix of your admissions in that time period, you will generate an average weight. And that average

weight may be higher than average, lower than average, or average. And you use that to adjust your actual value to get an adjustment mark.

And that adjustment mark is what you would use to compare yourselves to yourself over time. So if you had a large, changing case mix over a two or three year period, it wouldn't be right to say "Yes, I've done a great job with readmit reduction" or "I've done a poor job with readmit reduction." You would want to be able to adjust for a change in that mix over time.

And that's basically what this would do. And it also allows you, of course, to compare yourself to the other facilities with potentially different mixes.

You know, we don't solve the problem of low volume. If you have a really low volume, then obviously the validity of the result's going to be -- you're going to have a higher confidence interval in the results

than you would if we had a very large, very high volume facility.

MEMBER GLANCE: But it would just -- just to -- I just wanted to clarify, but again, you're using direct standardization, and the rates for the bins -- for those individual cells -- are calculated for each specific hospital. That's very important, because you're not basing those rates on the entire population. You're basing those rates for the individual cells on individual hospitals.

MR. STETTLER: The rates of cells, the actual rates of a given --

MEMBER GLANCE: Within each cell

MR. STETTLER: Based on the individual facility's report.

MEMBER GLANCE: Thank you.

CO-CHAIR LAZAR: All right. Well, let's move to the four areas that we have to cover. And obviously in the comments, particularly around scientific acceptability,

there may be other questions that come up that would require a concise response from the developer.

But let's start with the importance to measure and report. Have we -- you know, we did this earlier. I think we've gotten past the issue of, "Is this a valid outcome, this measure, or not?", and come to some accommodation on that issue. Does everyone feel comfortable with the criteria within this question, and are we ready to turn it over to Adeela for a vote?

(No response.)

CO-CHAIR LAZAR: Hearing no dissenters, Adeela, take it away.

MS. ADEELA KHAN: Again, on importance to measure and report, we're looking at high-impact performance gap in evidence. Was the threshold criteria, importance to measure and report met?

And we're voting yes or no, and you can start now.

(Whereupon, a vote was taken.)

MS. ADEELA KHAN: So we have 16 for yes. We're missing two people.

CO-CHAIR LAZAR: Okay. I assume we're comfortable with a substantial majority and we can move forward. Yes?

(No response.)

CO-CHAIR LAZAR: Okay. So let's move to the issue of scientific acceptability. And as we did earlier, why don't we tackle reliability first, and then we'll talk about issues of validity. Discussion around reliability?

DR. GHINASSI: I think this is in reliability. I hope it is. If not, then just keep moving. There was a phrase where -- and I'm paraphrasing, but it said "Risk adjustment calculation happens two ways: retrospective analyses of hospital performance determination, and real-time EHR environment."

I'm just unclear, in terms of reliability, how those two methods are going

to occur simultaneously across the concurrent national landscape, where there's such a mix between paper-based systems and electronic systems. And how does that play out, methodologically.

MR. STETTLER: That's a really good question. We've implemented it on our notification area. So UnitedHealthcare gets notification data from most of the hospitals in the country, and we get it in real time, or close to real time.

And so we've actually implemented it and compared the results to the claims-based version. One of the goals in this process was to make a single metric that could be applied in both environments.

And all of our reliability testing that you're seeing in the documents are based on the claims-based side. We have not yet completed the electronic portion, but we do believe that the weights may very well be different, because the data is -- may be less

complete, or at least it's going to be different on the electronic medical records side than it would be on our claims data.

And we think that, chances are, you will need a different set of weights for the EHR sources. So that was the main thing we were trying to get to with that statement.

CO-CHAIR LAZAR: Karen?

MS. PACE: So I think -- we didn't get these specifications, correct? So you're evaluating this measure as a claims-based measure, because that's what's before us now.

MR. STETTLER: Correct.

MS. PACE: All right.

MR. STETTLER: The documents that you have before you -- again, the point was that, in theory, it can be applied to the e-side.

MS. PACE: Right.

MR. STETTLER: That was one of the changes we made.

MS. PACE: Okay. The thing is, we

have some very specific requirements for electronic health record specifications and testing, which we don't have. So I think, just to get everybody on the same page, you'll be evaluating this measure as a claims-based measure.

CO-CHAIR KAPLAN: Can I ask a point of clarification from Karen? It's supposed to change to be translatable to an electronic environment as well. How much of a handicap for this measure is that criteria?

MS. PACE: I'm sorry, could you --

CO-CHAIR KAPLAN: Yes. If we are evaluating this measure on its electronic venue, if you will, how much -- as well, without data or evidence to support or plans for -- or specific detail of plans for, as we saw in other applications, how much are we valuing -- is that a limitation of the documentation that's been provided?

MS. PACE: Well, I think -- what I was saying is that I think that, although they

mentioned that it could be done in an electronic health record environment, we don't have the kind of specifications for that, or it hasn't been tested that way.

So I think we just won't even consider that, for purposes of this discussion and evaluation of this measure.

MR. AMIN: So this measure would have recommended for endorsement, it would be only applicable for administrative claims data, as specified in their submission, that the data source is administrative claims. It's been tested in administrative claims. There are plans, as described by the developer, although what's in front of you is a claims-based measure and should be evaluated as such.

CO-CHAIR KAPLAN: Can I ask one other follow-up question, procedurally? To what extent does the Congressional mandate include electronic health records? Or does it?

MS. PACE: To my knowledge, it does not at all include them.

CO-CHAIR LAZAR: Okay. Ashish?

MEMBER JHA: So kind of staying on that theme, I guess, should we factor in at all the ability or the fact that we may be able to move towards electronic sources of data at all into our thinking about this, or would you -- I'm seeing from Eliot the answer is no. Is that right?

CO-CHAIR LAZAR: I think the answer is no. I think we're strictly looking at this in terms of its criteria as a claims-based measure. And the promise of translatability or applicability to an electronic record, I think we're going to have to keep entirely separate, as tantalizing as that may be.

Other comments? Bruce?

MEMBER HALL: So I take that instruction to mean that this paperwork would actually need to be edited, that if we did

decide to pass 0329, we would not be passing the portions of it that are discussing the e-submissions, so to speak.

MS. PACE: That's correct. I mean, most of that discussion is descriptive, versus actual specifications or testing, so --

DR. HALL Right, but I wouldn't want there to be confusion about us -- we're sort of tacitly saying "0329 has been approved," and then someone going back and saying "Okay, well, there it is."

CO-CHAIR LAZAR: Yes, I think you make a very good point. Other comments about reliability?

Okay, Sherrie.

CO-CHAIR KAPLAN: I have a question, actually, about the precision using your past experience with this information. Because we didn't get as much information on that topic. What is your experience with the measurement error of what you're doing?

MR. SCANDRETT: So again, as Ron

said, we tested -- are you talking about the new measure, or the old measure that this is replacing?

CO-CHAIR KAPLAN: That's a confusion to me. I guess we're talking about the new measure in front of us, yes.

MR. SCANDRETT: I mean, we did testing that was very similar to what was shown under the previous, the last presentation, in that we tested to see how well it fits in terms of the intercept of one slope. I mean, it tests very well, whether you're comparing the actual path of the data to a hold-out sample, it tested very well. Whether you're comparing a two-year look versus a subsequent third year that was not included, in terms of a longitudinal ability to fit, it seemed to fit very well.

We also tested national versus regional, specific factors, and those matched out very well. So is that sort of what you're getting to, or --

CO-CHAIR KAPLAN: Well, let me push you a little bit on what "very well" means. Because what we've got so far, or what I heard you say, was test/retest reliability.

MR. SCANDRETT: Yes.

CO-CHAIR KAPLAN: So what is the measurement error proportion of what you get at the hospital level? Do we know what's the true score variability at the hospital level, and what belongs to measurement error? Because you've had some experience using this, right?

MR. SCANDRETT: Yes. I mean, that stuff we got to we didn't -- wasn't part of the original question. We have looked at that subsequently. Again, I guess this comes to how much of -- when you fit to a hospital-level result, how accurate do you want that to be?

I mean, you could create a model that takes out all the variance between hospitals and it would fit them very well, but

that naturally doesn't tell you much about relative quality performance of those hospitals, if you could exactly predict how that hospital performs.

CO-CHAIR KAPLAN: I wasn't actually asking the validity question, which is next.

MR. SCANDRETT: Oh, okay.

CO-CHAIR KAPLAN: I was asking the precision question, which is the reliability issue. You talked about test/retest, but I was wondering, what part belongs to measurement error of what you've done, and what's the real probable true score variation that you're seeing at the hospital-level measure?

MR. SCANDRETT: Yes. We did not test that at the hospital level. We were focusing on testing the individual buckets, the RRC categories that we divide everything up into, and then comparing those to see if those were accurately predicted over time,

repeatable over time, rather than at the facility level.

CO-CHAIR LAZAR: Okay. Comments about -- any other comments about reliability?

(No response.)

CO-CHAIR LAZAR: Let's turn to validity. Comments or questions about validity?

CO-CHAIR KAPLAN: Let's read the criteria again, just to make sure. Can you read them?

MR. AMIN: Yes. Validity is assessing whether the specifications are consistent with the evidence. The extent of the validity testing at the data element or at the measure score level. Basically where you just were.

Justification of exclusions, whether they relate to the evidence. Risk adjustment, the risk adjustment model. Identification of differences in performance and comparability of data sources and methods.

If there -- this is a procedural matter. If there is anywhere in the measure that you want us to focus on and bring up in this reading, we're happy to do so. Also, obviously, there are the preliminary evaluations here that you can reference in your discussion as well.

CO-CHAIR LAZAR: Okay. Comments about validity. Frank?

DR. GHINASSI: Just a point of clarification. On the document that we were sent, which was sort of a summary document -- it was on page 6 -- you go into a fair amount of detail -- great detail, by the way -- on quoting verbatim the specific guideline recommendations for what looked like process verification, aimed at obviously facilitating an effective transfer of care.

And it's a great list. I mean, it goes on to page 7, and there are ten or eleven sub-bullets. I was just confused about how that -- how the inclusion of that plays into

what it is you're measuring, and how does that relate to the validity of the measure that's being presented?

It's a wonderful set of processes, I'm just not sure how it all fits into what this particular instrument's going to do.

MR. STETTLER: So that's page 6, the specific guideline recommendations?

DR. GHINASSI: Yes, it starts on 6 and ends about midway through 7.

