NATIONAL QUALITY FORUM

+ + + + +

ALL CAUSE ADMISSIONS AND READMISSIONS
STEERING COMMITTEE

+ + + + +

MONDAY
May 5, 2014

+ + + + +

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

PRESENT:

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

KAREN JOYNT, MD, MPH, Brigham and Women's Hospital
PAULA MINTON-FOLTZ, RN, MSN, Harborview Medical Center; UW Medicine
PAULETTE NIEWCZYK, PhD, MPH, Uniform Data System for Medical Rehabilitation*
CAROL RAPHAEL, MPA, Subject Matter Expert
PAMELA ROBERTS, PhD, MSHA, ORT/L, SCFES, FAOTA, CPHQ, Cedars-Sinai Medical Center
ALISON SHIPPY, MPH, Consumer-Purchaser Alliance, National Partnership for Women & Families
THOMAS SMITH, MD, FAPA, American Psychiatric Association
RONALD STETTLER, United Health Group
CRISTIE TRAVIS, MHA, Memphis Business Group on Health

NQF STAFF:

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

ALSO PRESENT:

JANE BROCK, MD, MSPH, Colorado Foundation for Medical Care
KEZIAH COOK, PhD, Acumen

LAURIE COOTS, RTI

DEBORAH DIETZ, Acumen*

DAVID GIFFORD, MD, MPH, American Health Care Association

MELVIN INGBER, PhD, RTI

JACK KALBFLEISCH, PhD, University of Michigan Kidney Epidemiology and Cost Center

EUGENE KROCH, PhD, Premier, Inc.

YI LI, PhD, MS, University of Michigan Kidney Epidemiology and Cost Center

JOSEPH MESSANA, MD, University of Michigan Kidney Epidemiology and Cost Center

LAURA SMITH, PhD, RTI

URVI SHAW, MPH, American Health Care Association BETH STEVENS, MS, Colorado Foundation for Medical Care

* present by teleconference

TABLE OF CONTENTS

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

8:03 a.m.

MR. AMIN: Good morning, everyone and welcome to the National Quality Forum. Thank you all for all of the time that you have spent up to this point and for the two days that we'll spend on these complex measures, so it will be a very filled agenda, so we appreciate everyone's time in sort of keeping their comments directed and to help us keep moving the conversation along. Again, thank you for being here. And I'll turn it over to our two co-chairs for their welcome from Sherrie Kaplan and Bruce Hall.

CO-CHAIR HALL: Welcome. I don't have a lot else to say.

(Laughter.)

CO-CHAIR KAPLAN: Ditto. Moving on. Seriously, this is a beautiful day in Washington and thank everybody for -- many of you made a long journey to get here and we're all very appreciative of the hard work we're

about to do and we do want to make sure that we keep to the schedule. We've got a lot of measures to review in a fairly compressed amount of time to do it in, so we appreciate the spirit of conciseness.

MR. AMIN: Adeela, take it away.

MS. KHAN: Good morning, everyone. My name is Adeela. I'm the project manager on this project. I just wanted to quickly go over some housekeeping items for you all.

The restrooms are at the end of this hallway to the right, men and women. We'll have three breaks during this meeting, one at 10:45; lunch will be at 12:30 and then at 3:30. For those of you who have your laptops and cell phones, the WiFi information is also up there. The user name is guest and our password is capital NQF and lower case guest. And we do ask that you please mute your cell phones during the meeting.

MR. AMIN: Adeela, one other thing if we can -- for administrative purposes. If

everybody can turn their name tag to the front so that we can -- so the co-chairs know, can see everybody and make sure that they can recognize you by name. And as is sort of NQF tradition, if you would like to speak, just raise your placard to the side in this fashion and the chairs will note the order in which they've been raised and do their best to make sure that we go in order.

The second housekeeping item is that we can only have two microphones on at the same time, so please remember to turn off your microphone when you're done speaking. And secondly, it's really important to use your microphone because the meeting will be transcribed and obviously the meeting is open to the public and is webcasted. So thanks, Adeela.

MS. KHAN: Sure. And just to introduce our staff really quickly, I'm Adeela. I'm the Project Manager. We have Zehra Shahab, our Project Analyst over here.

Andrew Lyzenga is our Senior Project Manager. And then, of course, Taroon Amin is our Senior Director.

I'm going to turn it over to Anne who is going to be going over our introductions and disclosure of interest.

MS. HAMMERSMITH: Thanks, Adeela. I'm Anne Hammersmith and NQF's General Counsel. And for those of you who have been on any of our committees before and probably a familiar face as I do the disclosures at the first meeting of each committee.

I'm going to run through some introductory remarks to remind you of a few things and then we'll go around the table and you can introduce yourselves and tell us if you have anything to disclose.

If you recall, when you were nominated for this committee, you received a lengthy form that asked you about your professional activities. So what we'd like to do here today is have you disclose matters

that you believe are relevant to the subject matter that the committee will deal with today and tomorrow. We're not looking for you to recount your resume, but just to disclose things that are relevant to the subject matter of the committee today and tomorrow.

I want to remind you that you sit as an individual on this committee. You are here because you're an expert. You do not represent your employer. You do not represent any entity that may have nominated you to serve on this committee.

Another reminder is that disclosures aren't limited to financial disclosures or conflicts. NQF's conflict of interest process is a little bit different in that regard. It's because of the nature of the work that we do. You may have served on a committee as a volunteer for a professional society or some other group where the work that that committee did was relevant to what you're doing here today. It's not necessarily

a conflict, but we look to you to disclose that. And I want to stress that disclosure does not equal conflict. Part of the idea here is to be open and transparent for each of you to know where people are coming from for the public to know that. So just because you disclose does not mean that you have a conflict.

As I said, we are only looking for you to disclose things that are relevant to the work the committee will do today and tomorrow. We are particularly interested in your disclosure of grants, research, or consulting, but only if it is related to the subject matter of the committee.

So with that, let's go around the table, tell us who you are, who you are with, and if you have anything you would like to disclose. I always start with the chairs.

CO-CHAIR KAPLAN: My name is Sherrie Kaplan. I am Assistant Vice Chancellor for HealthCare Measurement and

Evaluation at University of California Irvine. I used to say I was a psychometrician by training, but then I got into trouble with the clinometric psychometrics kind of distinction, so now I say I'm a measurement scientist. It has clarified nothing.

So my prior NQF service, I co-chaired the All Cause Readmissions, Measure 789. I also served on the Composite Measures Advisory Group. And I am now a member of the Family and Patient Centered Measures Committee for NQF as well.

Since the last reporting period, two things have come up. I've received a grant from Patient Centered Outcomes Research Institute to measure children ages 4 to 12 functional status and well being using an animated touch screen based measure performance for use in perioperative anxiety in children. And I also serve on the Physician Compare Advisory Committee that has as one of its contractors Acumen.

CO-CHAIR HALL: Bruce Hall. Welcome again, everybody. I'm a Professor of Surgery at Washington University in St. Louis and a Professor of Healthcare Management for our Business School as well. I serve as a Vice President for our healthcare corporation, BJC Healthcare. I'm the Associate Director of the National Surgical Quality Improvement Program for the American College of Surgeons out of Chicago. And I've served the NQF on a number of different projects and committees in the past. Like Sherrie, I assisted with the 789 measure that was not too long ago reviewed. I've served as a measure developer for NQF on behalf of American College of Surgeons. No measures in front of us for this session have I been a developer on and one measure of the vascular surgery measure I was an expert for. We'll discuss that tomorrow. So as far as I know, I have no other issues or conflicts.

DR. AUGER: I am Kathy Auger. I

am an Assistant Professor at Cincinnati Children's Hospital. I'm a pediatric hospitalist and health services researcher. My primary interests are in pediatric readmission and this is my first NQF experience. And the only past funding that might be relevant is I received a grant from Blue Cross Blue Shield Foundation of Michigan to look at problems of pediatric readmission and risk factors for pediatric readmission. But no stake in the current measures.

DR. EDMUNDSON: Good morning. I'm Ross Edmundson. I'm a VP Medical Director in the Florida hospital system which is the Adventist Healthcare System. I'm not a researcher and this is my first time in NQF. I'm actually quite nervous, but honored to be here. I appreciate the opportunity.

In my capacity, I do -- it's a large hospital system, seven hospitals with about 2200 beds total and a large tertiary hospital, so I get intimately involved with

the readmission and admission related, on the ground, boots on the ground type of issues. But I have otherwise, I believe, nothing to disclose.

DR. GRIGONIS: I'm Tony Grigonis. Currently Vice President of Quality Improvement for Select Medical Corporation which is a for-profit, post-acute healthcare organization and I don't have any grants or any other committees that I've been on that would constitute a conflict of interest.

Our healthcare organizations owns long-term acute care hospitals and in-patient rehabilitation hospitals, so we have sort of a direct good purpose for having me on this committee. Thank you.

DR. ROBERTS: I'm name is Pam Roberts and I work at Cedars-Sinai Medical Center in Los Angeles, California and I am in charge of rehabilitation there. And I have served on the Member Application Partnership for Post-Acute Care and Long-Term Care for

NQF. I also was a tapped member for the Readmissions IRF Group and I recently received funding for the Centers for Rehabilitation and Research using large databases to study readmissions and stroke.

DR. HEIDENREICH: I'm Paul Heidenreich, a cardiologist and Vice Chair for Quality at Stanford, Department of Medicine. I also work at the Palo Alto VA and I'm chair of the Task Force on Performance Measures for the American College of Cardiology and American Health Association, but none of -- we did not address any of the measures that are going to be presented this week.

DR. BROOKS: Hi. My name is Jo Ann Brooks. I'm Assistant Vice President for Indiana University Health for Quality and Safety and we are based in Indianapolis, Indiana. We're a fairly large system.

I served on the previous All Cause Readmission Committee with NQF and I have no potential or other conflicts of interest for

this meeting.

MR. STETTLER: Ron Stettler. I'm Vice President for Healthcare Economics for UnitedHealth Group. And I have nothing to disclose.

MS. TRAVIS: I'm Cristie Travis. I'm the CEO of the Memphis Business Group on Health in Memphis, Tennessee and I work with the major public and private employer purchasers in our market. I also served on the other All Cause Readmission Committee for NQF. I serve on the Consensus Standards Approval Committee for NQF as well as the MAP Hospital Work Group. And I have no conflicts to disclose.

MS. MINTON-FOLTZ: Good morning. I'm Paula Minton-Foltz, and I'm an Assistant Administrator with the University of Washington Health System. Specifically, I'm at Harbor View which is the Trauma Level 1 there. We have eight entities in our system in Seattle, Washington.

I was on the 2011 All Cause Readmission also. So I recognize quite a number of you. I also am on the Washington State Hospital Association's HEN specifically for readmissions, but it's implementation and also the Governor's BREE Committee which is also representing in payers' interest in implementing some of these measures. Thank you.

DR. JOYNT: Good morning. I'm Karen Joynt. I'm an instructor in Medicine at Harvard Medical School in Health Policy at the Harvard School of Public Health. I'm also a practicing cardiologist in the VA. This is my first time at NQF and I don't have any disclosures. Thanks.

DR. FISHBANE: Good morning, Steve Fishbane. I'm a nephrologist, Vice President and Director of Research for the North Shore-LIJ Health System. We have two current grants to study readmissions. One of them is to look for methodology for avoidable readmission

testing for dialysis patients. The other is just related to risk factors for readmissions. I haven't worked previously with NQF.

DR. FIELDS: My name is Wes Fields. This is my second go round. I was part of the Regionalization of Emergency Services. It turned out to be a real interesting activity. My academic appointment is with the Clinical Faculty of the University of Irvine Emergency Medicine. I kind of split my time between Northern and Southern California. My main sort of industry handle is past chair and long-serving member of the largest partnership in acute care, so my group, my partners in hospital medicine have a lot of exposure to the readmission problem at about a hundred locations or so, primarily on the West Coast, but in other regions as well and I very much appreciate the chance to be with you.

MS. CENTENO: Good Morning, my name is Mae Centeno and I'm a Corporate

Director for the Baylor Healthcare System. I co-chair the Readmission Reduction Task Force for across 11 facilities and I'm involved in a grant right now that's about to complete looking at risk stratification and implementation of interventions for patients with heart failure and pneumonia. Thank you.

DR. SMITH: Hi, everybody. I'm Tom Smith. I'm a psychiatrist at Columbia University and the New York State Psychiatric Institute in New York. I'm a clinical researcher and mental health services researcher. We do a lot of work with the New York State Office of Mental Health. Our main project right now is developing the performance metrics for the behavioral health, public health Medicaid-managed care program that's being implemented in New York State. I have no other outside funding relevant to this work.

This is my first Measures Committee with NQF. I am on the American

Psychiatric Association Quality Council which is how I was referred here, although I do also sit on the NQF Readmissions Action Team at present.

MS. SHIPPY: Good morning, Alison Shippy with the National Partnership for Women and Families. It's an advocacy organization here in Washington, D.C. I don't have any disclosures to note. I do sit on the MAP Coordinating Committee though.

MS. HALL: Leslie Kelly Hall from Healthwise and I am a consumer advocate and spend time mostly in the meaningful use space and in Health Information Technology Standards Committee. I have nothing to disclose.

DR. GLANCE: Good morning. My name is Larry Glance. I am a cardiac anesthesiologist and a health outcomes researcher. My appointments are at the University of Rochester. I'm Professor and Vice Chair for Research. I also am Professor at Public Health Sciences and have a secondary

appointment as a senior scientist at RAND Health.

I have previously served at NQF. I was also a member of the prior Readmission Steering Committee. And I don't have any conflicts of interest, although I do serve on the American Society of Anesthesiologists Committee for Performance and Outcomes Measures. Thank you.

DR. CHEN: Good morning. I'm Helen Chen. I'm the Chief Medical Officer of Hebrew Senior Life which is an integrated senior healthcare organization located in Boston, Massachusetts. We're the largest provider of post-acute care in New England and we serve about 2,000 lives across the continuum of care from outpatient through inpatient rehabilitation through long-term acute care hospitals.

In my previous academic career, I was actually a Professor at UCSF. Recently joined Hebrew Senior Life and otherwise have

no other disclosures.

MS. HAMMERSMITH: Thank you. I understand there's a committee member on the phone, Paulette Niewczyk. Is Paulette Niewczyk on the phone?