MR. STETTLER: I'll be quite candid. I mean, we put in the guidelines that our clinicians use. We tried to follow the form, and tried to apply it here.

DR. GHINASSI: But I mean, does that play a role in this particular measure?

MR. STETTLER: I would have to read through them, honestly. I didn't put these in. The clinicians part of our team did.

DR. GHINASSI: I guess the technical question is if you could tell us

what role, and if it does, how we'd be asked to evaluate the measure with the testing.

MR. SCANDRETT: Yes, I think these are just references to what the guidelines are for, I guess, quality of care to avoid readmissions.

MS. PACE: That would just be information about the relationship between this outcome measure and positive instructions of care. So it's back on the idea of it being a health outcome, and are there structures and processes of care that can influence the outcome.

So that's just the list --

DR. GHINASSI: I just want to be clear that, as defined, the measure -- I didn't see it, anyway -- doesn't suggest any way for any of these variables to be included in the measure.

MR. STETTLER: That's correct.

DR. GHINASSI: So I was just confused about why they're there. That's all.

CO-CHAIR LAZAR: Okay. Other comments or questions on validity.

(No response.)

CO-CHAIR LAZAR: Okay --

CO-CHAIR KAPLAN: Mark is --

CO-CHAIR LAZAR: Mark, I'm sorry. I missed you.

MEMBER SCHUSTER: I don't know if I should bring this up now or at usability, but I was curious about the decision to go from -- zero to 64 seemed like a very wide age range, and I'm just not used to seeing such a wide range. And I'm just wondering how that played out with case mix adjustment, how this plays out for children's hospitals versus general hospitals.

Often it seems like case mix adjustment, just in general, not in particular to all-cause readmission, is just very different for pediatric populations.

MR. STETTLER: Yes, and that's one of the things we've actually -- we wanted to

make sure it was inclusive of pediatrics, because our prior method actually wasn't. It specifically disallowed under 18.

So we wanted to make this more specific. I will tell you that the OB and the peds diagnostic categories work out very well. They test well. I do think that one modification we may want to make is to split it into three, to add a 0 to 17 set of factors, as well as a greater than 65 and an 18 to 64.

But the way it works now, they actually do test pretty well, and they actually work pretty well. Again, I would go to stratification, would be my main emphasis there. Would I want to compare a pediatric hospital in with the rest of the general hospitals in my measure? I think I would go into a stratified process and compare them to others of the same build. That would be, I think, the way to implement it. But again, that's kind of a process question for who's

going to implement it, so --

CO-CHAIR LAZAR: Okay. Michael?

MEMBER LANGBERG: I'm the one asking this question, which makes me suspect the answer's really easy, but it's on 2.b.4, on the risk adjustment strategy. The increase that was in the packet we got was simply this measure itself is a risk adjustment. And I didn't really understand what that meant. I'm sure others do, but I didn't understand what it meant.

MR. SCANDRETT: Yes. I mean, I think that may have just been how we sort of interpreted the questions. The answer to that is that the questions were actually placed in there, but that is the entire purpose of what we've done here, to create these factors based on the categories, the diagnosis categories and so on -- the sole purpose of that is to, in fact, risk adjust everything.

And that's sort of -- risk adjustment data is probably spread out

throughout the other sections. It's not really -- creating a 30-Day readmission rate without impacting adjustment would, I guess, not be that useful. So it's really -- all the risk adjustment -- the entire purpose of all these calculations, and creation of these buckets, is to do the risk adjustment, so it's spread elsewhere in the answers in this document.

CO-CHAIR LAZAR: Okay. Paula?

MEMBER FOLTZ: I didn't see a risk adjustment for AMA or transfers, or rehab. Is there a reason for that?

MR. STETTLER: For transfers, do you mean same-day transfers? So, one of the most important methods about this, obviously, is defining the case itself, and making sure that you're looking at unique admissions and so on.

So one of the most important aspects is to try to remove, or to group the transfers together. So our recommendation is

that a case that's transferred isn't a readmittance, it's excluded. You have to have a space between the discharge and the admit in order to be considered a readmission.

I don't think that we -- so rehab is in there. And I don't think we excluded it specifically.

CO-CHAIR LAZAR: Just so I understand, so a patient who is transferred from, or discharged from an acute care hospital and admitted to a rehab facility, which is bureaucratically the way you have to do it, is considered a readmit.

MR. STETTLER: As long as it was done on the same day as discharge, it would not be a readmit.

CO-CHAIR LAZAR: Okay. It would not be a readmit. Okay. So I think that was the question. Okay, that's helpful, then.

MEMBER FOLTZ: But patients who went to a nursing home and then came back when they were ready for a rehab admit, would be --

MR. STETTLER: That would be a readmit.

MEMBER FOLTZ: Okay.

MR. STETTLER: As long as it's within 30 days.

MEMBER FOLTZ: Yes, we're being consistent with that.

CO-CHAIR LAZAR: Karen? Karen, did you want to --

MS. PACE: I'll wait until --

CO-CHAIR LAZAR: Okay, Bruce and then Richard, and then Tanya, and then Warren.

MEMBER HALL: Well, I raised my card to slow down the discussion, because I'm not that comfortable with the sophistication of the risk adjustment approach. I think I understand it, and I know Larry was asking about whether this was really direct or indirect standardization, and whatnot.

And I'm just going to speak bluntly. Frankly, it seems to me that this is a more crude response than we've been

presented with earlier this morning, and that because the discussion was more detailed earlier this morning, I think we were discussing at a different level of rigor.

As I understand, you've created several hundred categories, interacting age and procedures and diagnoses and whatnot, and then calculated a rate for those categories, and then I think you're crediting each institution with the appropriate rate based on case.

But it's not a hierarchical approach, and this morning we were talking about some relative merits of hierarchical approach, in which we were actually kind of condensing hierarchical approach with the aspects of shrinkage, which we probably shouldn't have done. But this does not sort of meet what I consider current standards of treating clustered data in a clustered fashion.

So I want to slow down this

discussion, and say I don't feel like we're holding this to the same level of rigor that we did this morning, and I would like to say I feel this is a less informed -- I feel like we're looking at a less informed, insightful approach.

CO-CHAIR LAZAR: Well, I want to be careful. Because the volume of discussion does not necessarily reflect the assessment of the measure. It could be that people have come to some kind of impression one way or the other.

MEMBER HALL: I agree.

CO-CHAIR LAZAR: I just want to be careful here that we don't --

MEMBER HALL: And for that express reason, I didn't want discussion to only be 10 seconds, and create an impression that everybody was, in fact, comfortable.

CO-CHAIR LAZAR: Richard?

MEMBER BANKOWITZ: Yes, I wanted to, along the same lines, just make sure that

I understand correctly when you use the term risk adjustment. It may be semantic, but it really is a case mix adjustment. We're adjusting to discharge -- diagnosis or procedure, right?

MR. STETTLER: Diagnostic, procedure, and --

MEMBER BANKOWITZ: But in terms of the co-morbidities, planned versus unplanned, other status -- other factors are not going to be in, by definition, those boxes, right? It's just age and discharge.

MR. STETTLER: Our understanding of all causes, it should be all causes. So we want to include all admissions.

MEMBER BANKOWITZ: No, I understand this. So I guess, to turn this back to validity, base validity is one aspect of validity, and we've had this in use. What have been the responses? What have been the -- what's the impression of the providers?

MR. STETTLER: We have -- I can't

remember the exact number, but we have a few dozen contracts signed that use this methodology in a performance-based construct. And we are very keen to make sure that it is linked with risk-adjustment, because we think that's important to balance it. We don't want people to start too early or late, and we have it be stable. So I think it's a good balance.

So we have it -- again, it's improving our internal readmissions, and we have convinced hospitals to contract with us based on this methodology. Whether or not -- I don't know how to -- obviously, we contract with thousands of hospitals, so it's a pretty small number at this point.

MEMBER BANKOWITZ: Okay, not to make too fine a point of it, but secular trends are decreasing. So for example, we can't just look at one population before and after, and say "Well, we've impacted the readmission." Or do you have any other external data to suggest that this actually

was important in reducing the readmission rate?

MR. STETTLER: We can't control for the secular variables, as far as -- you know, I don't think --

MR. SCANDRETT: Yes, I mean, it also depends on the interventions. You try and target it. I mean, this can be used to identify facilities that may be not performing. But then, once you've identified them, there's no sort of cookie-cutter solution to say "This is what we have to do for all these facilities."

It's really just a matter of using this tool to identify places where there may be room for improvement, and then you have to look at the specific case, to sort of say "How do we improve things?" So it's part of the package that we're doing.

CO-CHAIR LAZAR: Okay. Tanya?

MEMBER ALTERAS: This is a really dumb question, so I apologize, but on page one

of the Measure Submission Form, it says at the bottom "Note to committee: the measure includes planned and unplanned readmissions," but then it doesn't say it anywhere else. I'm just curious, why is it including planned? Or how does it include planned, in the way --

MR. STETTLER: Our basic -- again, we want to make the method be claims-based and dependent. Doing matched pairs, in our opinion, is a very difficult and time-consuming and almost impossible thing to do for individual facilities. Again, we wanted to include all cause, all readmissions. There's just no good way to determine what's a planned readmit on the claim, as far as our experience goes, that's reliable enough to be used in this context.

CO-CHAIR LAZAR: Okay. Laurent?

MEMBER GLANCE: At the risk of being redundant, I just -- my impression in looking at this particular measure is that it really doesn't level the playing field. There

essentially is no adjustment for comorbidities other than age. And with age, the patients are divided up into two age strata, which is incredibly crude.

So that is my major concern with this measure is: there is extremely limited risk adjustment.

CO-CHAIR LAZAR: Brent and then Ashish. And Bruce, are you still up, or is that from before? Great, thank you.

MEMBER ASPLIN: I'm just following this through. So essentially, just primary diagnosis and primary procedures. So Patient A, both had a STEMI and both had angioplasty, or something, but Patient A also had COPD, diabetes and major depression, and Patient B didn't. They would both be in exactly the same weighted risk readmission category. I'm just playing out what Laurent was saying.

MR. STETTLER: Again, we wanted to make it based on the individual hospital admission. And the secondary diagnoses, we

think, are not needed for the measure. If you want to do comorbidities, you should take a historical timeframe, pull all claims for the member, and identify every criteria they have. That's the only way to be fair in doing those comorbidities adjustments. If you only base it on the claim, you run the risk of being inaccurate.

CO-CHAIR LAZAR: Yes, I just want to be careful that we don't get into a debate about this. We're simply trying to clarify what the methods are, and then the committee members simply have to opine on them. So I just urge my colleagues to have caution. Ashish?

MEMBER JHA: I'm not going to get into a debate. I just want to understand and maybe ask a question about it. It seems absolutely right that if you get claims just from the hospital, it's going to be incomplete and it's not going to be perfect.

But I would think, based on the

other modeling data that are out there, that you get better than not doing it at all. And I guess the question is, can you talk a little bit more about your decision to opt out of that?

And just, since we were exploring options, is that a fixable issue? Is that something that you guys can consider going back and seeing if you could do this? And I realize that's not necessarily on the table. We're not here to recommend how you should build your measure.

But we're just trying to understand the thinking behind saying "Well, it's not perfect, therefore we're going to opt out of doing it at all."

MR. STETTLER: No, I think it's definitely something we'd consider. Again, I think it comes back to a balance between usability, feasibility, validity, and reliability. I think there's got to be some kind of balance. If you make it too robust or

too refined, then we hear "Well, two years old data isn't good enough. We can't wait that long. We need to see the other admission." And then nobody uses it, right?

And then on ours, we don't have enough of that, so we don't want to use that one either. So maybe there does have to be some middle ground here that we can agree to and get a model that works and is feasible.

CO-CHAIR LAZAR: Mark?

MEMBER SCHUSTER: Yes --

DR. BURSTIN: Put your mic on.

MEMBER SCHUSTER: Sorry. I'm so used to thinking about removing chemotherapy as an example, as an effort to sort of get at planned readmissions, is that something that you guys tried in the earlier version of it? Did you try it?

And we'll get to usability later, but has there been any push-back in the existing measure for not having removed -- or, we can never remove all planned readmissions,

of course, but to make an effort to remove some of the more obvious ones, I guess?

MR. STETTLER: Yes, there has been some push-back, absolutely. I think that, if you think about the way the factors work, the factors by default already include adjusting for planned readmits. The chemotherapy risk adjuster is very high, so there's a very high likelihood of readmit if you have a chemotherapy admission.

We have had some facilities who want to stratify and isolate out portions of the readmissions: transplants, chemo, and some of the peds and NICUs and things like that. So we have had some push-back there, but the nice thing about the metric is, you can simply, you know, choose to stratify the population and exclude it from the metric on a go-forward basis if you need to.

So it is flexible in that way. To completely exclude it and not even have it included as a way, as an ability to use,

means, I think, that a cancer hospital can't be compared to their peers, because we don't provide weights for the cancer hospital's admissions. So we just think it's a good idea to include them all at the initial outset.

CO-CHAIR LAZAR: Leslie?

MEMBER KELLY HALL: Just for clarification, you've only tested this on hospitals that have contracted with you? Is that what you just said?

MR. STETTLER: Correct, yes.

MEMBER KELLY HALL: And so it's only your data --

MR. SCANDRETT: Well, they have to be contracted with us to --

MR. STETTLER: Well, we have non-par data in as well. But the vast majority is going to be participant.

MEMBER KELLY HALL: And is there anything proprietary about the data that you're collecting, versus the claims data that would be collected by CMS?

MR. STETTLER: As far as we know, no. CMS has done a really good job of helping us standardize the data submissions, so the data coming in on the claims look very similar. We used a five percent sample, and no reason you couldn't do that, but we don't think that there's anything in our method that would be non-standard, that you couldn't deal with on anybody's claims data.

MEMBER KELLY HALL: Thank you.

CO-CHAIR KAPLAN: I just have one quick question for you. Somebody like me -- I'm a psychometrician, so my use of the term "factor" means a little bit different than it sounds like you're using. So by "factor," do you mean a variable? Do you mean a composite? Is there some method you use to approach the derivation of those whatever-they-ares?

MR. STETTLER: Yes, it goes back to the calculation of the risk factor for the readmit, into those 220 groups. And that's basically what I'm referring to. So it's: how

often do admits that look like this, that have this clinical grouping, get readmitted? And that's how we create the factor in the first place, and then we apply it to the hold-out sample to see how well it predicted it.

CO-CHAIR KAPLAN: In terms of what we call sampling from the domain of observables, what exactly -- if those clusters vary, those clusters of characteristics vary, how confident are you that those clusters reflect the same overall construct?

MR. SCANDRETT: Not sure I follow. So, in terms of the clusters of each of the condition buckets, is that what we're talking about?

CO-CHAIR KAPLAN: You collectivized, in my understanding, 200 or some odd number of buckets.

MR. SCANDRETT: Yes.

CO-CHAIR KAPLAN: So that, in order to sort of adjust away whatever it is you think is an unfair contribution, that's

not a reflection of quality of care, to that hospital's readmission, if you're going to vary the clusters, similar to sampling different questions for a math test, they have to at least all measure math.

MR. SCANDRETT: Yes.

CO-CHAIR KAPLAN: So to what extent do you think that that -- and have you tested that that kind of clustering isn't measuring something that you didn't intend to measure for one hospital, and something very different for another hospital?

MR. SCANDRETT: We have not tested that.

CO-CHAIR LAZAR: Okay. No further discussion or questions? Bruce, you had one more?

MEMBER HALL: I just want to know, quickly -- so again, here, there's no SES or other resource consideration with respect to the population, the community. That's also, again, completely absent from these

discussions.

CO-CHAIR LAZAR: Okay. Karen?

MS. PACE: Again, I just want to make a couple clarifications for the steering committee.

Although this measure is an update or change to a previously endorsed measure, as Alexis pointed out earlier, our current process of endorsement maintenance is that whether a measure was endorsed previously or is coming in new, it's held to the criteria that currently exist.

So you know, we have more guidance on reliability, validity, and looking at the risk adjustment methodology. Just to make sure that the committee's clear on what was presented, because it looks like some of the risk adjustment testing was actually put into the reliability and validity, so I'm not clear exactly what the reliability and validity testing were, but just to make sure that the steering committee has all that cleared in

terms of what gets presented.

CO-CHAIR LAZAR: Understood. Adeela, I think we're ready to vote. So this is the scientific acceptability of the measure properties, and I think we read through the criteria before. And the question is, was the criteria in scientific acceptability of the measure properties met?

MS. ADEELA KHAN: So you guys can go ahead and vote yes or no. Nicole, is Patricia back online yet?

OPERATOR: Not yet.

MS. ADEELA KHAN: Okay, thank you. So we have 18 nos, zero yeses.

CO-CHAIR LAZAR: Okay, Taroon, would you like to tell us where we are?

MR. AMIN: So the normal process here is to give a little bit of a rationale on the vote. What I'll do to try to move things along -- not to be complete, but at least to signal to the measure developers and for our own internal reports, as we sort of move

forward with this project, is to summarize a little bit at a high level what some of the major concerns were.

Some of the concerns were raised around the age range, the dealing with the planned readmissions, additional details about the risk adjustment model would have been warranted, questions about the reliability and validity testing, and SES considerations which were similar to previous measures.

So, as part of our procedure, if the measure does not pass scientific acceptability, it would not move on, move forward. And with the clear consensus here, we would not move on with this measure. So I believe we're at public comment.

CO-CHAIR LAZAR: So before we go to public comment, is there any rationale that a steering committee member would like to be put on the record supporting the no vote that Taroon has not already expressed? My sense is he gave us a pretty comprehensive list of the

comments that he heard, and I didn't hear anything that -- well, there's nothing I heard earlier that I think was missed, but I just want to give the committee an opportunity to express that, if somebody has a very strong opinion that was not captured.

(No response.)

CO-CHAIR LAZAR: Hearing none, or seeing none --

CO-CHAIR KAPLAN: Well, I have one comment, which is, in general, I thought that there wasn't enough -- we weren't given enough data. And so I think I heard a lot of questions that suggest, in the absence of data, we simply don't know. And that's one of the criteria for our vote. So if that is informative, great, if it's not --

MR. STETTLER: The one thing I would guess that the developers -- and I'm not what the other developers would say, but we were given three weeks warning to submit this. And again, obviously we should have had more

of our ducks in a row and so on, but I thought the process was very challenging, especially when you have a business to run. It's not as simple to get this stuff out as you would think.

And I know we had a lot more information we would have liked to have presented, but we did not have the time to do it. So I would say in the future, if you can give your developers just a little bit more time to actually prep for it, that would be very beneficial.

MR. AMIN: I mean, this is the first project NQF is doing which is an expedited review. We recognize that the timelines are very short for everybody involved in this project, including the staff, the developers, and the committee members who probably would have liked a little more time to review the measures.

We respect that concern. We'll take that back, and as we review further

measures for potential expedited review we'll consider that as well. And other measure developers had a similar timeline as the measure developers here.

CO-CHAIR LAZAR: Taroon, do we need to go for public comment?

MR. AMIN: Yes, so we'll ask Nicole to open up the lines for any public or member comment.

OPERATOR: Certainly. For public comment, please press *1 at this time.

MR. AMIN: And if there's anyone in the room who would like to address the committee?

(No response.)

CO-CHAIR LAZAR: Okay. We're scheduled for a break, but is everybody comfortable just working through? Okay, so why don't we do that?

MR. AMIN: Our measure developer needs five minutes.

CO-CHAIR LAZAR: Five minute

break, but only five minutes.

(Whereupon, the above-entitled matter went off the record at 3:11 p.m., and resumed at 3:19 p.m.)

CO-CHAIR LAZAR: Could I ask everybody to grab their places, and we are going to start with our third measure, and we are looking at 1768, Plan All-Cause Readmissions, from the National Committee for Quality Assurance.

So why don't we begin with the importance to measure. I am hoping, as the third go-round on this one, we can dispense with this rather quickly, and then get directly to the issues of scientific acceptability.

Does anybody have any comments or questions they would like to make around the issue of importance to measure. Is the threshold criterion, importance to measure and report, met? Any comments or questions about that? I think we have discussed this quite a

bit twice before. Are we ready to take a vote?

MS. ADEELA KHAN: Okay. On importance to measure, again high impact, performance gap evidence. Was the threshold criterion, importance to measure, met? You have a minute.

CO-CHAIR LAZAR: Feels like we have jeopardy music in the background

MS. ADEELA KHAN: Okay, we have 18 for Yes, and zero for No.