DR. BULGER: John Bulger. I'm the Chief Quality Officer for the Geisinger Health System. I don't believe I have any conflicts at the moment.

MS. HAMMERSMITH: Thank you for those disclosures. I just want to remind you of a few things before I leave, one of which is that we rely on all of you to have a successful conflict of interest disclosure process. So if you are sitting here and you think you may have a conflict, you think one of your fellow committee members has a conflict, you think someone is behaving in a biased fashion, please do speak up. We don't want you sitting in silence if you think something is up.

If you do want to raise something

like this, you can bring it up openly in a meeting at any time. You can go to your co-chairs who will go to NQF staff or you can go directly to NQF staff. So based upon the disclosures that have been made this morning, do you have any questions of each other, anything that you want to discuss or raise?

CO-CHAIR HALL: I note that two members are unaccounted for. Has NQF accounted for them? Have they said they would not be present? Frank will be here, okay. And Carol, okay.

MS. HAMMERSMITH: Okay, anybody else? Thank you.

MR. AMIN: Thank you very much, Anne. I just also would like to introduce Helen Burstin, our Senior Vice President in our Performance Measures Group.

DR. BURSTIN: Good morning, everybody. I just want to add my welcome and say thank you, especially for those of you willing to come back. Last time was a pretty

intense committee. I suspect this will be as well. This is obviously very high profile, but really, really important and I just want to thank you.

I also just want to mention that it's a thrill to have so many of you back including Sherrie as co-chair. We actually had a co-chair with Sherrie, Eliot Lazar, who some of you may know who is the Chief Medical Officer at Columbia who passed away. So I wanted to at least recognize his great service to both New York and to NQF and he's definitely missed. We are in excellent hands between Sherrie and Bruce and thanks for your support in advance for what will probably be an interesting process. Thanks.

MS. KHAN: Okay, so I guess it's back to me. So I wanted to talk a little bit about the role of the standing committee. You'll be chosen to either serve a two- or three-year term. During that two- or three-year term, you'll be working with NQF staff to

achieve the goals of the project which is to review all the measures, evaluate each measure against each criteria, indicate the extent to which each criteria is met and the rationale for the rating.

You'll also be making recommendations to the NQF membership for endorsement and you'll be responding to comments submitted during the review period. You'll also be responding to any directions from the CSAC which is our Consensus Standards Approval Committee and you'll also be overseeing the portfolio of should be readmission measures.

MR. AMIN: Adeela, before you move on on that, I just wanted to note for the committee that this is a change to the NQF process. For those that are new to NQF, and those that are returning, we are instituting standing committees so you'll be asked to be sort of supporting NQF in this area of readmissions for the next two to three years

and that will include to the extent that you're willing, obviously, that will include measure review, but also will include elements of follow-up activities, for instance, reviewing dry run results or reviewing updates to NQF guidance.

And so part of what we'll do today is review the portfolio because this will be an area that we would expect some level of input from the committee and then also relevant guidance related to technical issues related to measures. So we've also invited Karen Pace, our lead methodologist to walk through some relevant information that won't necessarily be relevant to the measures in front of you, but will be relevant in future efforts related to readmissions and admissions.

MS. SHIPPY: So just to clarify for the committee, over that two to three year course, conflicts can change. People get grants, things happen, interests are kind of

pursued, blah, blah, blah, blah, blah. So can you clarify for us what happens in how to deal with those conflicts as they occur?

DR. BURSTIN: So we recognize that and that's fine and that's -- we have actually as part of this process of moving to standing committees, also have a policy of recusal. So you no longer -- we used to have people if you had a conflict you couldn't even be on the committee at all. And now what we do is you can be on the committee, you just have to recuse yourself and not participate in any discussions or vote on any measures for which you have had a role. So it's actually -- it should work out fine. I mean if it works out that you are, in fact, on the overwhelming majority of the measures and you would be silent, we would probably ask you to just sit that one out, but pretty unlikely, I think, particularly in this space. Thanks.

MR. AMIN: So one thing that we neglected to do is during the introductions,

but we want to make sure that this is in the record, so we're going to around and ask you to draw a term and if you could just say your name and your term as we walk around the room. We'll try to do this as quickly as we can.

Your term, you'll either select a two or three-year term. It will be randomly selected.

DR. BURSTIN: And the terms are renewable. So even if you get a two-year term, we'd be delighted to have you for four years.

MR. AMIN: So Steve, if you don't mind, just starting with the term that you've selected just so that we have it in the record.

DR. FISHBANE: Yes, Steve Fishbane. It's a 30-year term.

(Laughter.)

It's a three-year term. Three-year term.

DR. FIELDS: Wes Fields. Three-

year term.

MS. CENTENO: Mae Centeno. Three-year term.

DR. SMITH: Tom Smith. Two-year term.

MS. SHIPPY: Alison Shippy. Two-year term.

MS. HALL: Leslie Kelly Hall. Three-year term.

DR. GLANCE: Larry Glance. Two-year term.

DR. CHEN: Helen Chen. Three-year term.

CO-CHAIR HALL: Bruce. Two.

DR. BURSTIN: I got a three-year term.

DR. BULGER: John Bulger. Three-year term.

DR. AUGER: Kathy Auger. Two-year term.

DR. EDMUNDSON: Ross Edmundson. Three-year term.

DR. GRIGONIS: Tony Grigonis. Two-year term.

DR. ROBERTS: Pam Roberts. Three-year term.

DR. HEIDENREICH: Paul Heidenreich. Two-year term.

DR. BROOKS: Jo Ann Brooks. Two-year term.

MR. STETTLER: Ron Stettler. Two-year term.

MS. TRAVIS: Cristie Travis. Three-year term.

MS. MINTON-FOLTZ: Paula Minton-Foltz. Three-year term.

DR. JOYNT: Karen Joynt. Two-year term.

CO-CHAIR HALL: Tonight after dinner, there's going to be a tug-of-war between the Twos and the Threes.

(Laughter.)

CO-CHAIR KAPLAN: Frank Briggs will have a three-year term. Carol Raphael,

a two-year term. And Paulette Niewczyk will have a two-year term as well.

MS. KHAN: So, going back to our meeting expectations, I just wanted to talk about some of the expectations for today's meeting. NQF is continuing to improve our committee meetings based on input from our multi-stakeholder membership. And we've made a few changes since some of you may have been here to our meeting process.

We want to recognize that we have our measure developers present and we'll be asking them to briefly introduce their measures for discussion. Once they've done that, selected work group members will then begin discussion of the measures in relation to the measure evaluation criteria.

We've provided the developers two seats up front. If you have other people from your team who would like to join us, you can actually go to the back table back there. We've provided the designated place for the

developers at the main table during the introduction and discussion of the measures. Here, they are easily able to respond to questions from the committee and correct any misunderstandings about their measures during our discussion.

Developers can put up their cards to indicate when they wish to respond to questions raised or correct any statements about their measures. During the measure evaluation, committee members often offer suggestions for improvement to the measures. It's important to note that these suggestions can be considered by the developer for future improvements. However, the committee is expected to evaluate and make recommendations on the measures for the submitted specifications and testing.

This multi-stakeholder group brings various perspectives, values and priorities, so the discussion and respect for differences of opinions and a collegial

interaction among the committee members and measure developers is expected.

Again, some ground rules for today's meeting. The agenda is quite full. All the committee members, developers, and staff are responsible for ensuring that the work of the committee is completed during the time allotted. During the discussion, committee members are expected to be prepared, having reviewed the measures beforehand. They're expected to base the evaluation and recommendation on the measure evaluation criteria and the guidance. We've actually provided the guidance tables for you. It's that colored document that's on the table.

Remain engaged in the discussion without distractions. Attend the meeting at all times except during the breaks. Keep comments concise and focused and avoid dominating the discussion and allow others to contribute and indicate agreement without repeating what's already been said.

Does anyone have any questions?

This is just a timeline for activities. So we have our in-person meeting today and tomorrow, May 5th and 6th. We'll be drafting our report after the meeting and we'll be posting it for NQF member and public comment, June 6th through July 7th. After that, the world can be in the steering committee again to respond to all of our comments. Once the committee has responded to all of the comments, we'll be posting our draft report to our website and the measures will go out for NQF member vote. They'll be reviewed by the CSAC in August and then we hope for the Board to endorse the measures by September. And we'll have a 30-day appears period as we do with all NQF projects in October. We'll have more specific dates for you as the project moves on.

I'm going to turn it over to Taroon now to go over the overview and evaluation process.

MR. AMIN: Thank you, Adeela.

DR. SMITH: Could I interrupt?

MR. AMIN: Yes, please.

DR. SMITH: Maybe I'm the only person that cannot get the WiFi access, but if I am, is there any IT Supporters, someone that could look at my computer?

MS. KHAN: Sure, we'll get you someone.

DR. SMITH: Thank you.

MR. AMIN: Yes, thanks. Actually, I was just going to ask, are there any questions from the panel? I know we're trying to move things along. However, also recognizing that a number of new folks joining us, so please, feel free to ask any questions that you may have. We want to make sure that you're able to fully participate in the conversation.

So I'll just go through a very, very brief introduction of what we're going to do throughout the course of this process. As

you know, we've gone through a call for nominations process that was open to across all stakeholders. We've had a comment period on your nomination to this committee. And we've gone through adjudicating those comments. We've had a call for the consensus standards which are the standards that are in front of you today.

Today, we are looking at the standards review which include the review of submitted and maintenance measures. Maintenance are measures that are currently endorsed and we require a three-year review of any endorsed measure for updated evidence or testing, and testing.

The committee deliberations during today's discussion and tomorrow's discussion will have a number of public comment periods in which we'll invite members of the public to provide input to not only the measures, but also the committee's deliberations. And we'll go to a formal 30-day comment period where

members of NQF and the public will be asked to comment on the committee's decisions and the recommendations of the group.

And we will have an adjudication call for those comments. We do expect that this project will generate a significant amount of interest. And so we'll have a significant amount of comments to review.

We go through a member voting process which we ask our members to vote on the measures that are in front of you. And then this information will go to our Consensus Standards Approval Committee which is a governing committee which will review the recommendations of the committee and then we go to the Board of Directors which ultimately ratifies the decisions of the committee.

Any of the measures in front of us can go through an appeals process if stakeholders feel that there has been a significant change in the field requiring a re-review of these measures. And the NQF

staff generally review these appeal requests and also the Consensus Standards Approval Committee.

So that's the general process that we'll be going through. That's typically what we call our CDP project. This is what we do for our general measures and this is what we'll be doing over the course of the few months that we'll be working together on this particular effort.

What you'll be doing today on the next slide is -- yes, please?

CO-CHAIR KAPLAN: So for those who haven't served before on the NQF committees, can you clarify the three year review process and how much wiggle room there is if someone wants to shorten that, for example, or if the committee feels like there should be a shorter length than three year.

And the second thing is is how public is public? Can John Q. Public dial in to these meetings, etcetera, etcetera?

MR. AMIN: Yes, so I'll start with the second one first. So these meetings are completely open to the public. The dial-in information is available on our website so it really could be any member of the public. Your friends could call in and listen in on the committee deliberations if they're so interested. And we would welcome comments from them.

And the three year maintenance cycle is that if there are no changes in the evidence or in the -- I would just say let's just say the evidence, we would have a three-year cycle and that's generally -- we try to have projects that fit within that three-year cycle, although they could span between two and four years at times. But if there is any significant change in the field or evidence with the measures in front of you, any member of the public can request an appeal, ad hoc review, sorry, thank you, Helen. An ad hoc review of any of the measures that are in

front of us.

Now as a standing committee, if there is, for example, in the last project there was some questions around threats to validity for some of the measures that were in front of you. You asked the developer and they agreed to provide some information back to the committee related to the dry run results of the measure, meaning as the measure was being implemented since it hadn't been implemented prior. If that is an agreement between the committee and the developer, that type of information can be brought back to the committee during their next review.

So in some ways, some of you are already acting in the spirit of the standing committee, but typically, it's a three-year cycle.

Are there any other questions from members of the committee on this?

Okay, thanks. So I'll just finish on this slide. Obviously, you guys have been

a very committee through our work group calls. I won't really spend a lot of time on the criteria, but again, just as a reminder, we have conditions for consideration before the measures are able to come to the committee requiring essentially a number of conditions. And then we have four main criteria importance to measure and report: scientific acceptability of measure properties, feasibility, use, and usability. We would expect for these measures since the majority of the measures that are in front of you are outcome measures that the bulk of the discussion will be in the scientific acceptability section of the evaluation.

Generally, for our more clinically-oriented process measures, the importance of the measure and report criteria involves a review of the evidence, the quality, quantity, and consistency of the evidence, justifying the measure focus. In the case of the measures that are in front of

you, since they are outcome measures, we will generally move pretty quickly through the importance to measure and report portions of the discussion. However, if there are comments that you have, feel free to raise them. But we'll likely move through that section of the evaluation relatively quickly.

Scientific acceptability include reliability and validity which is again, generally the area of most heavy discussion. And then finally, I don't believe we have a discussion in this panel related to harmonization. And I think that's all I wanted to point out here.

Is there anything else, Andrew, or Adeela that you want to raise?

Okay, so in the spirit of the standing committee, we wanted to have a quick discussion -- there's two sections of the discussion that we wanted to add now as we're moving more toward a standing committee. The first is to give you more of a macro

understanding of some of the other work that's going on at NQF and that's related to this work, but not corely related to the measures that are in front of you.

And we have invited Karen Pace, our lead methodologist, to give you a quick update on our work related to risk adjustment and SES. And then following that, I'll turn it over to Andrew, who will walk through the measures that are in the portfolio that ultimately the committee will be responsible for which include various different types of readmissions and admission measures. So we'll do a quick walk through of the measures that are currently in the portfolio that are within the purview of the standing committee.

So Karen, I welcome you to begin.

MS. PACE: Good morning, everyone. So probably many of you already aware that we're in the midst of a project where we're looking at the question of whether outcome measures and potentially some process measures

could include adjustment for socio-demographic factors. So initially, the project was labeled socio-economic status. Actually that's a key element, but also could potentially be other socio-demographic factors. That project is currently in motion. We just completed our 30 day public comment period. The expert panel will be reviewing those comments actually this Friday and deciding whether there are any adjustments to be made to their recommendations. And then it will go to the Consensus Standards Approval Committee and Board. So we don't anticipate that project really being completed until the end of June, potentially beginning of July, depending on how things go.