CO-CHAIR LAZAR: Okay. Why don't we move into the issue of scientific acceptability, and perhaps you could just -- Oh, I am sorry. Would you like to introduce yourselves?

MR. SAUNDERS: We would. So my name is Robert Saunders. I am a research scientist at NCQA. I am the technical lead for the development of this, and I have with me Jerry Gottlich.

MR. GOTTLICH: I am a senior

health care analyst at NCQA who helped develop the measure logic with Robert.

MR. REHM: This is Bob Rehm, the Assistant Vice President for NCQA for oversight of our measure.

MS. ALAYON: This is Dawn Alayon. I am a senior health care analyst. I manage all of the NQF measures for NCQA.

MR. SAUNDERS: I didn't know if we are still doing the three to five-minute comment, if that is still permissible.

CO-CHAIR LAZAR: Sure, absolutely.

MR. SAUNDERS: So I think we have had a very informative and enlightening conversation this morning as we have gone through the two previous measures here. I think, as we transition into our measure, one thing we wanted to highlight is the fact that, in the preceding measures, there is a hospital based all-cause measure, and the unit of accountability is the hospital.

As we look at the measure in this

round, we are an organization that is about monitoring health plan as the unit of accountability here. So one of the first things we wanted to point out is that there are hospitals that are the unit of analysis within the health plan, but the measures are aggregated across health plans. So multiple hospitals are reporting to health plans.

We think that this is really an advantage of this measure, because it starts to move us up from the hospital as the unit of accountability to a more population based approach.

I think some of our earlier conversations about who is accountable for this focus on that there are issues related to transitions of care, and we think that health plans are certainly a key component to doing that. They are also key to understanding the totality of services that may happen to prevent readmissions. So we think that a plan based approach is an important distinction to

make.

I think the other thing Taroon has mentioned at the beginning is that we should clarify some of the information that we have presented.

In terms of the supplemental information is that our measure development process has gone on for over three years. So we started the initial testing and development of this measure back in 2009. Our initial testing with commercial and Medicare Advantage plan based data was in 2010 using 2008 to 2009 data.

We have since gone into the field. So we have collected first year measurement from health plans to Medicare Advantage plans and commercial health plans. Those data are already in use at CMS, for the commercial side, for the 65 and older population, and we are in the process of -- At the time that this call came out, we were in the process of updating our models for our second year of

implementation.

So I think the initial tables that may have been in there might have reflected the first year regression weights for conditions, and then we now have data from 2008 through 2010 for Medicare weights, and so we have also modified the risk adjustment model to have separate risk adjusters and weights for the Medicare under 65 and the Medicare 65 and older population.

I think that would be it, I think, in terms of the clarifications about the measure and the material that you have. Thank you.

CO-CHAIR LAZAR: Thank you. Bruce?

MEMBER HALL: Could I immediately just ask you to clarify your last statement. So are you saying that there is information that we are not looking at which you would like us to consider?

MR. SAUNDERS: So we had submitted

separately an additional package with regression coefficients for the current specification model for a second year of data collection. I am not sure where that is in terms of the packeting.

MR. AMIN: That information was sent to the Committee subsequent to the original measure specification that you see. So the risk weights as referred to were --

MEMBER HALL: Okay, great. I didn't know if you were -- if there are other changes to the approach that are not incorporated here.

MR. SAUNDERS: No.

CO-CHAIR LAZAR: Okay. Should we open the discussion around the issue of scientific acceptability?

MEMBER BANKOWITZ: Can I ask a quick process -- or just a quick question to the developer. The measure type is listed as a process measure. Was this measure intended to be a process measure or an outcome measure,

because that would the difference in the evaluation?

MR. SAUNDERS: I think, given the conversation we have had here, I think outcome is appropriate. We were torn. We see the readmission as a process. Naturally, that occurs within the system, but I think the way we have talked about the measure conceptually, I think we would revise our answer to be outcome. We think it is consistent with that.

CO-CHAIR LAZAR: Ashish?

MEMBER JHA: I have just got a couple of clarifying questions. So this is a health plan level readmission rate, not a hospital level readmission rate?

MR. SAUNDERS: That is correct.

MEMBER JHA: Okay. And so I realize we are going to evaluate it based on that. Just thinking about sort of what we talked about in the morning around the sort of Congressional requests or requirements for developing a measure, that is really at the

hospital level.

MR. SAUNDERS: Correct.

MEMBER JHA: So this measure would not meet those needs, but we can still evaluate it for its validity, etcetera, etcetera. Do I understand that?

CO-CHAIR LAZAR: Yes, exactly right. Helen, clarified that for us offline.

MEMBER JHA: Okay. Can I just ask as a follow-on -- I sort of feel like I keep doing this, which is keep asking what else could you do with this measure, which is not the purpose. But could you potentially develop a hospital level readmission rate, given that you do have for the patients who are in the plans that you guys are covering or that are participating -- you have all the data. Could you do that?

I am not even sure why I am asking, outside of you probably know why I am asking, but --

MR. SAUNDERS: So I think that the

short answer is probably not. Given the way the data -- this just sort of speaks to sort of how we collect our measures.

So we are sort of a hands-off organization in terms of the data collection. Plans take our specifications and implement them either themselves or through their software vendors that calculate how many hospitalizations, all the transfers and so on.

They then submit to us the aggregated performance information, the numerators, the denominators, those expected rates. So we never see the data to know which hospitals are performing well within the health plan.

The hospital -- or the health plan itself certainly has that data, and they are in a position to do that, plus a variety of other investigational things, the way the measure is constructed having the CCs for the risk adjustment, categories for the index condition. They can certainly look at whether

psychiatric hospitalizations have a higher readmission -- propensity to readmit than other kinds of hospitalizations, and so on.

We treat that as a health plan responsibility. Our job here is to create a standard metric for quality monitoring and accountability of the health plan, and then leave to the health plan to work with its network of hospitals and providers or medical homes and other entities to figure out what quality improvement strategies will work, and help move that metric.

CO-CHAIR LAZAR: Jeff.

MEMBER GREENWALD: I appreciate the sort of different approach that you took relative to the prior two. I think it has some interesting implications. I wonder if you could comment, however, as to why we would want that to be the measure that we would endorse, given there are some practical limitations to having health plan level data in terms of -- and this is perhaps jumping a

little bit ahead in terms of usability, but conceptually and philosophically why we would want a health plan level approach, if we are going to, hopefully, find something that is intervenable.

MR. SAUNDERS: Absolutely. so I think the critical issue is, given that we were in a context of thinking in terms of 30-day readmission rates -- there are a variety of other ways you could go about measuring readmissions and thinking about that problem. There are certainly people that are going beyond that now that think about what that is, but if we start from the premise that 30-day readmission rates is what you are after, one might think that the accountability function might decay as you get further from the discharge date.

So if we are thinking about the kinds of things that happen within the first five to seven days of a hospital discharge are likely to be more in the hands of the

hospital, there is probably a role for primary care. There is probably a role for the health plan. There is probably a role for a variety of other people. But I think the underlying premise for most of the hospital readmission rates is the hospital is the unit to have.

As you get further out, in 14 days and 30 days, the health plan is in a greater position to act to deal with the coordination issues between primary care, the coordination between the nurse care manager who is -- or whoever is going to follow up about did you make your appointment, are you taking your medications in the right way, are you having problems getting your medications, those types of issues.

So we think that the health plan base measure is a complement to hospital based measurement. Certainly, we don't think of it as a substitute. We certainly use the Yale approach to the condition specific measures which were already in place and already

endorsed as our initial model for testing of this measure, but we look also at the United model and the DRE based risk adjustment.

We certainly -- You know, standing on the shoulders of people that have been very successful at this and looked at endorsed measures as their starting point, but we think there is a world separately for the health plan.

What we think sort of more broadly is that we are trying to -- if we are trying to effect improvements in quality of care, that you can come at it from multiple angles, that it is not simply a measurement issue of readmissions and is it hospital for readmissions; is it plan? There are a variety of other metrics that would be important to look at that would, hopefully, relate to any of our readmission measures.

So we had care transition measures. Linking that with service support and validity of the measures, that, we think,

is a field where to that stage adds in the development of readmissions, but we certainly, as a measure developer developing things in a variety of these areas, as we accumulate, think that those relate.

CO-CHAIR KAPLAN: I would like a little clarification, because I wasn't privy to the conversation that Eliot and Helen had. But for the purposes of those of us who are trying to serve -- We don't want to be comparing apples and airplanes here. We want to really sort of stick with fruit.

So if we are comparing the purpose and utility and validity and scientific rigor of a measure for one purpose, and we are trying to transport it to another purpose, and usability, certainly feasibility, and making that leap for some of us, I think I am getting a little bit confused about -- I can understand the unit of analysis here, but I am sort of trying to understand the unit of reporting, and to what extent are we

evaluating apples and airplanes?

DR. BURSTIN: Yes. So we did a broad call for all-cause readmissions measures, not just for the specific hospital program they have talked about. So we saw these as potentially complementary, although I think one question was raised is whether one of the other measures at the hospital level will be aggregated up to health plan. So those could be potentially competing.

So that is the way we have been thinking about it. I would, again, evaluate the measure on its merits, and we will return, I think, to those issues to follow.

CO-CHAIR KAPLAN: So it is merits for --

DR. BURSTIN: Endorsement as a national consensus center is all-cause readmission, and in this case for health plans.

CO-CHAIR LAZAR: More discussion?

MEMBER GLANCE: Just a quick

comment on that comment. I think, as we certainly consider harmonizing these measures, it might make a lot of sense to pick the best of the best, because two different risk adjustment models -- I mean, it is not that difficult to take these models and aggregate them either at the hospital level or at the plan level. If we get one model that does it at the plan level and one model with those at the hospital level, and these are different models that have different risk factors, it may end up giving us somewhat conflicting information which may not be very useful for benchmarking purposes.

MEMBER JHA: So the way I am reading it is the risk adjustment, not the hierarchical shrinkage issue, but the risk adjustment seems to me very, very similar between what you guys have at the plan level and, let's say, what the Yale-CMS group had at the hospital level. Is that right? Are they identical? Are they very close?

MR. SAUNDERS: They are not identical, but it was our intention to mimic as much as possible their work.

MEMBER JHA: I couldn't tell a difference, but maybe there is.