But I wanted to just mention where we're at, where the expert panel's recommendations fall, and some of the issues that we'll be confronting. But as Taroon said, you know, it's important to keep in perspective that the measures that you'll be

looking at today and tomorrow, you really are being asked to judge them against our current criteria. So I know that that's hard to do sometimes, but we do have to kind of move things along in a systematic way so that everybody knows what they're being held to.

That being said, the expert panel's recommendations are that outcome measures and potentially some process measures could be adjusted for socio-demographic factors. And I think it's important to realize that there's several conditions upon that. The first is and the recommendations are actually stated, when there is a conceptual and empirical relationship between the factor and the outcome of our process of interest. So the point is that it's not a blanket every measure should be adjusted for socio-demographic factors, that just as for clinical risk factors, one has to follow good, systematic, sound methods. And the first step is is there a conceptual reason that that

factor might be related to the outcome of interest? And when you look at those variables reflecting that factor, is there an actual empirical relationship with the outcome or process of interest and then to follow all of the other guidelines for developing a risk model.

Another key point is whether the distribution of that factor across providers varies and then you start looking at using these factors in risk models with your other factors and seeing what really works in terms of a risk model. So I just want to clarify because some have interpreted that the recommendation is that every measure should be adjusted for socio-demographic factors and that is not what the recommendation is.

The other thing that I think is important to clarify is that the expert panel did not abandon stratification for identification of disparities. Their recommendation is really two part, basically

to look at if you're having one computed score of a performance measure, they're recommending that if it's relevant, all the caveats I just mentioned, then it could include socio-demographic factors. But very much identify that if we want to identify disparities and actually work to improve them, that we will need to be doing risk adjustment -- or I'm sorry, stratification so that we look at the outcomes by the various factors, you know, it could be income, it could be homelessness, whatever the socio-demographic factor that's in play for that particular performance measure.

So I think those are the key recommendations. Certainly what follows from those is that the NQF criteria which currently states that statistical risk models should not include socio-demographic factors -- or basically it says should not include factors that are related to disparities. And the guidance is that stratification is preferred.

So that kind of prohibition against having socio-demographic factors would be removed from the criteria.

The committee also made specific recommendations about measure submission and the information that steering committees such as yourself, would need to have available to really look at the adjustment process, what factors were included, how that decision was made, what those contribute to the model, etcetera.

And then there are some -- and that the guidelines that are used for selecting risk factors apply to the socio-demographic factors, so again, the emphasis on sound methods. And then there's some additional recommendations related to improving data collection, some suggestion that NQF look at its stance and potentially start looking at providing implementation guidance for endorsed measures and some other clarifications. But the key things were

related to socio-demographic factors. So I'm going to stop.

Helen, is there anything you want to add to that?

DR. BURSTIN: No, that was a great summary. The only thing I'd add is again considering that it is still in draft form, it is not going to be finalized until July, that's not the principles under which we are operating for this group. We may need to return to these measures, if that moves forward, but at least to just keep in mind that is still draft. We've got 670 comments. So it wasn't a report that sort of people didn't notice. There's a lot of discussion. The committee has a four-hour call on Friday to review those comments and see if there might be some modifications in some of those recommendations to seek common ground.

Again, as a consensus-based organization, our goal is not to have reports that move forward with -- where we don't, in

fact, try to reach consensus. We will try our best over the next couple of months to reach that, but we are still operating under our current guidance for now.

MS. SHAHAB: Right, so as Helen said, you know, our process is to attempt to resolve objections. And so we will be addressing the comments that came in that were opposed to some of the recommendations and seeing, as Helen said, if there's any way to resolve those and have the expert panel really examine those and identify whether there are any potential modifications before they move forward with their recommendations to the Consensus Standards Approval Committee.

MR. AMIN: So I will turn it over to the chairs, if you have any introductory comments, then welcome some discussion on the topic.

CO-CHAIR HALL: Thank you. I'll actually ask two questions of Karen.

Karen, as you stated a minute ago,

there's more or less a prohibition currently on such adjustments, but we know that one or more of our measures in consideration to today or tomorrow, is adjusted or has a form of adjustment. So are we to consider that actually still prohibited?

MS. SHAHAB: Good question. So we actually, NQF actually has endorsed some measures that have socio-demographic factors in the adjustment model and those have to be well justified. You know, I think we'll have to look at them on an individual basis and really try to sort that out. But it has been -- as the criteria state that it should not include factors related to disparities and the preference is to stratify, but as you all are well aware, different situations may call for different considerations and that happens on a measure by measure basis, but we try to stick to the criteria as closely as possible. And we'll certainly be interested in your questions about that and discussion with the

developers so that we can try to resolve that.

CO-CHAIR HALL: Okay, thank you. My second question is let's pretend that the new recommendations would be finalized in July. Will the NQF make a statement that this group of measures should be examined or will it invite developers to submit any comments or should this group be thinking about some sort of statement in the process of our work today and tomorrow saying if such recommendations were to be finalized, we would recommend that one or more measures be reexamined?

MS. SHAHAB: Yes, I think that would be certainly within your purview and would be of interest to NQF. Some of the comments that came in that is probably more for NQF than the expert panel is what -- if these recommendations are upheld and approved by NQF, what would be the process of looking at previously endorsed measures. So that is something that NQF with the CSAC will need to grapple with, but I think, you know,

statements from this committee, if they think that's a primary issue that we would certainly welcome those.

DR. BURSTIN: And to that, obviously, the way they're -- assuming they stay the way they are, it does require a conceptual and empiric relationship so some of it is you'll be looking at it today without the information, in fact, on whether there is at least an empirical relationship. You would probably infer something about the conceptual relationship. So it would really just be that these might be measures that you would want to think about asking for those additional analyses, but you certainly look at it and say yes, it should be or shouldn't be because you don't, in fact, have that data.

CO-CHAIR HALL: And I would add to that in reading the recommendations, as they've existed so far, that they're actually pretty burdensome. So they actually grate on the developers quite a bit of demand for their

work and I think it would be unfair to expect the developers to have reacted to something that is not finalized yet.

CO-CHAIR KAPLAN: Let me just reiterate what I heard so we're all clear. This is going to come up. It's come up in the all cause for admissions discussion as well. So one is that we will evaluate the measures under consideration as presented by the measures developer currently. So what you see is what you get. And the guidance yet coming from the Socio-economic Status Committee isn't there yet. It's on the horizon.

So we are not to consider socio-demographic adjustment for any of the measures unless specified in the document by the measures developer, one. And two, the stratification by hospital type, I just want to -- whatever the unit of comparison is is okay to consider for the purposes of enveloping in that stratification of hospital type, comparisons that are fair, but not based

on socio-demographics, based on other considerations like hospital size or whatever the stratification. But it still has to be as presented by the measures developers. Is that so?

MS. PACE: Right, so when you're talking about stratification of the hospitals, let me make sure I'm understanding. I think the committee last year recommended for implementation that like hospitals be compared. And last year MEDPAC came out with a recommendation specifically about I think it was all cause admissions and looking at -- comparing hospitals within deciles of income that they're serving.

So that type of implementation guidance, the steering committee can make some statements, but as you're well aware and that is one of the recommendations from the panel for NQF to consider what is the role of NQF in making implementation guidance.

So certainly the committee can

talk about that. It doesn't necessarily -- it's not part of the measure specifications that are being endorsed and so that it's not something that is hard and fast in terms of part of the endorsement, but the committee can certainly make those kinds of statements.

Taroon, did you want to add something?

MR. AMIN: No. I think the committee has done that in the past and specifically in terms of evaluating the dry run results along certain criteria in the past. So if that type of future analysis is needed and agreed upon by the developer, that might be --

MS. PACE: But I think getting more at your -- maybe the real specifics of your question is since we currently say that generally we don't expect those to be in the risk model, is it fair game then to say to do something about that post-measure score development? It's a gray area. So I think it

really relates more to what Helen was saying.

Since you're not going to have the empirical information, you can make some statements about what you think the conceptual relationship is and what would follow from that, but it would be hard to make more specific recommendations, not knowing exactly how that will play out. But ask again. It's a gray area. It's hard to give black and white answers.

MR. AMIN: Alison?

MS. SHIPPY: Thanks. I have a question from a historical perspective. So you did acknowledge that there are some measures that have been endorsed when those variables have been included in the risk adjustment model. How does that kind of play into the CSAC? I mean they're kind of considered the stewards of the evaluation criteria. Is that kind of flagged as a deviation from the evaluation criteria? I'm just curious from that perspective?

MS. PACE: It's been pretty infrequent and the ones that come to mind, that was flagged for the committee, they asked for specific analyses for justification and so that kind of information would then also go to the CSAC.

CO-CHAIR HALL: Kathy?

DR. AUGER: Thank you. Could you just give some guidance about how broad of a net socio-demographics really encompasses? There are certainly things like race and income, but what about things like payer status? Because dual eligibles or Medicaid patients are indicative of socio-economics as well?

MS. PACE: Right, so it's pretty broad. So for example, of course, the kind of three major aspects of socio-economic status are income, education, and employment or occupation. And there's a recognition that we don't really collect on a standard way income and it's probably not individual income. It's

household income and so we do tend to use at least in the current environment and the data available, use things like Medicaid status, dual eligibility, insurance coverage status. There's also been some work done with address and geocoding to census track or census information which actually has gotten pretty good traction and results. But I think the -- so all of that is fair game, but in terms of the expert panel's recommendations, there's also the reality check that what data are currently available. And that's why one of the recommendations is that NQF, along with other stakeholders such as IOM, who has been doing work in this area, ARC, CMS, really need to come up with some standard core data, definitions, and collection process and so what is even possible is going to change over time. So right now should these recommendations move forward, the first step might be using insurance status. The next step might be more broad implementation of

address and geocoding and probably the longer term is really having good standard data collection.

So the expert panel recognizes this is going to be evolving and like many areas of measurement, it's not going to be perfect, especially starting, but they also kind of saw it as a chicken and egg thing. That actually, this could have an impetus for better data collection for us really identifying disparities and doing something with this data as well. So it's going to evolve over time.

CO-CHAIR HALL: Wes.

DR. FIELDS: Yes, I just want to point out a couple of obvious things. First of all, I really appreciate the interim report, very useful for an emergency physician. But just a couple of specific examples. You now have a country where Medicaid eligibility varies dramatically across blue and red states, something like 26

states have yet to pick up the extension of Medicaid that was part of the Affordable Care Act. That's essentially a political activity and so using insurance status as a metric is problematic. It suggests that you need another way to stratify that may actually be based on income, since the income variation is so radical between states these days.

The same thing to a lesser extent is probably true about people with private insurance purchased through exchanges. Young adults who opt not to purchase it because they can come to my emergency department, that's another way of suggesting or thinking that insurance status may not be truly useful as a way to differentiate strata, patient, and problems.

MS. PACE: All good points and things that the expert panel discussed especially about the Medicaid status and the insurance status as you mentioned. So you know, it's definitely something that's going

to have to evolve over time. We can't use data that don't exist and we have to continue to move forward and try to get better data, but definitely good points need to be considered.

CO-CHAIR KAPLAN: Yes, I just want to underscore that the committee is still working on the report and that we will go over some like disproportionate share hospital status and other kinds of variables that have attempted to address this issue in the dry run results tomorrow. So we're going to get a little bit of a chance to see how these various use of different indicators -- we can argue more or less satisfactory indicators of socio-economic status -- have played out empirically in the all-cause readmissions data dry run for tomorrow.

So keeping us on track, where are we now?

MS. KHAN: I just want to note that we have all of our committee members, so

I would like Frank and Carol to introduce themselves and we also have Paulette on the phone. So if you could do an introduction and just do a quick disclosure if you have anything to disclose to the committee.

So why don't we start with Carol.

MS. RAPHAEL: I'm Carol Raphael and I'm actually the chair of the MAP Post-Acute Care, Long-Term Care, Hospice and Palliative Care Work Group at the National Quality Forum. And was for more than 20 years the CEO of the Visiting Nurse Service of New York and after that a Fellow at Harvard University. And I'm currently a Senior Advisor at Manatt Health Solutions and the Chair Elect of the Board of AARP and was appointed by President Obama to the Bipartisan Commission on Long Term Care and on a number of other boards.

So in terms of I guess what I should disclose, I'm the Chair of the Board of Long Term Quality Alliance which is an

alliance that tries to promote and raise the bar on quality in the field of long term services and supports. And at Manatt Health Solutions, I'm working with a number of systems on how they can better integrate post-acute care into their service delivery.

And the other thing I should just mention is that I'm the Chair of the Health Information Technology Board in New York State and we are working as well to try to integrate care across the state and set up an information highway. And I'm on the National Quality Forum, the group that is looking at emeasures and bringing technology to bear.

MS. KHAN: Thank you. Frank.

DR. BRIGGS: Good morning. Frank Briggs from West Virginia University Healthcare. I'm Vice President of Quality and Patient Safety at the hospital for 12 years now.

MS. KHAN: Nothing to disclose, I'm assuming?

DR. BRIGGS: Right, nothing to disclose.

MS. KHAN: Thank you. Paulette, are you on the phone?

DR. NIEWCZYK: Yes, hi. I'm Paulette Niewczyk. And I'm with Uniformed Data Systems for Medical Rehabilitation in Buffalo, New York. I'm also with the University of Buffalo. And I'm their Director of Research and responsible for establishing and developing the psychometric properties and some of the same derivatives as well as managing the inpatient rehab data associated with the (inaudible) as well as some of the outpatients as well as skilled nursing and other large data repositories that use our tools.

In addition, I am a Professor at Daeman College which is one of the local private colleges in the Buffalo, New York area. And my background is in public health and epidemiology and I also have a doctorate

in sociology where my research is focused heavily on access to healthcare.

My only disclosure is I'm employed at UDSMR.

CO-CHAIR HALL: Thank you, Paulette. So why don't we have --

CO-CHAIR KAPLAN: I have one more thing.

CO-CHAIR HALL: Yes.

CO-CHAIR KAPLAN: First of all, thank you, Karen. That was a nice summary of -- on behalf of the committee of where we are and where we aren't with respect to SES guidance. But for the group, if we have specific queries that the committee feels strongly should be sent to the SES Guidance Committee, it's okay for us, I understand to sort of summarize our issues and forward them on to that group so that we don't leave ourselves terribly, terribly frustrated with respect to where we are right now with respect to the SES issues, is that correct?

MS. PACE: Yes, that would be great and you know, you're going to be in the midst of it over the next two days, so I think it would be very relevant and helpful for us.

CO-CHAIR HALL: And I was just going to ask if any other committee members have any questions for Karen? Not seeing any.

MR. LYZENGA: Great, thanks.

I'm now going to quickly run through the admissions and readmissions portfolio. As Adeela and Taroon mentioned, we have transitioned to a standing committee system. This is as opposed to seating a new committee every time we started up a project. This was done for a number of reasons, among those increasing the consistency of decision making across time, achieving some efficiencies at the process level in terms of project startup, but also so that the standing committees can gain a familiarity with the topic area and the measures in the portfolio and start to kind of steward that portfolio

over time and provide input into the sufficiency of the portfolio in terms of addressing the topic area.

So what we'd like you to do is to just sort of keep that in mind as you're reviewing the measures today and tomorrow, provide input on the measures that are under review, that are included in the portfolio, consider issues of standardization and parsimony when considering measures in the portfolio as a whole. Identify measurement gaps. Raise awareness of other measurement activities for the committee and other stakeholders. Be open to external input and provide some feedback to us on how the portfolio should evolve over time and for developers as well.

Right now we have about 27 endorsed measures that are related to admissions, readmissions, or length of stay. In terms of this committee's purview, for the purposes of maintenance, we've classified ten

measures, currently endorsed measures, as being in this group's purview. Three of those measures, three of those ten are up for maintenance under this project and we've gotten 15 new measures for review and endorsement so that will sort of bump the -- if we do pass those measures, that will bump the number of measures in the portfolio up significantly.

Just to give you sort of a visual depiction of the framework we're using here, admissions are sort of -- can be considered in that sphere on the left side. From the community or non-hospital setting, post-acute care, decline in health status, so a need for a higher intensity of care leads to a hospitalization. That's one sort of area of measurement, that movement from a lower intensity care to higher intensity care. Then we have measures down there at the bottom measuring length of stay in the hospital. And then once discharge occurs, patients go back

to the community or post-acute care and you have that readmission on the right side there. That's what we're measuring with readmissions measures, the decline in health status that requires a higher intensity of care, return to acute care of the inpatient setting.

I'll run through these pretty quickly just in the interest of time. I know we've got a lot to get through today, but I just wanted to give you sort of a quick sense of the nature and scope of the measures currently in the portfolio. The ones in orange are measures that are up for review in this project, so today and tomorrow. Is that right, Adeela? Okay.

So just -- we've broken these down into a few broad categories, one being what are sometimes called all-cause. I prefer to call them all-condition measures. I think all the measures that were being considered today are all-cause, actually, but that sort of becomes the lingo for it. But these are sort

of agnostic in terms of the condition or procedure that a patient has undergone and just measure admission. These are sort of addressing issues at a public health level. We'll skip out to the next.

Here's a few of the measures in the portfolio. I think many of these have actually been assigned to the Health and Well-Being Committee, not as opposed to this one.

Length of stay, sort of the second broad category. A few measures here, one of which we're considering today. Readmissions. Again, we have a few broad categories, all condition measures. Condition or procedure specific. Some of the groups of those are cardiovascular pulmonary conditions, surgical conditions or procedures.

Then we have a number that are setting specific, that are looking at readmissions from a particular setting back to the hospital and those settings include skilled nursing facilities, home health, long-

term care hospitals, inpatient rehab facilities, dialysis facilities, ASCs, etcetera.

So I would open it up at this point to any initial thoughts on the portfolio or questions or comments that you'd like to raise as we move forward into evaluation today.

DR. BURSTIN: I was just going to say, Andrew, just one perspective. I'd be curious if Cristie or others from the MAP, but there was also a lot of discussion at the Measures Application Partnership when they discussed the readmission measures in the past that there was a desire to, in fact, get readmission measures across the broad spectrum of settings, so it's actually very positive to see that movement towards looking at readmissions, not just from a purely to a hospital setting, back and forth between hospital to home, the other settings included as well.

MR. LYZENGA: Any other questions or comments? Leslie.

MS. HALL: Helen, you had mentioned harmonization earlier. Is that something we need to consider in our deliberation today?

DR. BURSTIN: Probably not today. We'll see how -- what we usually do is go through the measures on their own, the right first. If they both make it through well then the staff will walk you through an exercise to look at issues of harmonization. We usually don't do that because sometimes one of them doesn't make it through, so why invest energy to do that in advance. And we'll also have an opportunity to do that after this meeting and I'm sure others will help you through that.

MR. LYZENGA: Paul?

DR. HEIDENREICH: Yes. I noticed that some of the admission measures are dealt by other committees, and going forward though will this committee address those?

MR. LYZENGA: My understanding is that they actually will remain split across some other committees. This is an issue in a number of other topic areas as well. I work on patient safety and that's kind of an issue, a cross-cutting issue and there are a number of measures that are in other committees including surgery, long-term care, other things. It's sort of an artifact of our process of categorizing measures and we would certainly welcome input from you as well if you think that some of these measures would be more appropriately considered by this committee. We can certainly consider that.

DR. BURSTIN: And just to add, these ARC measures that you see listed here, they're prevention quality indicators are community level measures at the community MSA level and they have traditionally been looked at as part of our more population health focus. But again, I think you're looking at some today that are sort of on that border as

well, so we'll try to reconcile those going forward. Logically, they would probably go together in the future.

MR. LYZENGA: Paula?

MS. MINTON-FOLTZ: Thank you. I was curious about the length of stay measure that was put in here. Is that counter measure? I understand your Venn diagram and that there's a relationship, but you know, it kind of just feels like an odd duck in this admission or readmission. So I was wondering about that inclusion?

MR. LYZENGA: I don't know if Helen, you want to speak to why this was assigned to this --

MR. AMIN: Yes, I'd like to say that there is -- that the way that every measure has been assigned to a project is conceptually appropriate. I would say that we've gone through a process with the new standing committee process of really just trying to figure out who best and how best to

assign some of these measures. And I think the length of stay measure in this group is sort of a reflection of that. It's not -- we have to make some artificial decisions when some areas are between two projects and some don't necessarily have a clear home. And so I think that might be potentially part of the reason. So we've adopted them.

CO-CHAIR KAPLAN: Helen, with respect to the title of this committee as a standing committee, because it includes admissions and if some of these are preventable admissions and that's a community-based measure, and some of them are hospital -- could we work on the title of this? Otherwise, the volume of measures coming to this committee could get daunting.

DR. BURSTIN: Again, we look to the committee to give us some suggestions, what is the right framing of this, I mean to the question as well about length of stay. I think we wanted to have something that

collectively embodied this set of measures that are often highly related, but you know, artificial separations don't help either. So --

MR. LYZENGA: Okay, so moving on. I think, Adeela, you're just going to give some general notes before we jump into the actual measures that are in front of us?

MS. KHAN: Yes. So we have a list of the measures that we will be going over today. I won't read them off to you. They're actually listed out in your agenda, but I guess we can start off with 2502. So I believe our developers, RTI, if you want to bring your team up here.

CO-CHAIR HALL: So while they come up to the table any final opening questions or concerns from any committee members while we welcome our first developer up to the table? Looks like we're good to go.

MS. KHAN: I do want to add, actually, that we have an outline for everyone

to kind of follow further discussion and so if we could sort of stick to that structure. That would be the best way to kind of be able to evaluate each criteria independently and then we'll be voting on each criteria as well. You should have a copy at your table. If you don't, let me know. It looks like -- I actually don't have a copy.

CO-CHAIR KAPLAN: It looks like this. Two pages. So you can kind of follow along and make any relevant notes to yourself as we go along.

CO-CHAIR HALL: And just to reiterate what Adeela just said, this is a bit of a departure from some previous committees. We'd like to walk through each criteria individually and then take a vote on it and then move to the next criterion and vote on that and so on as opposed to having the full discussion and then voting all at one at the end.

With that, we will invite RTI to

introduce themselves and introduce their first measure.

DR. INGBER: There it is. My name is Mel Ingber with RTI and we're discussing the inpatient rehab facility measure.

MS. COOTS: And my name is Laurie Coots, also at RTI.

DR. INGBER: You've already heard a lot about the family of measures that are readmission measures. So this is a member, a first cousin to some of the others, but -- and the name of it is long enough to tell you everything about it. It uses up my entire time. all cause, unplanned readmissions measures for 30 days, post discharge from an inpatient rehab facility.

It's in the all cause family that was discussed, but with a little extra part about the unplanned part of it. There is an attempt to account for readmissions that are, in fact, are expected for this population. There are often things left in people that

have to be removed and things that have to be adjusted and we're trying to remove some of those and that's been done with some of the other readmission measures as well.

The focus is on the 30 days post-discharge, so I think fairly clearly we're dealing with the period in which people are transitioning from a relatively high intensity situation, inpatient, to a less intense often not inpatient at all situation. These transitions, coordination of care are clearly what's going on in addition to whether or not they got good or bad care in the hospital, but the focus is what happens afterwards.

There's been, of course, the hospital -- the acute hospital version of this which you're going to be reviewing again tomorrow and we're in the post-acute part of the world in this case.

We attempted to make it "no-burden-able" and that means we're using the claims-based measure in Medicare. And

Medicare counts for vast, vast majority of these patients, so it's not going to be a bad measure of what's going on in these IRFs. There are very few exclusions to this. When we use -- once we go beyond the part of it being Medicare, we do have an exclusion which is for patients who received only nonsurgical care for cancer. And that was carefully researched by the acute group. And we went and used their information to exclude them here, too. They have a very different trajectory of their condition afterwards. So that was excluded. So generally speaking, aside from pediatric patients, pretty few, everybody is there.

And the planned readmissions were basically reviewed by our technical expert panel and they're in a state of evolution. There's a pretty good list of them now and as you all are aware with ICD-10 upon us, we're reviewing the next generation of codes for these conditions. So we're not -- we actually

were expecting to have to do that, but now it's been pushed off officially by a year.

The measure is a risk-adjusted measure and that means we take into account a reasonable set of patient characteristics that we can get out of claims data and eligibility data from the Medicare system. We use diagnoses. We use whether or not they were surgical patients. We use length of stay, ICU. There's a bunch of patient-specific things that we use. And for the IRFS, we actually make use of the fact that they're already case-mixed groups that the IRFs themselves classify patients into and those have proven to be fairly useful as well. So each readmission measure has similarities, but they're all specifically addressing a particular facility when specific forms of information become available. Reviews the prior acute stay as a source of a lot of the medical information and even information from acute stays prior to that.

A lot of the post-acute care data when you look at claims are coded with a different mindset than what's in the more acute facilities. So we picked up a lot of other kinds of information than we would have if we just used the IRF claims themselves as a source of information.

So we're picking up with the risk adjustment, the morbidity of the patients and trying to level the playing field quite a lot across facilities. The statistical approach is similar to what we've seen in some of the others. Those of you who haven't looked at these before it's not one of these. We look at the observed readmission during the period divided by some expected or some number like that. It uses a model in which you estimate the probability of a patient being admitted or readmitted in a period. And that model contains in it all these patient characteristics, plus indicators for what facility you were in. The patients are

clustered in the facilities.

Then when you try to account for what's happening, what's the facility's piece in all of this, you take the ratio of what you're predicting for the facility including all those characteristics, including the estimate of what the facility effect is and then you divide that by what it would be in the average facility. So it's a different kind of numerator denominator, but it does make some sort of sense. The facility effect is included in the numerator and then removed in the denominator. That ratio is then multiplied by the average readmission rate across this population of facilities, so that you get a number like 1.1 times 13 percent or .8 times -- that would be a good place, .8 times 13 percent, some number like that.

The risk adjusters then are set up in one equation. The facility is in there along with the patient characteristics. The one interesting thing about the modeling which

we're using is that it's the hierarchical, multi-level model statistically. And there is a tendency for this ratio to be a little compressed, depending on the strength of evidence. And what I mean by strength of evidence is that if you don't have very many patients, it's not going to take your raw number which is likely to have a lot of variability because we have very little data on and it will combine that with the average to produce this, what they call -- well, shrunk meaning -- there's a shrinkage estimator that's involved here. And it isn't terribly unusual. We didn't bring it in from outer space. It's certainly been used before.

And the whole thing then is expressed as the standardized readmission rates. We've looked at the rates. Obviously, we look at the actual rates that come out. We look at their distributions. You have probably seen them in our thousand pages of material here. And we have pretty good

differences in the high and the low predictions. You can get very high predictions for patients, very low ones. It's pretty good discriminatory power on that basis. And there is not an unreasonable spread across facilities, although not as wide as you get in some other kinds of facilities. But once you get past that middle quarter, inter-quartile range, they do kind of spread out. So you can see ones that are doing a fair amount better than others. And we are looking for that in the measure.

So that said, that's the overall picture of what we've got. And we're open for discussion.

CO-CHAIR HALL: Thank you. Thank you very much.

Sherrie, you were listed as a discussant. Do you want to take over?

CO-CHAIR KAPLAN: Yes. I was primary reviewer of this and I have a couple of quick questions. First, with respect to

the varying shrinkage estimation modeling problems, your data actually do show evidence that there is a varying shrinkage problem for low-volume hospitals. That is, your low-volume hospitals more closely approximate the national average as opposed to your high volume -- 82 percent of your high-volume hospitals don't fall close to the national average. So that's one and it's come up again and there's a fair amount of controversy as you're aware of how that works out and what the issues are, but can you at least speak to that. I'm going to make a list here for you. So varying-shrinkage estimations is one that I'd like you to at least talk to the committee about, alternatives that you may have tested and not shown us, whatever, that -- alternative approaches.

Second thing is the C statistic. For those of you who don't know, 50 percent is like chance. So you've got 19 percent over chance. Can you tell the committee how common

that -- because you look at that and say whoa, you're only 19 percent better than chance. Tell us a little bit about how those C statistics -- is that normal statistics? Is that normal? It is, because we want you to talk to us about that.