MR. SAUNDERS: I am sure that we will be able to tell you the difference.

CO-CHAIR LAZAR: Is there anymore discussion?

MEMBER GLANCE: Well, I was going to make a comment on that comment. The Yale group has got five different models, and this is one model. So the coefficients are going to have to be different, I would think.

MEMBER JHA: My understanding is that the Yale group has one model, meaning one set of covariates. They let the coefficients vary across the five models.

So the coefficients in this will surely vary, but that is because at the plan level different things matter to a different level in terms of adjustment, but as long as

the same covariates are in, I feel like the signaling value to everybody about here are the things that matter for risk adjustment gets achieved, which I think was sort of your goal. Right?

MEMBER GLANCE: Yes. I think that, if you have five different models, although the covariates are going to be the same, the coefficients are going to be different. So you are probably going to see different results based on whether you are using their model or the Yale model when you are looking at the same unit of analysis, and I think that is where things may get a little confusing for benchmarking purposes.

CO-CHAIR LAZAR: Further discussion? Frank?

MEMBER GHINASSI: Just a question, again to be clear about what we are voting on. We are not voting yet, but the way it reads it says "results about the quality of care." This is at the health plan level. So we are

not really talking about the quality of care directly anymore. We are talking now about the coordination of care by a payer within a given market.

Is that what we are -- I just want to make sure I understand. That is what this sort of measures. Right? How a payer coordinates care through incentivizing interaction, care management, on and on, within a market or within the reach of their payee group. Is that what we are now voting on or considering? Right?

DR. BURSTIN: Care, though -- I mean, care management programs, that is care.

MEMBER GHINASSI: Those are all debatable. I would think of it less as care than I would the coordination of others' care, but you could argue both.

MS. PACE: Well, it is quality of care for the enrollees in a health plan, which could be done by a few number of hospitals or lots of hospitals, depending on who that

health plan contracts with. But it still is, I think, supposed to be indicative of quality of care. It is just that you are looking at it across a health plan now instead of at individual hospitals.

MEMBER GHINASSI: And so the data would be useful then -- Again, I am just trying to think about the validity of using it. It would be useful first to the plan in how they would go about setting their quality improvement methods, and then we would be relying on those plans to make that useful by the individual institutions?

MR. SAUNDERS: That would be our expectation for it, that the health plan would know the information about who they are contracting with, and they would have detail in their own data system to identify which hospitals are responsible or which individual providers are potentially contributing to that.

CO-CHAIR LAZAR: Christine, then

Leslie, then Tanya.

MEMBER TRAVIS: I want to be sure that I understood the last discussion. The way I looked at this was that it was really looking -- because it is a health plan measure, that it is looking at the underlying contracting and network provisions through the hospitals as well as those services and programs that health plans may lay on top of that to manage the population.

So, to me, it is still looking at the quality of care provided in the hospital arena, but it does add that additional layer of looking at what the health plan itself does to manage the population within it.

CO-CHAIR LAZAR: Leslie?

MEMBER KELLY HALL: And just to clarify, this would be just for those covered members within your system, and so would be then further stratified to cover people in your system. Correct?

MR. SAUNDERS: That is correct.

So our measure is focused on Medicare Advantage enrollees, and we have a separate model for the under 65 population, the non-elderly disabled, Medicare 65 and older, and then the commercial 18 to 64 population. Yes, they would all be treated separately. It is contingent upon having one of those class of coverage.

MEMBER KELLY HALL: Thank you.

CO-CHAIR LAZAR: Tanya.

MEMBER ALTERAS: This gets more into the usability discussions. So I will save this for later, but I think this would be hugely useful for purchasers and for consumers, and especially with the implementation of the health insurance exchanges, and having quality information to consumers at this level, I think, is just going to be enormously meaningful.

CO-CHAIR LAZAR: Richard?

MEMBER BANKOWITZ: So I am trying to put myself in the position of a plan

administrator. So this is going to be an exercise in imagination here for me.

It would seem to me that I would be worried that, if you are going to compare my plan to Laurent's plan here, I would want to know that my service mix of patients was about the same, so that I had the same number of conflicts, chronic patients, and the same number of patients, chronic disease, and the same number of sort of well patients. But I don't think in your model you accounted for any of that. Is that true? And if not, then why not?

MR. SAUNDERS: So we have -- To the extent that we have accounted for that, we have accounted for it in -- So we are adjusting for the patient attributes of care or other patient attributes, demographic in particular, that are likely to contribute to readmission. So we have the age. We have gender. We have presence of major surgery. We have the index condition, so why you were

in the hospital in the first place, and then your past year's worth of comorbidities.

So to the extent that we are comparing plans, we are adjusting for that set of attributes.

MEMBER BANKOWITZ: And what about looking at the specific hospitals, because I would be -- I mean, again as a plan, I would be concerned. Well, I've got this one rogue hospital that I just can't control, but I've got 99 that are doing real well, but I am going to be dinged because we've got some differences of opinion about how to manage. Did you look at that?

MR. SAUNDERS: For the data we have, certainly, the data submitted to us by the health plans in the first year of submissions, we don't see that level of detail. In the development datasets, we have the two data warehouses we use for Medicare data, we did not investigate that.

That is, I think, more strictly

sort of an NCQA sort of thinking bout the problem as the health plan has to decide how to handle the rogue hospitals or rogue providers. They have the option to contract with that hospital or they have the option to take some other corrective action, if it is through some quality improvement area.

Eventually, we are going to target resources doing that.

MEMBER BANKOWITZ: Last question: Did you consider looking at admissions in addition to readmissions, because I am thinking of, if I have 100 asthma patients, and so does Laurent have 100? I might admit quite a few, but not readmit any, and there might be three chronic patients that tend to get readmitted, and his look worse than mine even though I might be really using much more of a hospital than he was.

MR. SAUNDERS: We definitely think -- We did not specifically task that, but we definitely understand that the readmissions

issue is not an independent entity, that there is a whole complex issue of awardable hospitalizations generally, and the hospitalization we are going to award is the index hospitalization as much as the readmission hospitalization.

I think we have a series of resource use measures that focus on hospitalization. We are looking at risk adjustment and specific clinical condition, diabetes, cardiovascular care, that are coming online this year.

So we feel like that is measured in other aspects of our portfolio. That portfolio is not treated as part of the scope here.

CO-CHAIR LAZAR: Sherrie?

CO-CHAIR KAPLAN: I have a sort of follow-on question to that. To the extent that you choose hospitals, hopefully, that provide good care for your health plan recipients or clients, the selection bias

seems to me like a real concern here.

The other issue we had a rather lengthy discussion about here is the variability in the precision of your estimates by hospital.

So, one, can you talk to us a little bit about selection bias and what that does from an NCQA standpoint, and the other thing is, for the precision of estimates, how are you taking that into account in your potential for feeding back to the hospitals about their readmission rates?

MR. SAUNDERS: If I might ask you to sort of go a little bit further on the selection issue. I want to make sure I am answering correctly. Selection for an individual patient looking to select a hospital or a selection issue for comparison between health plans?

CO-CHAIR KAPLAN: Plans, hopefully, channel their patients to hospitals that provide good quality and efficiency, but

quality is one of the hopeful criteria that they channel their patients to.

So plans are then selecting hospitals for patients that provide high quality care. We know that there is a volume/outcomes relationship. Hopefully, the plan is thinking about that, too.

So I am worried a little bit that the selection bias here when you are talking about providing this at the hospital level or feeding back to hospitals information, that that gets taken into account along with the variability, potential variability, in that information at the hospital level.

MR. SAUNDERS: I think what I will probably fall back on may not be satisfactory, but I think for us it is for the health plan, that is who we are feeding the information back to or we are feeding back to the consumer the information about what that health plan's performance is on handling readmissions, and it is both a measure of the

quality of care provided by that health plan and is both a measure of accessibility, services, that we would not -- not thinking specifically about how that would feed back directly to the hospital.

We think that is a conversation between a hospital and the health plan, and we think that they would have those discussions and would certainly make those arguments about whether they are being unfairly penalized in some way.

Let me jump to the precision question. Could you repeat the precision question?

CO-CHAIR KAPLAN: Well, depending on how the plan level data are sampled -- for example, you get n patients per plan to estimate the plan's performance. It gets pretty think, if that number is 400, to the extent that you can give any kind of stratified data.

Again, if your intent is not to

feed back to the individual hospital, I understand, but you sort of said that in your opening remarks, that that was a potential something you could have done.

One, the selection bias with respect to participation in NCQA and, two, you know, how does this hospital readmissions issue play out with respect to your current sampling structure?

MR. SAUNDERS: I am not sure of the specific numbers of the proportion of health plans that have reported to NCQA, but it is quite high. It is high like it is -- You know, we've only got 20 percent of health plans here. I think it is in the neighborhood of 90 percent of health plans. We don't think there is going to be a systematic bias in that form or fashion.

In terms of precision, you are right. We are definitely worried about comparisons on small numbers for -- you know, if we think about special needs plans, and we

will get those individual contracts. Some of these have maybe 100 hospitalizations and, certainly, would be reluctant to make those kinds of comparisons.

We are in the process of figuring out where to set that threshold for public reporting. We are not setting a threshold, though, for submission to us. We are not saying, if you have -- you have to have at least 500 hospitalizations to calculate this and send it to us. We are saying, submit it to us, and if there is a lower end for what is reliable, whether that is 100 or 500 or 2,000, whatever that threshold is, we will set that for reporting.

We think of that as a reporting issue. We certainly understand it as directly influencing the reliability issue for the fairness of the comparisons between health plans, but I think that it may be in the data that we presented, just in terms of the actual denominator size of our health plans, that

almost 75 percent of our Medicare Advantage contracts were able to report at least -- I think it was -- We know that at least 96 percent were able to report, but I think at least 75 percent of our Medicare Advantage health plans had at least 7500 enrollees, so having a decent volume of hospitalizations within that group.

We are not so concerned about small numbers on the commercial side.

CO-CHAIR LAZAR: Ashish, and then Jeff.

MEMBER JHA: Because we torture our Yale colleagues on this, I figure I can't let you guys go without talking about SES.

How have you guys thought about it, given the variations in SES make-up of the different plans? It sounds like you came out against including that in your model, but if you could just sort of give us some of your thinking about that, it would be really helpful.