The third thing is in distributional scoring somebody always loses, so you can be 95 percent and if the average is 95 percent and the standard deviation is 5 percent, somebody in that bottom five percent is going to lose. So can you talk to us about the wisdom of distributional scoring?

And finally, dual eligible -- well, your inter-quartile range is 13.0 to 13.9. That's not after risk adjustment. So your risk adjustment did something big to tighten the variance around the mean. And your actual range is 11.1 to 16.1. So that's five percent after adjustment. That means that we don't have a lot of room here to kind of look at who's good and who's bad after risk

adjustment. Five percent sounds like a lot, but in fact, it tightened up considerably after your risk adjustment.

And, finally, dual eligibility looked like it moved your marker around as well. Dual eligibility looked like it had an impact, and the group is concerned about socio-economic status and even though you elected not to include dual eligibility as a risk adjuster since you did look at it, can you guide the committee a little bit about what you would do if it was your zoo for the next round of comments?

MR. LYZENGA: Just a point of process here. I may ask you to hold off on answering a few of those questions for the time being. We'd like to really kind of focus on each criterion as we go through the measure, so we want start off -- I know a few of those issues can probably be covered under scientific acceptability, and a couple of them can probably go under importance as well.

But right now, we'd like the committee to consider the question of whether there is a credible rationale linking the outcome being measured to specific processes of care or care interventions. So that's the sort of question for discussion at this particular moment. The evidence behind the measure.

CO-CHAIR KAPLAN: Okay, I goofed right away.

(Laughter.)

Just to own it.

CO-CHAIR HALL: That's my bad. I told you to start talking --

CO-CHAIR KAPLAN: Yes, don't ever ask -- so yeah, so with respect to the evidence then, the evidence is all based on inpatient readmissions, as opposed to readmissions to the inpatient rehabilitation facility. So you, kind of, cite the same studies over and over and over again. So could you talk to us a little bit about the

evidence?

DR. INGBER: I can't invent it. The evidence pretty much has been where people look, you know. There's a street lamp, and that's where you look, because that's where the light is. People have not spent a lot of time looking at readmissions after post-acute care. It doesn't mean that post-acute care, inpatient care, you should just take the patients and throw them on the street, of course. There's something to be said for transitioning from a high intensity -- and so it's not an acute hospital, but if you go into an IRF, they're going to do stuff to you.

It's a high-intensity program and then you go out and you have to be transitioned somewhere else. And IRFs, in particular, have a very wide range of patients. They have a wide range in the sense that you have stroke patients, TBI patients at the same time you have hip and knee patients that have actually very different

characteristics. The question is, when you send them out the door, you expect them to be discharged to a location appropriate for the kind of patient they are.

So I don't have any particular evidence to show anything, but I think common sense, kind of, tells us there's going to be some sensitivity to how you transition patients, and certainly in other work that I have done which hasn't been distinctly related to this, in which we're working with long term care patients, and the way their decisions are made of who gets readmitted back to hospitals and all that, there's a lot going on in process once you move the patient out. So no, I don't have any information, just words.

CO-CHAIR KAPLAN: So any comments from the rest of the committee?

DR. FISHBANE: Yes, I wonder, since this is the first measure for today if we might speak a little bit about threshold of evidence. I'm concerned. I mean this will

apply, I think, to a number of measures and I don't want to keep beating the same drum.

Maybe we can deal with it now? That, you know, there's a classical article from the New England Journal Of Medicine in 2010 by Chassin et al. This was on using measurement to promote quality improvement, and I thought that there was a lot of wisdom there. They speak to things like the need for a strong foundation of research showing that the process addressed by the measure, when performed correctly, leads to improved clinical outcomes. Strong foundation means more than one study, not an expectation that it will come from randomized trials, or that most of them will.

And the opinion of the authors, and this was a review piece, was that there should be a high bar, one that exceeds the typical standard used, for example, in clinical practice guideline development, which a lot of us have been involved with. So what

exactly is the threshold for the voting on evidence that we'll be doing throughout the day today?

MR. AMIN: Do you want me to take that? Okay. In the Importance to Measure and Report section of the evaluation, so this evidence area, what we're really looking for in terms of evidence is: first, there's the primary bifurcation is whether we're looking at outcome measures or process measures. For readmission measures that are in front of you, we consider those outcome measures. And so all we're really looking for is that there is a process or structure that the accountable entity can have to influence the outcome, that there's a rationale. So we're not looking for a full quantity, quality, and consistency review of the body of evidence.

We're really looking for, simply, a rationale that the facility that you're holding accountable can have some type of intervention that would influence the outcome.

For instance, if you're looking at hospital readmissions that proper discharge planning has demonstrated that that would reduce readmissions. So that's the bar and it's relatively low, I would say, in terms of what we're actually looking for in terms of evidence.

For more process measures, we're looking for a systematic review of the evidence which would include an evaluation of the quantity, quality, and consistency of the evidence that exists to justify the relationship between that process and an outcome that's important to patients.

So for the purposes of our evaluation here, the evidence review should be actually pretty straight forward, and from the fact that if there's an agreed-upon rationale that's provided by the developer, we should be able to move on from this section of the evaluation.

MS. KHAN: You can also, just to

interject, use your algorithms that we printed out for you to kind of understand a little bit better what we're talking about when we are looking at outcome measures for evidence.

DR. FISHBANE: Can I just ask one follow-up question? So are you saying that there's not a need for evidence, but just a need for rationale?

MR. AMIN: Yes. For outcome measures. For outcome measures that are important to patients, we're not looking for a systematic review of the evidence. Karen or Ellen, if you want to add to that?

MS. PACE: No, I just want to maybe explain a little bit of the basis for that. NQF really does have a hierarchical preference for outcome performance measures, because those are generally the things that patients are seeking and providers are seeking to do. And we ask that there be a rationale that outcome is related to at least one healthcare structure process, intervention, or

service.

So the idea is that for outcome performance measures, that there is a relationship to healthcare service and that that is the criteria for NQF endorsement, versus, as Taroon has already pointed out, for the process measures.

DR. FISHBANE: I'm sorry, I'm going to follow up again on this. So a rationale, even if there has been no study of that rationale would be sufficient evidence?

MS. PACE: Yes. I mean most of the rationale that is going to be that there's -- have been some evidence. It's not going to be necessarily a large body of evidence. But the evidence that readmission, for example, and rationale, that readmissions are influenced by discharge planning practices, connecting people to primary sources of care, discharging them in a clinically stable situation versus not. There are a variety of things that are either evidence based or

strongly -- strong rationale that there is some connection and that is what is required for an outcome performance measure.

DR. FISHBANE: I'll be quiet after this one.

CO-CHAIR HALL: Please don't.

DR. FISHBANE: What I'm hearing is then that you do expect studies that demonstrate a relationship. You're not looking for a systematic review necessarily of a large number of studies that have been conducted, but you do want a demonstration of evidence that indicates that there is a linkage between process and improved outcomes.

MS. PACE: So we don't require that they submit that. That obviously makes it stronger. But we do require that they present to you the rationale, the things that they think influence that outcome, and obviously if there are references to cite, studies to cite to include those if possible, but it's not required. Helen?

DR. BURSTIN: To build on that, and some of the logical of this, when this went to our evidence task force a few years ago, was that, in fact, at times outcomes measures come forward, are put out for public reporting and in fact processes follow that focus. And so the idea that you would always, in fact, know the processes up front to influence the outcome as a requirement to putting forward an outcome that's important for the nation was something that, particularly, our Board felt strongly we shouldn't do.

People often cite the example of center line associated blood steam infections, where in fact, the public reporting preceded a lot of the evidence-based processes and in some ways perhaps spurred on the search for some of those evidence-based processes.

So the rationale, at least gives the committee an opportunity to talk those issues through, but we don't want to make it

such that you can't move forward a really important outcome measure unless you actually know for sure exactly what processes will influence that. We recognize that's uncomfortable, particularly for some providers and clinicians among us, but I think that we've seen enough evidence that, in fact, pushing forward the outcome at times could be a really important impact as well.

CO-CHAIR HALL: Paul.

DR. HEIDENREICH: I guess I want to also discuss, I think, what is going to be an issue for most of these measures, and that is first stating this as an outcome measure, I think is a little problematic. I say -- while I agree with the central line infection, I think all patients are going to avoid that, there are some admissions that are good, probably some readmissions that are good, and I think most patients and I would view this as an intermediate clinical outcome. And to the extent it's associated with mortality or

health status, it's a good thing. And so we would want more evidence and say we would went for a true -- which I wouldn't consider this a true outcome measure.

Having said that, I agree that we always can't wait for data if there's a compelling indication. In this case, especially when we're looking at 30-day readmissions, I think there now is some data, and I think it's concerning. I know this application discussed the 24-week Naylor Article Readmission rate that was reduced, but there was a meta-analysis or a systematic review of 30-day readmission published by Hansen Annals of Internal Medicine 2011 of 16 randomized trials. I think four were significant in reducing all cause readmission, but the problem was there wasn't consistent pattern. There wasn't a reproduction that if one study did, say, home-monitoring, the other studies couldn't show it.

And their conclusion was that

there was no particular process that now having been tested was consistently associated with improving readmissions. So I think there actually is some evidence. I just think that it's poor, and I think that -- I'm not sure exactly how to weight that, but I think that -- I think that's going to be an issue for all of these.

CO-CHAIR HALL: Larry.

DR. GLANCE: So just as a comment, an additional comment, I think that beyond looking for a rationale to measure, I think that if you accept that readmission, hospital readmissions, or readmissions to other facilities is a reasonable outcome to measure, and if you can demonstrate, and I think there is evidence demonstrating this in many cases, that there is significant variability across hospitals or across facilities in readmissions, then the fact that you have that variation, after accounting for differences in patient case mix suggests that there's an

opportunity for improvement.

And I would suggest that that is probably the strongest rationale for going forward with the measure. If you think it captures a dimension of quality, and if there's variation in quality across hospitals, then that presents an opportunity for quality and performance improvement.

CO-CHAIR HALL: Leslie.

MS. HALL: Actually, follow up to that is Sherrie's initial comment about the five percent differential doesn't leave a lot of opportunity when the processes are vaguely identified or not pinpointed in a way. So therefore is there evidence enough to determine whether or not this promotes effective change?

CO-CHAIR KAPLAN: I think the performance gap criteria gets to the latter, establishing how much of the variability is mutable is kind of tucked underneath that performance gap criterion, as opposed to the

strength of the evidence. But I think that the concern is that the threshold, the criteria is a body of evidence that would at least suggest a causal link between process and outcome. And for those of us for whom that constitutes a fairly high bar, the evidence is probably not going to be there for -- well, it's going to be stronger for some of these measures in making imprints across hospital readmissions to some part of the -- readmission to some part of the hospital is probably less of a gap than -- a conceptual gap than between, for example, home health care and some of the other kinds of measures we may be considering today.

So I think that moving the target of inference still within the hospital is not as much of a leap as it is for some of these others, and therefore I was sort of interested in how that worked for people who were developing the measure, what kinds of evidence they're really relying on the readmission to

the hospital, versus readmission to one part of the hospital in the evidence basis for their analysis, a little more development about the thinking about the causality link there. But again, that is sort of beyond now.

We're taking it a little bit beyond what NQF guidance is giving us. They really want us to look at some plausible link between process and outcome, and even if the evidence, what we would consider causal, statistical evidence may or may not be there.

CO-CHAIR HALL: Paula?

MS. MINTON-FOLTZ: Just a point of clarification. Is the accountable party the discharging hospital? Because there's quite a rigorous -- or is the accountable party the rehab? Because there's quite a rigorous vetting of facilities to even take these patients from hospitals. Thanks.

CO-CHAIR HALL: I don't know -- actually clarification with our NQF colleagues: I know we're supposed to be

talking about evidence. It's hard in my head. I feel like we're talking about everything, which is hard to avoid. So I don't know how we move forward voting category by category, versus open discussion.

MR. AMIN: Yes. I think, again, for the purposes of kind of keeping things structured and making sure that the committee can move along in a pretty structured way, the important thing to keep in mind is that the Importance to Measure criteria is more really not necessarily into the measure specifications, but that the measure outcome, which is looking at inpatient rehab facilities into the hospital is that -- that is a help outcome, and is there credible rationale for the relationship between the structure, process and that outcome. And I think we should just kind of discuss that.

If there's any more comments, or go to vote and look at whether there's a performance gap, conceptually, and then vote

on that. And then we can get into the science of acceptability, which is where the majority of these comments are really focused.

CO-CHAIR HALL: And so Paula, the accountable entity is the inpatient rehab facility, but it almost feels like that's part of the scientific specification. So we'll return to that.

I certainly endorse everyone's concerns. There are 23 of us here in the room as experts and another on the phone, and in many ways we are to sit here like a jury using our best judgment and finding some reasonable level of comfort. That's about the only guidance I can pull out of having done this several times before. So we'll try to refocus on the category of evidence, but Carol, I see that you have a question.

MS. RAPHAEL: I just wanted to say two things from the point of view of the MAP Coordinating Committee and Work Group that has struggled with this. One is I do believe that

there is great variation in performance and there is an opportunity for significant improvement. And it isn't one process that's the key to unlock this door. We have to work in many dimensions. It has to do with communication. It has to do with how medications are managed. It has to do with followup, with primary care and specialty care as well as the discharge process and the degree to which the patient and family are prepared for what comes after.

And secondly, this may not be methodologically correct, so I plead guilty, but from the work group's point of view, we really need to move toward uniform measures, because in the post-acute care sector there are different assessment instruments. There are different case mix adjustments. There are different payment systems and when we're trying to rationalize why people get placed in certain places and what's the clinically

appropriate place for them to be, we have to begin to have some uniform way of measuring these different sites and the performance of the different sites.

And when we looked at readmission measures they were all over the place and it really handicapped and inhibited our ability to take a look and compare the different sites of service. So those are my comments before we move ahead here.

CO-CHAIR HALL: Thank you, Carol. Points well taken.

So before us we are asked to judge whether there's evidence that this intermediate outcome as Paul has highlighted, this outcome, which is also an intermediate outcome, has an adequate conceptual link in our heads to something that the accountable entities can do to improve this. And that will be tightly linked to the second category of vote, which is whether there's also an opportunity to improve that. I think that's

where we are. We will get back to the scientific acceptability shortly as well.