MR. SAUNDERS: Sure. We are saying that there is likely a relation between SES and performance on this measure. For most NCQA measures, we don't do any risk adjustment at all. We have such a good denominator that we think of the quality issue as, if you don't do this, you are doing something wrong. Readmissions is one of the first of our measures that has kind of gone beyond that, to worry about it.

So as a procedural issue for us, we don't have a really good way to evaluate the SES within the community. The health plans don't report that information to us. There are a variety of things that I would like for them to report as well that are not in there, doesn't seem to be in their capacity to do.

So there is a technical limitation to we can't get the SES data, but I think as a practical issue in terms of the comparisons, you know, this would mean something if you

were comparing a health plan in Minnesota to a health plan in Florida, but those decisions that are being made are on a more local scale.

So if we are comparing the health plans in the Washington Metropolitan area, we don't have any reason to believe in advance that there is a SES difference between the health plans unless they are particularly good at dumping or identifying ways to dump their people into other plans. They are certainly able to deny them-- or used to be able to deny them coverage, I think. We will see how far ACA gets along here, but we don't think that there is likely to be SES differences between health plans.

To the extent that we have the market segmented between Medicare Advantage plans, commercial plans, we did not specify this for Medicaid. There are a variety of issues we could sort of go into about that, but we think that there would be, certainly, homogeneity within those product lines.

The real issue is we don't have the data to do anything about it.

MEMBER JHA: That sounds reasonable. I personally think there probably is a lot of heterogeneity in most plans, but we have had a conversation about why it might be reasonable not to include it anyway, but it doesn't matter. But thank you.

MR. SAUNDERS: And we don't, certainly, have the data to know about it, but the plans are probably jealously guarding that.

CO-CHAIR LAZAR: Okay. Jeff, and then we are going to have to come to a decision about scientific acceptability.

MEMBER GREENWALD: Earlier there was discussion about this concept of the rogue hospital that had sort of outliers in performance. It strikes me that, when you have aggregated data essentially at the plan level, you might have rogue hospitals that would, in fact, not stick out and might get

buried in the data.

I wonder how -- Since you have frequently argued that, on a population health level, that there is a lot that the health plan can do in the health care access and follow-up pieces that you pout out, but the hospitals individually play a big part in that, and if there is a rogue hospital and it does get buried, how does your measure help to allay those concerns?

MR. SAUNDERS: Well, I think, if most of -- Certainly, NCQA is a voluntary measurement process. So I think we are -- That is an issue where we have to rely upon the diligence of the people that are being measured or volunteer to be measured on the issue, that they are looking to ferret out these types of issues.

I think our thinking is that, to the extent that there is a way to distinguish yourself on performance, and whatever incentives it takes to do that, whether it is

an incentive that I want to be a high quality provider on readmissions or a financial incentive to get that performance, that somebody is monitoring the measure and doing the additional quality improvement activities at the health plan level to identify those where, if the administration is kind of asleep at the wheel, then that is a problem for everybody and for any measurement strategy.

I think our -- I will just stop with that.

CO-CHAIR LAZAR: Let me make some sort of summarizing comments and, actually, I am going to borrow a phrase that Sherrie used before, and that is you can't measure weight with a ruler.

I think the point here, at least that I have taken away from this, is that the call was for all-cause readmission measures. Obviously, that did not exclude measures that were not at the hospital level.

So what we have before us is a

measure that is at the plan level, and our task is to decide whether it meets scientific acceptability and feasibility and the other criteria for what it is.

Now we may have issues with it -- maybe; maybe not -- as to whether there should be, as Laurent said before, different methodologies with a measure we might expect to be used at a hospital level versus something at a plan level but, frankly, that is something that probably can be discussed tomorrow in terms of the comparisons across.

So I just want to make sure we don't trap ourselves into thinking about this in terms of answering questions that it really wasn't designed to answer. It is really at a plan level.

I do have, if I may, just one quick closing question, and it may have been discussed on occasions, and I missed it. But does it account for the movement of patients from plan to plan? So is there the equivalent

of a universal PIN or patient identification number or some such thing?

MR. SAUNDERS: It does not.

CO-CHAIR LAZAR: Okay.

CO-CHAIR KAPLAN: I just have one more question. I apologize. I forget your response to Larry's question about the hierarchical modeling, because now you have added another tier to an already complex nested issue. And your decision was?

MR. SAUNDERS: So it is actually even worse than that. It is not even another level above. It would be nice if it was all nested. In fact, it is pretty extremely cross-classified here.

So we have patients going to multiple hospitals, multiple hospitals contracting with multiple health plans. So it would be -- First of all, it would be quite beyond my means and my HLM ability to get the cross-classified model to estimate. I think I could do it, if I had the data.

The second issue is don't put out the data. So if we think about where kind of Yale's model is working off of the 100 percent CMS file, so at the totality of hospitalizations and the totality of admissions and readmissions going on here.

We have a development database that has, in the case of the commercial population, a million hospital admissions. It is nationally -- It contains cases across the nation, but is not the universe of commercial hospitalizations.

Likewise, for Medicare Advantage, we have a database that has about -- the first year it was 22 percent of Medicare Advantage enrollees. For the second year, we have about 45 percent of Medicare Advantage enrollees.

So we are taking a data model that is not the universe here, and then trying to -- It is not the universe of health plans, is not the universe of Medicare Advantage enrollees. So we couldn't come up with an

estimate that would apply to each and every individual health plan coming in.

So what we have to rely upon is the fact that we have a large enough proportion of Medicare Advantage enrollees that we are avoiding serious selection issues, having these half-enrollees represented in this, that we are not doing something crazy untoward in our inferences, and likewise for commercial. We don't think that there is reason to prefer a market scan to the Ingenix warehouse that we use for estimating on the commercial side.

So we are putting the best number that we can, given that we don't see hospital-specific hospital numbers and have the ability to link hospitals to their specific health --

CO-CHAIR LAZAR: Okay. Two last two quick questions, Leslie and then Laurent -- okay, Laurent, and then we are going to call the question.

MEMBER GLANCE: So one of the

questions I asked earlier -- and I didn't quite get the answer that I wanted, and I am going to ask you this question: Did you look at calibration curves when you were evaluating -- when you were validating your model?

The reason I ask this is we really haven't talked about this too much, but I really do think it is very important, is that if you are going to do proper risk adjustment and level the playing field and compare apples to apples, it is really important that you look specifically at the model performance at different levels of risk.

So, for example, if a model systematically under -- or, say, over-estimates mortality -- or not mortality; readmissions, excuse me -- readmissions in high risk patients, and different hospitals have different case mixes, then that model -- the poor model calibration may lead you to make conclusions about hospital quality that are completely unwarranted.

In our last discussion we looked at a summary measure of calibration, and those summary measures can hide an awful lot of information; whereas, calibration graphs really do a great job of showing us how well the model works at various risk levels. So did you look at that?

MR. SAUNDERS: I do not have that available, but I could certainly pull that out. I think, if there is discussion of additional material to put forward later, I think we can put that together, but we did not look at that.

CO-CHAIR LAZAR: Okay. Adeela?

MS. ADEELA KHAN: Okay. So scientific acceptability properties: Are both reliability and validity rated moderate or high? Again, we are looking at 2(a)(1) for specifications, 2(a)(2) testing; 2(b)(1) specification consistent with evidence; 2(b)(2), testing; appropriate methods and scope with adequate results, and threats to

validity are adequately addressed, 2(b)(3) exclusion; 2(b)(4) risk adjustment and stratification; 2(b)(5), meaningful differences; 2(b)(6) comparability of data sources.

Was the criterion scientific acceptability of measure properties met? Vote one for Yes, two for No, and you can start right now.

Patricia, are you online now?

MEMBER McDERMOTT: Yes, I am.

MS. ADEELA KHAN: Did you want to cast your vote?

MEMBER McDERMOTT: I will listen to the response.

MS. ADEELA KHAN: So we have eight for Yes, and 10 No. Sorry, can you repeat that?

CO-CHAIR LAZAR: Well, I think that leaves us in a position where we are essentially stopped for the moment at this point. As we did for the second measure, I

think Taroon and the staff have been collating some of the comments that they have heard, and in terms of rationale. Taroon, I don't know if you are ready to go over that.

MR. AMIN: We have some. Just for the sake of completion, I will hold off on it until tomorrow's discussion when we sort of collate all of the information together.

CO-CHAIR LAZAR: Okay. What we have talked a little bit about is, you know, what to do at this point, and we have one measure that we voted and sort of decided we were going to hold in abeyance, if I read the sentiment correctly. I forget exactly what the numbers are.

We have two other measures where we were sort of stopped at this particular point in the process. We certainly need to have the comparison discussion tomorrow.

Sherrie and I think that a fair bit of the comparison discussion, particularly in view of some of NCQA's comments around a

fundamentally similar methodology to what the Yale folks presented earlier, is what flexibility is there and, based on some of your very detailed recommendations this morning, what tweaks and modifications to the system you heard in depth or detail from Yale, and then essentially here as well, can be made?

I think one of the things I heard Laurent say at the outset of this conversation was concern about approving a methodology at one level, i.e., the health plan level, and then having yet a different methodology at the hospital level.

So if tomorrow some of the concerns that the group have or had about the Yale methodology can be addressed and perhaps fine tuned a little bit to make it a more acceptable model, then we can probably revisit a discussion with NCQA in the hopes of harmonizing the two, with the understanding that you need something at the plan level, and

we, obviously, would like to get something at the hospital level.

Have I been clear about that? I hope I am not too convoluted. I tried to lay it out as crisply as our discussion from before. That would put us in, I guess, the enviable position of finishing early.

What we would do is adjourn now, give the staff some time to talk, possibly with NCQA and certainly with the folks from Yale, so that we could start tomorrow morning with an informed discussion about what flexibility there is in addressing some of the concerns from the group.

CO-CHAIR KAPLAN: I was just going to reiterate what you were going to say anyway, but in the sort of spirit of a fully throated discussion, I think Helen wants us to continue on the last two criteria, since there was a close margin on this criterion.

CO-CHAIR LAZAR: Okay. Yes?

MEMBER SCHUSTER: Could you

explain a little bit about -- I mean, I read the materials about what harmonization means, that in the end there is a single measure or is it NQF has endorsed a combination, but the two entities might still sort of put their measures out there separately?

CO-CHAIR LAZAR: Well, I am going to free associate here a little bit, because I think, you know, NQF has to decide how they want to handle it, but what I think I heard is that there may be need for a plan level measure as put forth by NCQA.

There certainly is a need for a hospital measure, as put forth earlier by the folks from Yale, in that at least some of the members of the committee expressed earlier that, if we are going to have either or both -- if we are going to have both measures, that it would be nice to have a methodology that harmonizes.

So it would be, I am assuming, eventually two separate measures but,

hopefully, two separate measures using very similar methodologies, to the extent that they can be used. Does that sound reasonable?

MS. PACE: Yes. I think, you know, to the extent possible, when we talk about harmonization, we are talking about the measures being as closely similar as possible on definitions, who is included, who is excluded, risk factors, those kinds of things.

Now how close we get on any two measures is a factor related to the data availability, may be special circumstances with a special population, but generally, we would like to have measures be consistent when we need more than one measure.

So as was brought up earlier, you know, sometimes you -- We ultimately would like the broadest applicable measure. So ideally, if you had a measure that could be used at multiple levels, that would be great. Sometimes that is just not possible, and then we try to get measures that will be as

consistent as possible.

CO-CHAIR KAPLAN: I have a follow-up question for Helen and NQF. If you are estimating different levels of performance of a health care system, the idea that you can do one size fits all may be a sketchy assumption.

So to the extent that these risk stratification issues are different, how much -- I am a lumper by training. I do lumping. That is what I do, but this may be a real splitter kind of issue.

MS. PACE: And then it is much more dicey with outcome measures that we are talking about with some methodologies. If you have a straight process measure where there has been a risk adjustment, it is simpler, even though they are terribly difficult, maybe simpler than outcome measures. So each one presents differently.

CO-CHAIR LAZAR: Richard?

MEMBER BANKOWITZ: Is it time to start a discussion of feasibility?

MEMBER FOLTZ: I had one more question. The other thing that I wanted to talk about in the harmonization is, looking at these two, the burden is going to be on the hospital either way, and both, because the health plans will come back to hospitals if they are not performing.

We just don't want it to be a double impact to the hospital. The hospital has to look at it this way, and now this way. So I just want to put that out there.

MEMBER JHA: I guess, given that scientific validity is such a key part of this, and I am a little -- It sort of shows how in tune I am with our group here. I am a little surprised at the vote.

Were we going to get a sort of a synopsis of what the main issues were on scientific validity before we move on to usability, just even like two or three points of what were the main concerns, because I didn't hear them in the discussion as clearly.

MR. AMIN: And I think that was actually one of the -- So I do want to like at least have some time to go through some of the notes, but some of the main issues that were brought up that I heard -- actually, Sherrie made a lot of them, and maybe she can elaborate -- but concerns around selection, potential selection issues; the precision questions around the liability testing of the measure; and potentially additional more information around the measure performance at various levels of risk.

That isn't intended to be the complete list. I mean, I am sure there is more, but that seemed to be the sentiment of the group. So I will leave it at that.

CO-CHAIR LAZAR: Okay. Shall we move on to usability? And again, remember, we are not talking about usability at the hospital level here. We are talking about usability at the plan level, and for the public it would also be at the plan level,

presumably in the region. Rich?

MEMBER BANKOWITZ: So my fundamental concern with this measure would be usability, because at the plan level we really do have a potential to make an impact in terms of consumer choice. Consumers often don't choose the hospital, but they often do choose the plan. Sometimes they can't choose the plan, but oftentimes they can choose a plan.

So I sit here and think, would this actually help me choose a plan? I know it is -- you know, measuring readmissions is involved as a surrogate of coordination of care, but coordination of care goes beyond coordinating those patients who were admitted to a hospital. We would like to coordinate care so patients don't have to get into the hospital, and so they don't have to get into the ED.

To me, it would be more fundamentally useful to know if, let's say, I had asthma or CHF, how good are you at keeping

me out of the hospital? How good are you at keeping me out of the ED? A plan can measure at that level. A hospital can't.

So it just strikes me that this is perhaps not unimportant, but it seems like a very small component of how I would use a plan.

MEMBER JHA: So you are basically saying that you think that readmissions is not that important of a measure at the plan level? The sort of importance of this measure is the fundamental issue as opposed to its fuller or its -- Maybe it is the informational content of the measure.

I am not sure that -- Again, to me as a consumer, it conveys a lot of information, because I would certainly want to know that the plans are comparable in terms of the underlying populations, which we don't know, and I would certainly want to know for my particular condition how good are you overall in coordinating the care, not just

once I get into the hospital, because I really don't want to go to the hospital in the first place. So it is more informational content.

CO-CHAIR LAZAR: Tanya?

MEMBER ALTERAS: Well, I respectfully disagree. I think that, this measure would be reported low. I am assuming it would probably be reported as part of the greater HEDIS measures, which does give you information if you have asthma or CHF or some other chronic disease.

I think, for patients with multiple chronic conditions that really have to make some significant calculations in their mind about cost and just their health status when they are choosing a plan, that having this type of information would be critically important and usable to them, if they do have a choice.

As I mentioned before, you know, in 2014, hopefully, people will have greater choices in health plans through the exchange.

You know, there aren't that many -- There aren't too many measures right now that are really relevant to purchasers, especially private purchasers. I think that private purchasers who are making decisions about contacting a health plan, to find a measure like this is really important as well.

It will raise all sorts of questions for them to be able to talk to health plans about, to talk to brokers about, in terms of what type of care coordination strategies are you using; how are you going to make sure that my premiums are -- you know, those sorts of conversations, as well as giving important information directly.

So I think that the usability, to me, I never questioned that on this measure.

CO-CHAIR KAPLAN: Can I follow up that with you, Tanya? Will Manning, when we were at RAND, did this -- in those days, the old study. If you plan changed by three dollars in terms of the premium, you will dump

it. It doesn't matter, these quality issues. You will change plans on the basis of a three dollar difference in premiums.

So I am wondering, do consumers really get it, the high readmission rates? Consumers might look at that and say, oh, good, if I need to, if they send me back to the hospital, oh, that is a great plan. Do we really know a lot about how consumers use readmission to hospitals?

MEMBER ALTERAS: Well, unfortunately, I think what we do know is that people do look at it as rationing care, if you are trying to reduce readmissions. You are right. That is a big flaw, and there is a greater concern here, which is how do you engage consumers in understanding the quality information that is out there, and we are trying to push for that with all the information about the exchanges coming out.

It is all part of a great continuum, but you are right. My theory is,

if we get the best information out there that, hopefully, people will get engaged.

MR. SAUNDERS: We have done some work on this, not specific to readmissions but in terms of looking at resource use in combination with quality. It is the same type of problem of perception that more is better and indicative of quality, which we know is often not the case.

Judith Edwards' work has sort of looked at the impact of the context of information and how you present it. So if you can provide the information in a qualitative way, provide it in the way that is nuanced to emphasize the direction that you want to be nudging people, that there are ways to present the information to encourage the right type of behavior.

We think that our measure would be capable of doing that.

CO-CHAIR LAZAR: Thank you. Brent -- or let's just go right up the line this

way. Are you done?

MEMBER ASPLIN: Yes. I am with Richard on the usability of this. You know, at the big picture level, we are talking about a three-legged stool, and we haven't even talked about the third leg at all today, which is probably as important as the hospital. It is the ambulatory care system.

I think the hospital and ambulatory care are much more important than the plan. To the extent that plans have historically been important, it is because the ambulatory care has failed in some of these coordination issues.

So just philosophically, I am not sure this as a complementary measure would be all that usable. If it was usable, you would really need to know -- to look at this hierarchical analysis and understand for each -- not only the final outcome, which is plan to plan comparisons, but for each plan look at the hospital level data.

That would be the only way to really make a meaningful conclusion, I think, about not only how the different plans in your market support this work, but now for health partners, what do the different hospitals look like versus -- because they are not apples to apples at all, because it is not a nested analysis. It doesn't have fee for service Medicare. It doesn't have a lot of different populations in it.

So if you have the hospital measures for your market sitting here and the plan measures for your market sitting there, you make a lot of, I think, erroneous conclusions, because they are not nested.

CO-CHAIR LAZAR: Right.

MEMBER ASPLIN: So that, to me, from a usability standpoint doesn't work that well. I am having a hard time getting my head around that part of it.

MR. SAUNDERS: If I can just quickly respond to that, that has not been an

issue that was raised in our any of our public comment by any of our health plans. They are not concerned about those types of comparisons or unfairness.

CO-CHAIR LAZAR: Thank you.

MEMBER ASPLIN: The other biggest thing is then we get this. We get this. So readmissions aren't improving in our market. This measures is saying it is a plan issue, and that is because the plans aren't doing their job. WE have this issue saying that the hospitals aren't doing their job, and we don't get any better. That, to me, is even a bigger philosophical concern.

CO-CHAIR LAZAR: Yes, I concur. That, for me, is an NQF issue, because that is beyond the scope of our charge, I think. But thank you for raising that issue. Frank?

MEMBER GHINASSI: A very short series of ifs. I've only got this halfway formulated myself.

One of the concerns I had about

the previous measures were everything you have already heard about the focus on one part of this complex system, so hospitals, readmissions.

There is something compelling about the focus now on the plan level where, within at least that universe that the plan represents, you've got this compilation of data that includes both the senders and the receivers and this overlay that was described before by Christine, which, I think, has promise.

There have been a relatively finite number, although they are very important. I don't want to minimize this, but there has been a finite number of methodological recommendations that we heard, the specific round, you know, the way that the data has been massaged. You have asked the calibration kinds of questions, I think.

I don't know that any of those are not addressable, and there seems to be

something potentially compelling about finding a way to address those for both of these in a way that allows for the identification possible at the hospital level which locates data where it needs to be for quality improvement activities, and harmonizing that simultaneously with data at the plan level.

The devil is going to be in the details on this, but I am hoping that enough was captured in all the recommendation -- I'm sure it was -- but specifically around trying to remediate these, that there may be a win here if these can be modified in a way to address those to the satisfaction of this. I just don't know if that is doable in a day and a half.

CO-CHAIR KAPLAN: Right. I thought you were headed for a best in class. It is different classes. So, you know, I thought that is where you were headed, but I -- - Respectfully, I think that is for comparability tomorrow as a question.

Christine, and then Leslie, and then let's call it a day.

MEMBER TRAVIS: Yes, I appreciate this. I think usability is usability by whom, and I do think that Tanya brought it up, that from the purchaser's standpoint, their primary relationship and contracting relationship is with the plan.