I know Sherrie has questions. I have some scientific-type questions. I'm sure many of you do as well.

MR. AMIN: Yes, let's move to voting. Zehra

MS. SHAHAB: So before we start, just a few, I guess, like instructions. Please point the clicker towards me, towards this laptop when we're voting. Does everyone have clickers? I believe I handed it to everyone.

Just click on the number for the vote. You don't need to click send. And you will have 60 seconds to vote. You may click the button multiple times, but it will only register once.

(Laughter.)

CO-CHAIR HALL: Zehra, is it going to register your last click or your first click?

MS. SHAHAB: Yes, your last one.

CO-CHAIR HALL: And are you going to tell us what numbers mean what?

MS. SHAHAB: Yes. I will read out the numbers. I will read out what we are voting for, and if anyone has any questions.

CO-CHAIR HALL: And then will you ask Paula for her vote or will she be sending that separately?

MS. SHAHAB: Paula is going to be sending to me via the chat and I will enter it for her.

CO-CHAIR HALL: Thank you.

MS. SHAHAB: So is everyone ready to start? First, we're going to vote on importance to measure and report, 1a, evidence. One is yes and two is no. So I'm going to start the timer and --

CO-CHAIR KAPLAN: And that means that the evidence criterion is met in this case for the health outcome that there's a rationale that there's a healthcare structure

process, intervention, or service linked to that.

MS. SHAHAB: Does everybody have that? One is yes, the criteria is met; and two is no.

CO-CHAIR HALL: Paul.

DR. HEIDENREICH: It sounds like you're saying we should not -- we should only be going on number one, healthcare rationale and the second line should be removed from the slide. Is that basically what you're saying? That this is a mistake slide?

CO-CHAIR KAPLAN: Yes, yes.

MS. SHAHAB: So, 60 seconds begins now.

(Pause.)

CO-CHAIR KAPLAN: When you send, a green light shows up in the upper left corner.

MS. SHAHAB: So we're waiting for one more. That is all votes and voting will close now.

So 21 people voted yes and 3 no.

So we will move on to the next vote.

CO-CHAIR KAPLAN: So now we're on to performance gap. And I'm going to just query the measures developer about the after-adjustment issue. So initially, your range was considerably greater. After adjustment your range went to 11.1 to 16.1. And the inter-quartile range is after adjustment 13.0 to 13.9. So help us understand, because this gets at the issue of causality and mutability. Now you've got after adjustment, 5 percent difference between the lowest and highest folks.

Is that credible, in terms of -- in your opinion from the data you've looked at, etcetera? Is that credible with respect to how much variation is mutable? Is there a big performance gap here? Are we looking at, still, noise and if we were only able to risk adjust tighten we would sit everybody right around -- right up against the mean?

DR. INGBER: There are judgments

to be made here, of course, when you say big and small, that term is relative to what, of course. And I will point out that this particular facility type has a relatively low readmission rate. The nature of their patient mix is very different from the one we're going to see a little bit later, the long-term care hospitals.

And so the range you're going to see will be reflective, to some degree, of where that mean is. And the mean at 13 percent is a good deal lower than the roughly 16 percent for acute care hospitals and 20X percent for the LTCHs.

So you have to put it into that context, as to how much it can be moved. Five points, is it big or small? I don't think that if we were to say the difference of five percent of the patients is -- and that's not everybody. But the difference is not negligible and does allow for some people to be doing a lot more in that period,

apparently, to lower their rates, assuming that the risk adjustment is working properly, that we could have some of the others move up, if not to the top, at least to the middle two points. And two points, I don't sneeze at when it comes to readmission rates.

As a percentage of the 13'ish number we're talking about, it isn't totally insignificant. So I mean I can't convince you of what's good and what's bad. The other issue will come about later, which is, relative to how you use the measure and the distribution in terms of saying oh, you're going to get unrewarded and you're going to get rewarded. There are lots of ways to use the numbers that don't just arbitrarily, and I really dislike the method of just looking at your decile, because deciles can cover 1 point or 10 points or 50 points. So without getting into that right now, the issue of how narrow it is is somewhat related to the fact that you can use the information in different ways.

You're not fixated necessarily on one quartile or another quartile.

CO-CHAIR KAPLAN: Questions from the group? Thank you for that. Questions from the group? Wes?

DR. FIELDS: Just a quick point. I think if you adopt this as is, with a narrow spread, and you endorse the concept that it's really a measure of the facility's overall performance, which I think is what we're getting at, if I was only one or -- if I were one or two quartiles below the hospital, the acute rehab facility and I wound up a quartile or two below my competitor, the rational response would probably be to change the cohort of patients we accepted into our acute rehab facility, which really wouldn't be a process improvement. It would be a way of managing service lines.

I think, ultimately, part of my distaste for this one, or my discomfort with it, is that I really would love to know how

facilities are doing by service line. I think you've got a better chance of getting at what you want from this in terms of performance and reducing of variation. So I'm speaking with some level of discomfort about having this as an outcome measure for facilities with such a narrow spread, when the easiest way to perform at higher level in the future, if we adopt this as is, would be to actually restrain access to higher risk rehab services, say, for stroke patients.

I mean if I want to do well, I probably just kind of rehab hips and knees and make the hospital across town deal with the strokes. And I don't think that's the intent of NQF, or CMS for that matter.

CO-CHAIR HALL: I know our developer wants to comment, so please make a note. But Steve, do you want to --

DR. FISHBANE: Just a brief comment. When you have a measure, from any source, where from the unadjusted to making

adjustments for covariates and there's a lot of movement after that adjustment. I'm very concerned that because we don't measure all variables or the richness of all variables that, with a lot of movement, it indicates that there is a high probability that there's a lot of residual confounding that's probably not being captured. And I'd just be curious in this case about what the developers' comments might be.

CO-CHAIR HALL: Before we invite that, and I think your point is correct, and it's an example of a point that addresses not just the performance gap that we're presented with, but then also the scientific spec of the measure. So we'll certainly return to your concern, Steve. I know Taroon wanted to make a clarifying --

MR. AMIN: Just a quick clarification, Chris, in terms of the criteria that we're looking at variation in the performance, but we're also looking at whether

there's an overall lesson, optimal performance at an absolute level as well. So it's not purely just variation, but I just wanted to add that into the discussion.

CO-CHAIR KAPLAN: Could you also comment, or Helen, about the policy with respect to NQF on the use of these measures following approval and endorsement, versus how they're implemented and used etcetera versus the endorsement process itself?

DR. BURSTIN: Yes. That is an important issue. And actually, it's very helpful today, because you have several members of the MAP at the table so the Measures Application Partnership, at least currently, is the group that after you guys do the okay on the scientific acceptability of these measures, we'll then make a determination about which programs they're appropriate for, in terms of use. So we try to keep them separate, try to keep you guys really focused on the science. It is often a

very difficult task.

I will also say that one of these were actually actively doing which, Taroon is helping us lead because we actually are doing an internal lean effort to try to, in fact, better integrate the work between those groups, so that it's very clear when things reach the MAP exactly what the science is and try to make it less about an artificial divide, and even considering something moving forward where we will actually move away from binary yes-no endorsement and look towards some way of differentiating levels of endorsement for intended use. That's still a work in progress.

CO-CHAIR HALL: I don't see any other cards up, so I'll ask our developers if they want to make any brief comments about what they've heard on this topic, please, performance gap.

DR. INGBER: Yes, on the topic of gaming the system, as it were, by picking only

the patients who can play football or whatever, we did attempt to address it by actually including the service lines in the model, so that the stroke patients, or the traumatic brain injury patients, we take into account their probabilities, and to the extent that the hip and knee folks have a much lower probability of readmission, it will self adjust and so it's easy to game as it would appear.

Now, let's see, the other issue was, I thought, somewhat related, but I think a lot of it has to do with the risk adjustment approach of things. And the fact that there is randomness left. I don't know about you guys, but I've been modeling health stuff for a long time. There's a lot of randomness. You cannot adjust for it all. We don't have the luxury of having everybody's lab values, DNA, and everything else we can have in the system right now, and so the measure is using what information it does have and some of the

uncertainty is, of course, taken care of by the very nature of the modeling which has to do somewhat with the fact that small facilities look like they're near the mean, but even if you didn't make them near the mean by using another statistical approach, which we can talk about, they would come out with such large standard errors that you wouldn't know what they were no matter what. You couldn't say that they were very different one way or the other from the mean.

So there are different ways of getting to the conclusion that says, I don't know. There's not enough information here. So we do adjust for what we can and I mean it's my judgment this particular measure fits in the acceptable range from what I've seen of models that work in this dimension.

So I think the risk adjustment such as it is, we put in a lot of variables and we went through a lot of clinical review and what not. So all of that is in our pages,

but it's working. I will not say it's working perfectly. And I certainly have not had the experience of any modeling yet that did.

CO-CHAIR KAPLAN: We're going to come back to this, the scientific acceptability. This was on performance gap. So your position is that there is a performance gap, and even though there is a fairly tighter mean for these kinds of facilities, there remains a performance gap between good and poor and high quality facilities. Okay, so are we ready to vote? Okay, so.

MS. SHAHAB: So we will vote on 1b, performance gap. Date demonstrated considerable variation or overall less than optimal performance across providers and-or population groups. One is high. Two is moderate. Three is low. Four is insufficient. And the time begins now.

(Pause.)

We're waiting for one more

response. CO-CHAIR KAPLAN: Do it again really fast just to make sure you got your thing in.

MS. SHAHAB: So we have all 24 votes and I'll close the voting. Three voted high; 13 moderate; 8 low, and zero insufficient.

MR. AMIN: Again, I'll just point out for this next section on high priority, this is really a conceptual question of whether or not the measure addresses a high priority area of looking at readmissions from inpatient rehab facility. Again, it's much more of a conceptual question, and I think we've had some of these discussions already, but if there are other conversations I welcome them.

CO-CHAIR HALL: So comments on whether this is a high priority. I'm not seeing any comments. Questions? So we'll move to vote.

MS. SHAHAB: We're going to begin

voting on 1c, high priority. Addresses a specific national health goal priority or data demonstrated a high impact aspect of healthcare. One is high. Two is moderate. Three is low. Four is insufficient. Time begins now.

(Pause.)

We're waiting for two more responses.

(Pause.)

Ten seconds.

So we didn't get all 24 votes. So we will have to revote.

CO-CHAIR HALL: Paulette, don't forget to send in your vote.

MS. SHAHAB: I know. We're missing two votes. The results are 6 high, 13 moderate, 3 low, zero insufficient.

CO-CHAIR HALL: So we'll return now to scientific acceptability and Sherrie, you might want to start by restating your concerns or questions.

CO-CHAIR KAPLAN: I think I'm probably going to restate some of what's already been said, but one is the varying shrinkage estimates potentially overestimate the low- volume hospitals and your data sort of show evidence of that, so can you come back to like 40 percent, like twice as many of those have an estimate that was near the national average, as opposed to 82 percent of the high volume hospitals were way away from the average. So can you come back to the varying shrinkage?

I know we're not going to resolve this, but if you have some information that would help us understand, as you said, the low-volume hospitals are going to have a huge standard error around their means, so interpreting their differences is going to be problematic, but maybe did you look at them compared to each other?

The second thing is is the performance adjustment, you know, it's going

to shrink and just what you were talking about earlier how much is in this. I think you had 204 variables in your risk adjustment model, so I mean adding the 205th, 6th, and 7th may not be a plausible thing to do, and may not have helped us understand what residual remains after you've adjusted for everything else. Is it true score variation.

And finally, distributional scoring is always going to put somebody on the bottom, so if you were indicating that you may have done some threshold analysis or something like it, if you could give us a sense of if you do those kinds of analyses are we getting around to the point where the distribution is so tight that we might want to consider some other kind of scoring than distributional scoring?

DR. INGBER: The issue of what happens when you use these kinds of models, it is the nature of these multi-level models to assume that the cluster, which is the

facility, is distributed in a bell-shaped curve around some average and by building that in, you basically have statistically said well, this particular facility -- I was using the word strength of evidence earlier -- that has only 27 patients and this other one has got 300 and what not.

And so what's the probability that they're really out as far as they appear to be? And we had some pretty pictures that we did submit to you on what the raw distributions were and there was some with 100 percent and some with zero percent, and none of which did we find believable as a good characterization of a facility.

So the nature of this kind of estimation is to say we're going to take into account how little evidence there really is for you being that far out and bring you in. So the numbers do tend to bring them.

There's a danger, then, if you use a pure distributional that some larger

hospital that's out here is declared to be particularly good or bad. This other little one has been made to look better or worse. It goes both directions. And so therefore, you're making a bad judgment.

So the question really is: how do you apply what you've got? And nothing in the measure tells you how to use it. What we did find, if you look at the section that talks about who is statistically different from the mean, is that fewer of the small facilities are -- not all of them are out there, but fewer of them make it to be statistically different from the mean. And the bigger ones, justifiably, you can say are different from the mean. Now is the mean your ideal in any way? It becomes a different question.

So where is your benchmark for what you're going to consider good or bad or how are you going to make use of this information? And one way that it has been used. I don't know, I mean CMS has to decide

from a policy point of view, and I don't know if Paul wants to speak about it, but how these are going to be applied. But, naively, you don't want to just rank folks and say you're good, you're bad, when you know you have confidence intervals that are saying well, you look good, but the confidence interval is a mile wide and so we're not so sure.

So how you use it has to be done carefully, and I can't tell CMS what to do with it. I just work for them. So -- they pay me, really.

(Laughter.)

So at any rate, yes, I say use these things with caution and understanding how the numbers are derived. We did supply for you a whole little section that was a supplementary doohickey, and it had in it a picture of what if I used the different measure? And it shows what the raw rates are for the smaller, the middle size and the bigger ones. I didn't take the smallest or

the biggest to do these.

And then what happened with a so-called fixed effect model which makes no assumption as to what the distribution of the facilities is, it just says whatever it is, that's -- after we risk adjust you, where does it go? And these measly-looking charts that you have do show that there's a very wide spread of raw rates that are kind of all over the place. And that risk adjustment does tend to bring them toward the center. In some cases it even crosses the line from a below average to an above average, or vice versa. So the risk adjustment is fairly effective working with it.

Now, if you wanted to use a pure fixed effect model we don't have in here the sort of error bars that that tell you that yeah, you see a different number than we're telling you as the final number with the shrinkage estimator, but that error bar would be very, very large. And you can't quite see

that. So whether you would actually draw a very different conclusion about a facility from this, you can just see there are differences and you can see that the shrinkage is much greater for the smaller facilities, as just what you'd expect because they have relatively weak evidence.

We are using two years of data which made the facility sample sizes larger. We hesitated to use three years, because there's always the straight off between timeliness --

CO-CHAIR KAPLAN: Sorry, for the record, can you tell us which page you were looking at when you were looking at the diagram?

DR. INGBER: Okay, this particular area is in something called "Assessing the Impacts of Risk Adjustment and Shrinkage," and it's a relatively small addition that we sent.

And in there we have lots of numbers which say how much change was caused

by pure risk adjustment and how much change was caused by the shrinkage in addition. So we do present to you with some information on that to work with and we do present the differences between what's going on with the larger facilities and the smaller facilities so you can say yes, of course, there is small evidence, you get larger shrinkage and you also, sometimes, have wilder raw effects. You have to be really careful with raw effects. So the risk adjustment is helpful.

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

Larry, did you have issue?

DR. GLANCE: I think that the use of shrinkage estimators is really a cross-cutting issue, and we went over this and spent a lot of time on this last time around. Certainly, the use of shrinkage estimators gives you more stable estimates of performance, possibly more reliable as well.

The disadvantage of using

shrinkage estimators is that you are implicitly or explicitly making the assumption that low volume providers are performing in an average way. So that from the standpoint of public reporting and accountability, shrinkage estimators may not be completely ideal, because we know that in many cases there is a volume outcome association, and that low volume providers have worse outcomes than higher volume providers. We also know that there is some literature in the case of readmissions to suggest that is the case.

On the other hand, if you don't use shrinkage estimators, you can get some fairly wild estimates of provider in hospital quality. So you're kind of stuck between and this is a very technical term, between a rock and a hard place.

(Laughter.)

And we spent hours talking about this last time. So --

CO-CHAIR HALL: And I'd like to

build on that, what we're building on Larry's comments. When you think about using shrinkage in this type of application, performance assessment potentially for reward and punishment, what we're really saying is we're less willing to penalize people when the information, quality or volume is low. Now that comes at a cost. The cost is, as Larry said, that means some people who might, should be penalized are not going to be penalized. That's equivalent to Larry saying we're going to give those small-volume hospitals the benefit of the doubt and push them back to the average. What we're really doing is saying, because the volume or quality of the information about those facilities is low, we're not willing to penalize them at present time. That's the implication of what we're saying.

CO-CHAIR KAPLAN: And this again transitions over into the use of these measures after the fact and I think, Helen, is

that accurate the MAP folks have and are dealing with this issue? So what we're looking at is basically for the measure as proposed, is it scientifically sound? Any other comments from the group on what that constitutes?

CO-CHAIR HALL: I don't see any cards up yet, so I'll throw up a couple of my concerns where first of all, just to highlight, this is unplanned readmissions. The algorithm is very similar to the unplanned algorithm that applies to several other measures coming out of the CMS, so that's a pretty common algorithm that we have some exposure to and if you have any questions about that, please feel free to ask our developer.

The time period here is proposed to be 24 months and in the past deliberations people have raised whether a 24-month evaluation is really an evaluation that you can act on in a timely fashion. So that's one

concern. Obviously, that's done so that the volume numbers will be higher than they would otherwise be. But that's a tradeoff.

This also in the information submitted by the developers, the reliability is really adjusted with split sample consistency, rather than sort of a signal to noise emphasis, if you will. The split sample consistency seems good. The correlation, the inter-class correlation numbers are only moderate at best.

And then perhaps the one question that I do need answered, which is open for me, is why transfers from the IRF to short stays were excluded within a day? So in other words, we're saying if you're discharged from the IRF, and you go home and you're home for more than a day, you're not eligible to be readmitted. But if you're transferred directly from the IRF, or if you go home and within a day go to a short stay care, you are excluded. Those seem to me to be just as

powerful indicators that something went wrong. So I was curious as to why the immediate transfers back to short stay or the transfers after only one day or up to one day were excluded?

DR. INGBER: Yes, on that: one day is a standard way of handling the issue of Medicare data where sometimes you cross to midnight and it's really a transfer, and we didn't want to get into that debate about whether or not it was really a transfer, but the clock ticked and it went over.

CO-CHAIR HALL: Then it reduces to why your transfers are excluded?

DR. INGBER: Yes, exactly. And you can think that maybe there's another measure that might be in the works to take care of that. But we wanted to make sure that we were looking with people demonstrably in the community one way or another.

DR. AUGER: So a couple of questions. One is about the planned exclusion

criteria. There's something that mentions that they were modified, I believe, so I didn't know how so or what the justification was for that, and then how we know whether or not it remains to be a valid sort of algorithm for the modification?

DR. INGBER: The modification was done having started from what was done from the acute hospital measure. They had done a lot of work looking at planned readmissions and we don't throw out perfectly good work. But we did bring it to our own technical expert panel and they know what patients they see. They know what patients they're sending back to hospitals, so we did add, and what happened subsequent to that is that the other measure, which we'll be seeing this afternoon, actually, in their considerations came up with some of the same things we do. So it's something that is always a judgment call, and our folks just added a few more procedures to the ones that were excluded to begin with.

And it is understood that if some of these procedures occur in the context of an acute admission, as noted by the diagnosis on that next inpatient readmission, you can have something done just incident to some acute admission, and because it's done that way it doesn't get you off the hook. So if you really went back for your heart attack or whatever it was that happened to you, and you happened to have some wound debridement done, that wouldn't take you off the hook. So it was kind of done by just a technical panel. There's no magic to working this one out.

CO-CHAIR KAPLAN: Thank you, Helen, and we're ten minutes into our break here, so just for those of you who need caffeine, we're going to work through this and be a bit more concise as we go on. This is kind of our first measure so we're working a little bit harder here.

Helen?

DR. CHEN: Just quickly. I

understand that the issue of transfer, going back to Bruce's comment might need to be evaluated as a separate measure. However, this exclusion, and similarly in the LTAC measure troubled me, because I think it could actually lead to some potential for people transferring patients who are marginal in terms of their ability to be safely discharged towards the end of their stay, knowing that it wouldn't count against them as a readmission. So there's a potential for an unintended consequence and some degree of gaming the measure.

CO-CHAIR KAPLAN: Thank you, Helen. I think monitoring that might come, if approved, that might come under what gets monitored in the endorsement in the three-year run. Kathy, did you still have an issue?

DR. AUGER: One other quick question is about the heterogeneity of ages. So clearly Medicare eligibility, over 65 is included, but also younger people with

disabilities, therefore there might be more younger people that have been in inpatient rehab facilities. I believe you guys adjusted for age, but I'm wondering if this sets up sort of a situation where you may have some rehab facilities where you're comparing some rehab facilities that primarily deal with things like stroke rehab, and that kind of thing, to facilities which are very different which primarily deal with rehab for like younger patients with other sorts of disabilities.

And I guess part of it is, sort of, a I don't know whether or not there are different types of facilities like that, but also are those appropriate hospitals to compare?

DR. INGBER: Assuming that's a question, we do have the risk adjustment that, as you said, does have age in it per se. But if they tend to fall into one of the CMGs that is more characteristic for younger people --

we are doing our best to adjust for what kind of patient are you, on top of just how old are you, and not only that, but what your diagnoses were. We mentioned 200 risk adjusters. We actually worked very hard to pull in -- there are over 100 CMGs out there in the world.

So and most of them came out, we had to condense a few because they're quite small, some of them even in the national data. Guillain-Barre is not exactly your most popular disease. So we did collapse a few of them. But to the extent that particular age groups of patients are found in certain CMGs or have certain things, we attempted, I can't swear to you that we have taken all the variants out, but we tried.

CO-CHAIR KAPLAN: Thank you. I have just one more quick question for you and that is would it be your guidance, for example, the standard error of measurement is the standard deviation times the square root

of one minus the reliability. And what it does is it gives you guidance, like a confidence interval does, on how much noise there is in a measure. Would it be possible if this measure is endorsed to kind of, as we go forward, start providing the standard error of measurements around these measures for whatever groups we're kind of comparing?

DR. INGBER: It's just a little tricky, because you can produce standard errors. But these kinds of measures, the way we got confidence intervals was entirely differently, because the very nature of the ratio estimator led us to have to do the bootstrap because it doesn't lend itself well to the standard calculation. So we could produce numbers, but we'll have to be careful about how we use them.

CO-CHAIR KAPLAN: Any other comments? Are we ready to vote? So what are we voting on first? So first we're voting on reliability.

MS. SHAHAB: 2a, reliability including precise specifications, and 2a2 testing. So one is high. Two is moderate. Three is low. Four is insufficient. And the time will begin now.

(Pause.)

Will everyone just press it one more time, please? We're trying to get to 24.

(Pause.)

So we have 23 votes and the results are 3 high, 16 moderate, 4 low, and zero insufficient.

CO-CHAIR KAPLAN: Okay, now reliability spoke to the issue of consistency, reproducibility, etcetera, etcetera. Reliability is consistency. Validity is accuracy. So reliability: is it reproducible. And I love to give the example of my bathroom scale. I step on my bathroom scale every morning. It tells me exactly the same answer. I love my bathroom scale. It is completely wrong. But I love the answer it gives me. So

validity speaks to the issue of was my bathroom scale correct? Which is different from reproducible. So we're about to vote on validity.

MS. PACE: And validity also includes the -- what we talk about as threats to validity, so issues about exclusions, risk adjustments, etcetera come in this section.

CO-CHAIR KAPLAN: So is it right or is it wrong?

MS. SHAHAB: So now we're going to vote on 2b, validity. One is high. Two is moderate. Three is low. And four is insufficient. Voting will begin now.

(Pause.)

CO-CHAIR KAPLAN: Do it again.

(Pause.)

MS. SHAHAB: Okay, so we have all 24 votes. One, high; 16, moderate; 6, low; and 1, insufficient.

CO-CHAIR KAPLAN: Okay, now we're going to talk about feasibility, and I,

basically, don't have any comments about feasibility, other than they use the data that was existing, so at least we know with those data you can do something. But if there's any other comments from the committee or questions, is it feasible? Okay, we'll vote. Let's vote.

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

(Pause.)

We have all 24 responses. Eighteen, high; 6, moderate; zero, low; and zero insufficient.

CO-CHAIR KAPLAN: Okay, we're on to usability and use. I did lie to you. We are not ten minutes into our break. The break is actually 10:45, but it was a strategic move. I'm going to declare it as a strategic move.

(Laughter.)

Technically, as Bruce pointed out, we are behind, because we started early at 9:20, so okay. So on usability and use, it's sort of where potential usability and use are, are public reporting, quality improvement, and quality improvement both for purposes of benchmarking and for institutional internal use where the potential uses of this thing -- I think they were specified, but if not, in the document itself. Other comments on usability and use? Tony?

DR. GRIGONIS: Yes. I think in your response to one of the issues that was raised in the preliminary analysis, you've noted that these measures, and this will come up -- I'll be doing the LTAC one,, so it will come up in the long term acute care hospital measure also. And that is this measure is not really intended to track change over time. And if that's correct, is that really problematic, since what we're trying to endorse is a quality measure that can measure

improvement?

DR. INGBER: Whereas it does not track change over time explicitly, you just can't take the number in Year 1, Year 2, Year 3, and easily track it for a facility. It has the virtue of accommodating itself to changes in the way people are treated over time, because each time you're looking at a new world each year and new treatments come up, new ways of doing things come up, so that it actually allows you to say, given the current situation, you're doing a lot better over -- whatever, than other people.

The way you can track time is a little bit different because it requires sort of tracking that raw number, because that raw number of 13 percent could go down to 12 percent let's say, and then everybody is clustered around this new mean. So you get kind of relative to the mean of what's happening, how are you doing? So yes, each facility over time, if you just track the

numbers without knowing what's going on overall, it's not really great, but it does accommodate.

For instance, I'll just make this because it's an obvious one. You mentioned LTACs. Something strange is going to happen in the LTAC world in 2016, according to the new rule that just came out this week. And it's going to change the patient mix. Well, the measure will accommodate that because it gets reestimated each time, so that you don't worry that the patient mix is suddenly changing, and the same thing could happen in the world of IRFs in which they have some administrative change in some of the rules, 60 percent of this or that, you know.

And it accommodates itself to that. So there are pluses and minuses. I can't say it's great for tracking over time, but I can say, because of the way it works, it accommodates itself to the change in the universe over time.

DR. GRIGONIS: But just to clarify, that universe would be a two-year time period. So, I'm bringing this up as a cautionary note of how the measure may be used after it's implemented. For example, if it's used to change payment structure or penalty or something like that, I wouldn't want to see it misused by shrinking it to a year or six month time period or something like that, when you've clearly established that its utility is based on a two-year sample.

DR. INGBER: Utility is based on having enough data. Whether -- you want to do it for one year, you can do it on half the facilities reliably, and say anyone with fewer than N patients we can't really do this, and so administratively we're going to throw you out of the picture. There are a lot of ways of sort of dealing with that.

We estimated it to get the full distribution and to get their sample sizes up and actually, we don't have a lot of tiny

facilities. We managed to get a bunch of them up, and that's why we have as many as we do that are statistically significant from the mean, even in the small group. We were kind of surprised to see that the confidence intervals actually said they're not average to 20 or 30 percent of that little group at the bottom, as opposed to everybody down there is average looking.

We don't really mean they're average when we say they're average. We mean we can't differentiate them from the average. So I do want to be a little careful.

CO-CHAIR KAPLAN: Paul.

DR. HEIDENREICH: Just in the voting on 4c with the term evidence, and it's in the further description, evidence of unintended consequences. Now in the cases where we haven't actually implemented it and we don't have the evidence yet, does that mean you don't consider it, or would it be if there's a rationale for unintended

consequences that we should consider, even if there is no evidence?

MS. PACE: This really applies mostly to measures that are coming back for endorsement maintenance, rather than initial measures. But certainly it's something that you can talk about if you think there's a big issue with potential unintended consequences.