Therefore, looking at the plan's performance, looking at this measure at the plan level, is a critical component, because, to be quite honest, they don't want to have to get down into the thousands of hospitals that may be in the plan across the country, down to that level of detail.

We have started using readmission measures and Evaluate, which is a health plan performance tool that from the National Business Coalition on Health, and it has just been using some NCQA data that is really just now descriptive data versus really analytical data. But I will say, in the first year of

trying to use this information, it has been very important to the purchasers to say this is a measure of really population management around this one issue.

To Tanya's point, this isn't the only measure we would look at. We would marry this with other measures, but it is something that, I think, employer purchasers especially will hold their plans accountable, because it is more of a population based approach where we layer the plan's programs on top of the hospital's. So this is very important to the purchaser community.

CO-CHAIR KAPLAN: Thanks very much. Leslie?

MEMBER KELLY HALL: I think the question is which population. My concern is that, if a hospital is being measured in one way and then the plan can reflect a different measurement about that hospital, perhaps that hospital with limited resources only attacks those patient populations that are covered in

both entities.

So what happens to those patients that aren't covered in any plan or are underserved or are in a safety net environment? Do they put resources only again where there is high coverage and high visibility and not resources where there is underserved?

CO-CHAIR KAPLAN: Excellent point. Okay, I think it is time to vote. Adeela?

MS. ADEELA KHAN: Okay. On usability, we are looking again: 3(a) equal, understandable and useful for public reporting and accountability; and 3(b) meaningful, understandable and useful for quality improvement.

So was the criteria on usability met? You can go ahead and start voting. We have one high, two moderate, three low, four insufficient.

We are missing one person. So we have three for High, six for Moderate, seven

for Low, and one Insufficient, and Patricia, did you want to cast your vote?

CO-CHAIR KAPLAN: Duly noted. Okay, now we are moving on to the feasibility issue. Feasibility, the criteria are: Clinical data generated during the care process. All data elements are in electronic claims. How susceptible are they to inaccuracies and unintended consequences, and the data collection strategy can be implemented.

Comments? No comments from this group? You are tired. This is fatigue. So if I can ask one quick question. So to the extent that NCQA is a known entity, they clearly can deal with clinical data. What happens with electronic claims? Are you -- There's vendors all over the place in electronic claims.

How confident are you that all the data elements you need to estimate readmissions, including the risk

stratification variables or risk adjustment variables, are in the various bunches of claims vendors that you are dealing with?

MS. SAUNDERS: We are very confident that we -- So as I mentioned before, we have collected our first year data with the health plans. We were able to get 424 Medicare Advantage contracts submitted, 314 commercial health plans. So we know that they are able to calculate and submit that information to us.

How we know that it is of high quality and reliability, first of all, is we have our certified software vendor program. So we have 12 test datasets for the commercial software vendors who are certified by NCQA to come up with can you implement the measure specifications? Are they getting the right calcs out of the test dataset to evaluate, and the implementation of all the programming logic and identification of all the related services, and how they are handling the

primary data, as I said, is transferred from this setting to this setting types of issues.

The second layer of quality control that we have is our auditing process. So we have certified auditors throughout the country that attest to the performance of the health plans and the submissions.

They were key to us in our first year analysis. We weren't sure that they would implement the risk adjustment process correctly, that they would apply the wage or that they would find all of the comorbid conditions and the professional services, that they might just focus only on the hospital sets.

Through our auditing process and our discussions with -- We had focus groups with certain other key auditors around the country, so that there were no problems in terms of implementation by identification of the services or the calculation of the measure.

That gave us some confidence to proceed on with the improvements that we are making in the second year to divide the risk adjustment process for the Medicare population and to have separate plans. We felt that they were capable of doing the work before, that they were able to do this next step.

CO-CHAIR KAPLAN: Thank you. Other questions? Go ahead.

MEMBER LANGBERG: Does the data collection that the -- I will say at the hospital level require CPT codes in addition to ICD-9?

MR. SAUNDERS: It does. They are using CPT/UB codes. It's everything.

MEMBER LANGBERG: So would this represent a different class of data collection and coding?

MR. SAUNDERS: I don't know whether the hospitals do but, hopefully, we will be able to identify that for all the other settings of care that you receive

treatment in. So there is still the opportunity to get those comorbid conditions.

CO-CHAIR KAPLAN: Okay. I think we are ready for a vote. Adeela.

MS. ADEELA KHAN: Okay. So feasibility: 4(a) Do they generate that during care; 4(b) electronic sources; 4(c) susceptibility to inaccuracies, unintended consequences identified; and 4(d) data collection can be implemented.

So to what extent was the criteria on feasibility met? One, high; two, moderate; three, low; four, insufficient information. You can start your vote.

So we have six for High, nine for Moderate, two for Low, and zero for Insufficient, and Patricia, are you online?

CO-CHAIR KAPLAN: Okay. So that wraps it up for today. Right? So for tomorrow -- Oh, we need an overall vote. We need to summarize.

So now we give an overall vote.

Correct? So we are going to review the criteria briefly for an overall vote, and give our response. Wait, Bruce had a question.

MEMBER HALL: I am confused about that, because we have -- for scientific acceptability, we said no, and then we did three and four at Helen's request, but why are we doing an overall vote?

CO-CHAIR KAPLAN: Helen?

DR. BURSTIN: I think it just still wasn't clear where the committee was really landing on this measure. I think it was so close on scientific acceptability. I just wanted some clarity between what we did for CMS and NCQA. That's all. Just you can have your discussion tomorrow.

It is not binding. It is not as if you are saying you recommend the measure. It is just basically does it meet all the criteria.

CO-CHAIR KAPLAN: So the way I understand it, if the usability/feasibility

discussion shaded over your overall vote to move it beyond our discussion of scientific acceptability, then --

DR. BURSTIN: Right, although scientifically it must pass criteria. This was so close that we thought it was worth at least finishing up the analysis.

MS. PACE: No, I was just going to say -- I mean just reiterate that. You know, basically, if you felt it wasn't scientifically acceptable, then that should be reflected in your overall vote.

MEMBER LANGBERG: Could you re-read and see what the votes were?

MS. PACE: In terms of overall?

MS. PACE: Or maybe we don't need to do the overall.

DR. BURSTIN: Maybe we just don't do the overall, and we will just deal with it tomorrow. That is fine.

CO-CHAIR KAPLAN: Okay. All right. Everybody is fine with that? We are

not going to give an overall.

DR. BURSTIN: We will summarize the vote.

CO-CHAIR KAPLAN: Okay. So now we have another option, but I think this is the right option.

MR. AMIN: For consistency with the preliminary votes that we had for CMS, we should do the same for NCQA. There is a clear will of the Committee on the UHC measure. So that is just justifying the process there.

What we are doing here is going all the way through. We will put the preliminary vote in, and then we will have the discussion about competing and/or harmonizing, if that is what the Committee feels, tomorrow.

We have -- If you could read off the votes, that might be helpful. We will read them out.

MS. FORMAN MORGAN: Sure, I can read them. For importance, it is 18 Yes; No, zero. For scientific acceptability, it was 8

Yes, 11 No. For feasibility it was three for High, six for Medium, seven for Low, and one for Insufficient. For feasibility, it was six for High, nine for Moderate, two for Low, and none for Insufficient.

CO-CHAIR KAPLAN: Okay, Adeela.

MS. ADEELA KHAN: So if we voting on overall suitability for endorsement: Does the measure meet all the NQF criteria for endorsement? And again, it will depend on competing measures.

So one for Yes, two for No. You can start your vote. I think we have everybody. The vote is six Yes, 11 No.

CO-CHAIR KAPLAN: Okay. So Taroon is going to give us a little bit on tomorrow's agenda, because many of us are very intrigued by the process so far, and have reached some fuzzy, if that is the right word, conclusions about what is going on.

MR. AMIN: So the two measures that were evaluated all the way through the

process -- What we will do is that we will actually have a discussion around the comparisons.

Really, what we evaluate is, first -- actually, if you have the slide. This is the sort of the box on how we think about related and competing had harmonization of different measures. Really, if you are looking at the same target process, which this would be readmissions, hospital, and the same target populations, these measures in theory would be competing. However, based on today's discussion, if the Committee feels strongly that there needs to be different measures for the health plan versus the hospital, and also having discussion with both measure developers on whether the Yale measure -- I mean the CMS measure could be rolled up to health plans, all those types of discussions will occur tomorrow.

If you do decide that both measures are appropriate, then we will -- Then

we will have the justification for that decision.

The structure of tomorrow's discussion: We will collaborate here and try to figure out how we are going to run that, and we will give you a briefing on that tomorrow.

CO-CHAIR KAPLAN: Any final -- Laurent?

MEMBER GLANCE: I don't know if this is really feasible, but in my mind -- and I think the fact that we are still talking about both measures tomorrow -- Is that correct?

MR. AMIN: That is correct.

MEMBER GLANCE: so both measures are still in play, and what I would find extremely helpful in terms of personally being able to make a decision comparing both measures would be if we had a little bit more information on model performance.

Specifically, what I would love to

see from both measure developers are just calibration curves. Okay? I think that would be extraordinarily helpful. Thank you.

MS. PACE: Let's just ask the measure developer, is that something that you could discuss with us in the morning, calibration curves?

MR. SAUNDERS: We will do our best. If it is in our existing output, we probably will be able to, but we are dependent upon our programmers at the software vendors, and they may not be on the plat right now. so we will do our best to have the answer for all of that. I think we probably have the output from something to give a calculate.

CO-CHAIR KAPLAN: Thank you. It is my understanding now that we have to go to public comment before we wrap up.

MR. AMIN: Nicole, can you open up the lines, if there are any members of the public that would like to address the Committee?

OPERATOR: Certainly.

MR. AMIN: Any members here in the public that would like to address the Committee?

CO-CHAIR KAPLAN: Okay. Thank you very much for all your hard and diligent work. This was not an easy task, at least on my end of the table, and I am sure everyone is about at the point of cognitive exhaustion.

So I would welcome you back refreshed and ready tomorrow morning at eight o'clock sharp to begin our work on this competing measures kind of issue.

Thank you again.

(Whereupon, the above-entitled matter went off the record at 4:39 p.m.)

F

S

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Readmissions Steering Committee

Before: NQF

Date: 12-05-11

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

Court Reporter