DR. HEIDENREICH: The reason I bring it up is because the way 4c is described, it really is sort of a summary of your overall view of it is the benefit of having this measure. Does that outweigh the potential harms of having the measure? So in some ways it could very well apply to all the measures. Maybe that wasn't the intent.

MS. PACE: The intent, and you're right, you specifically noted that it said evidence, so it was mostly intended for endorsement maintenance. But again, you know, if there's particular issue you can certainly raise it, but overall, after you go through

each of the criteria you'll do an overall vote on the measure, but certainly this would be the place to discuss it, but that criterion is mostly relevant for the endorsement maintenance measures.

CO-CHAIR KAPLAN: So by that you mean empirical evidence after the application of the measure of unintended consequences, not anticipated unintended consequences that we could all conceptually kind of --

CO-CHAIR HALL: I don't know. I think that we're all called upon to use our judgment, and if there are concerning potential unintended consequences in your mind, then you're obliged to consider them in your voting. That's how I feel. Pam?

DR. ROBERTS: I just have a clarification question, based on what Tony said. So if it went forward, it would go to the MAP for the implementation issues, is that correct? So the issue that Tony brought up would be, if this does get endorsed, it would

be a really important issue for facilities to understand for future implications because it's buried in there and it's not something that people readily think about.

DR. BURSTIN: When measures go forward to the MAP, we'll, of course, we do try to make sure they stay true to the endorsement, and if the endorsement is around 24 months of date, then that would need to stay true to that.

CO-CHAIR KAPLAN: So the committee could make a recommendation to MAP that they consider the issue that Tony raised or no?

CO-CHAIR HALL: Tom?

DR. SMITH: I think I'm summarizing what people are saying, so for 4a and 4b, it doesn't say if new credible rationale or plan. So wouldn't the implication be for 4c that potential unintended consequences ought to be considered as well?

CO-CHAIR KAPLAN: I think that's

reasonable. That's fine. Any other comments? Ready to vote?

MS. SHAHAB: We're going to vote on usability and use. One, high; two, moderate; three, low; four, insufficient information. And the time starts now.

(Pause.)

CO-CHAIR KAPLAN: Do it again.

(Pause.)

MS. SHAHAB: Okay, we have all 24 votes. One, high; 14, moderate; 8, low; and 1, insufficient information.

We can go ahead and also vote of overall suitability for endorsement. Does this measure meet NQF criteria for endorsement. And note, this may not yet be a recommendation for endorsement. Final recommendation for endorsement may depend on assessment of any related and competing measures. So one is yes and two is no. Time begins now.

(Pause.)

Two more.

CO-CHAIR KAPLAN: Do it again.

(Pause.)

MS. SHAHAB: We have all 24 votes. Sixteen voted yes and eight voted no for measure 2502, all cause unplanned readmission measure for 30 days post discharge from inpatient rehabilitation facilities.

MR. AMIN: Before we break, I just wanted to point out one thing related to voting, because I assume we're going to get here at some point. Adeela, could we scroll back to the slide related to the gray zone voting?

So one of the enhancements that we've done in NQF over the last year and a half is -- one of the challenges that we have is sometimes when the measure just falls right below the 50 percent threshold, the conversation on the measure didn't move forward and the measure didn't go out for public comment. And so what you'll find is

that what we've implemented now and approved by the Board is that if it falls between the 40 and 60 percent range, we'll continue to move the measure forward to the evaluation of the remaining criteria and we'll move the measure forward for public comment, just so that there's not an absolute 50 percent threshold.

Obviously, this process includes this committee deliberation, public comment, voting by the members, and then we'll ultimately make a decision with CSAC on the measures if there are measures that fall within that range going forward. So you will hear, potentially, over the next day, I'd be surprised if you didn't, if we describe something as falling in the gray zone if a consensus is not reached.

CO-CHAIR KAPLAN: We were on track for our break until that last comment.

(Laughter.)

DR. FISHBANE: Gosh, I'm costing

us the break. Just the second one, 40 percent combined for high and moderate, so that's kind of a composite of taking all of the votes together. Thanks.

MR. AMIN: Thank you.

CO-CHAIR KAPLAN: Break. Eleven o'clock, start again.

(Whereupon, the above-entitled matter briefly went off the record.)

MR. AMIN: On the next measure, which is 2512.

(Pause.)

CO-CHAIR HALL: So we will move on with Measure 2512, All-Cause Unplanned Readmission, 30 days post-discharge from LTCH. We'll invite RTI to give their two minute introduction to this measure.

MS. COOTS: Thank you. My name is Laurie Coots from RTI, and I'm going to give a brief overview of the measure. So this next measure, the All-Cause Unplanned Readmission measure for 30 days post-discharge from long-

term care hospitals or LTCHs, is also among the same set of readmission measures specific to post-acute care that RTI is the measure developer on behalf of the Centers for Medicare and Medicaid Services.

So this particular measure is quite similar to the last measure that we reviewed, in terms of the methods and the statistical approach. In terms of the background, given the large proportion of readmissions among the Medicare population, and in particular for beneficiaries post-discharge from long-term care hospitals, CMS has proposed to monitor the readmission rates in order to reduce rates that are inappropriately high, with the aims of improving patient safety and quality of care.

This particular measure estimates the risk standardized rate of unplanned all-cause hospital readmissions for Medicare fee for service beneficiaries discharged from LTCHs who are readmitted to a less intense

setting, or back to another --

I'm sorry, readmitted to a short stay acute-care hospital, or back to an LTCH, within 30 days of the LTCH discharge. So again to clarify, this is post-LTCH discharge measure. It's based on inpatient Medicare claims data, and again we use a rolling two years of LTCH discharges, the measure exclusions here were also minimal and most often related to data limitations.

The risk adjustment variables include demographic and eligibility characteristics, principle diagnoses, types of surgery or procedure from the prior short-term stay, comorbidities, prior acute length of stay and ICU utilization, again from the immediately prior acute term stay, as well as number of prior acute admissions in the year preceding the LTCH admission.

Specific to this measure is also a claims-based indicator for long-term ventilator use in the LTCH. Again, we

performed a variety of statistical tests to evaluate this measure, and found that the measure demonstrated good reliability, and our risk adjustment model had good model fit and reasonable predictive ability overall.

So for example, we conducted testretest analyses using a split sample. We also evaluated the models by computing several summary statistics, including tests for the calibration, discrimination including predictive ability and c-statistics, distribution of residuals and the model's chi square.

We conducted bootstrapping in order to obtain multiple estimates for LTCH's risk-standardized readmission rate, or RSRR, and estimate confidence intervals around facility's RSRRs.

Results of the bootstrapping analyses suggest the ability to discriminate between providers with higher and lower than average readmission rates. So in closing,

this is an important measure for this patient population. There are no current-- there no competing measures, and this measure was harmonized to the greatest extent with similar measures.

The measure will provide information to providers and patients that is not easily available to them currently.

CO-CHAIR HALL: Thank you very much. I just want to note that whereas we had 90 plus minutes on the last measure, we have 35 this time around, and I'll invite Tony and Carol to kick of the discussion as discussants.

DR. GRIGONIS: Thank you. Obviously the first area we had discussed extensively for the inpatient rehab measure and that is the evidence or lack thereof, showing any kind of process improvement change relating to the change in readmission rates, and similarly there were really no studies done in LTCHs on this issue. But I'm just

opening it up. If anyone has any further comment, we could probably vote on that initial. Yes, Steve.

DR. FISHBANE: Yes. Just one suggestion for the National Quality Forum here. What we're talking about with the definition that we're using for voting is not clinical evidence. It's not scientific evidence, and I think that if we use that term as we put forth the results of our voting to the public, we may be accidentally misleading to suggest that we voted based on evidence.

In fact, what we're asked to vote on here is that there's a reasonable construct that exists, which is fair. I mean that's acceptable. But I think, you know, in the interest of transparency and clarity of definition, that the word evidence shouldn't be used, in terms of the specific criterion for outcome measures.

CO-CHAIR HALL: Thank you, Steve. So again, are we going to -- do we want to

move just with evidence? We'll take a vote. But any other comments, more or less limited to the topic of evidence?

CO-CHAIR KAPLAN: Carol.

CO-CHAIR HALL: I know I'll ask Tony and Carol just to raise their hands whenever they see fit, because they're the main discussants. But others. Wes?

MS. RAPHAEL: Are the other -- I guess Tony you can comment on this as well, that I think was discussed was whether 30 days was the right time frame, because these are very sick people. They don't have short-term episodes. But I have to say to be consistent in really looking for uniform measures, I think we came back to the 30 days.

CO-CHAIR HALL: Wes.

DR. FIELDS: I'm sorry. I'm used to yelling at people, I'm sorry. What is the evidence that a readmission to an acute-care hospital is the same outcome in terms of measurement to readmission to an LTCH? To me,

those seem like very different interventions for different problem sets. So why are we treating them as if they're comparable in value as outcomes?

CO-CHAIR HALL: Wes, can you clarify what you mean? I mean, this is a measure about LTCHs being readmitted to short stay acute care. We considered a separate measure --

DR. FIELDS: I'm just reading this, so forgive me if I'm wrong. But what it says is discharge from a long-term care hospital. Let's see. Readmitted to a short stay acute care hospital or an LTCH within 30 days.

CO-CHAIR HALL: Right, I'm sorry. I misunderstood your comment. I thought you were comparing this to the last. So your comment is well-placed now. Others, or Tony, do you want to add to that? Were you trying to raise your hand Tony or no?

DR. GRIGONIS: Pam.

CO-CHAIR HALL: Pam, I'm sorry. I missed somebody. Okay. All right, great. I apologize. Anyone who can comment on Wes' concern? Developers.

DR. INGBER: Yeah. I just want to say that to a degree that it's parallel to an acute hospital, I think a readmission to an acute hospital where we have an LTCH, which we take somewhat seriously the LTCH terminology, although we don't always use it.

They are treating very ill patients, and we're treating the readmission back to that facility as being also an unfortunate circumstance which comes after discharge, equally bad to having to go back. I mean most patients, if they do go back, will go to the acute from the community. That's where they're going to get sent first.

But we have some who go back to the LTCH, and we wanted to make sure that that was taken care of in a measure.

DR. GRIGONIS: I think you bring

up a good -- I think you bring up a good point. I don't think there's evidence in the data to suggest what kinds of patients would be different. It was no stratification done. That speaks, I think, more to the scientific part. So we may want to bring that up during that segment.

CO-CHAIR HALL: Any other comments or concerns before we vote on evidence?

(No response.)

CO-CHAIR HALL: I'm not seeing any.

MS. SHAHAB: So we're going to vote on 1A, Evidence. Rational supports the relationship of health outcome to at least one health care structure, process, intervention or service. One is yes, two is no. Voting begins now.

MS. SHAHAB: We have all 24 votes. 20 said yes, and 4 said no.

DR. GRIGONIS: Okay. The next area is opportunity for improvement. The

assessment of the performance was made using the retrospective data. No analysis was conducted for any current use of the measure. Does anyone have any issues related to the, sort of the ability of the measure, performance gaps?

CO-CHAIR KAPLAN: Actually, I think that in this case, the range is much more impressive in terms of, you know, how much the risk adjustment spreads people out in the distribution, and it makes it more, considerably, in my mind, a more compelling argument compared to the previous measure we just discussed.

DR. GRIGONIS: Right, and also the rates are higher, a lot higher than they were in the inpatient rehab facility.

CO-CHAIR HALL: Any other questions before vote?

(No response.)

MS. SHAHAB: We're going to vote on 1B, Performance Gap. Data demonstrated

considerable variation, or over all less than optimal performance across providers and/or population groups. 1 is high, 2 is moderate, 3 is low, 4 is insufficient, and voting begins now.

MS. SHAHAB: We have all 24 votes. 14 is high, 10 is moderate, 0 is low, 0 insufficient.

DR. GRIGONIS: The next area is Priority, that this measure is an important issue related to health care quality and-or the cost of care. Any comments? I think it's similar to the previous inpatient rehab measure, as far as importance.

(No response.)

CO-CHAIR HALL: I don't see any questions raised. Questions? None.

MS. SHAHAB: 1C, High Priority. Addresses a specific national health goal priority, or data demonstrated a high impact aspect of health care. 1 is high, 2 is moderate, 3 is low, 4 is insufficient. Time

starts now.

MS. SHAHAB: We have all 24 votes. 12 high, 12 moderate, 0 low, 0 insufficient.

CO-CHAIR KAPLAN: We're getting better at the voting process, you notice how? Yeah. We can make the thing connect with them. We were pretty normally distributed there for a while. Now we're kind of skewed.

DR. GRIGONIS: Okay. Now we move on to the Scientific Acceptability specifications.

CO-CHAIR HALL: I'll ask Tony and Carol to make their initial comments about Scientific Acceptability.

DR. GRIGONIS: One comment I had about the planned versus unplanned, are you still taking suggestions as far as what would constitute a planned category, because some of my colleagues across many hospitals have made suggestions.

So there's not a big list, but maybe four or five more that could be looked

at, could be examined as a potential. We see a lot of our patients being sent back for certain procedures that we know that they would have to go back for. So it's not really an unplanned situation.

DR. INGBER: From the developer's point of view, we are certainly open to improvement all the time. So we expect actually, just talking generally about these measures, they have to evolve over time, and that's one of the pieces of the evolution, yeah.

CO-CHAIR HALL: Tom.

DR. SMITH: Just to make sure I've got it clear in my head: so these are people in a long-term care facility, discharged presumably to the community but not transferred to a hospital, but in the community, and then are either admitted to an acute care hospital, or back into a long-term care facility within 30 days. And if they go to an acute care hospital and they're not

transferred back to a nursing home, they don't get counted twice?

DR. INGBER: That's correct. We look for the first event to occur, and then we stop looking past that. And any other kind of post-acute care facility we're not looking at here. It's just, are you having another admission to either the LTCH or the short-term acute.

DR. SMITH: Within 30 days of leaving the long-term care?

DR. INGBER: Yes.

DR. SMITH: Okay, and I still -- I struggle with the, I think it was Wes, the conceptual issue of equating an admission back into a long-term care facility with an admission to an acute care facility.

CO-CHAIR HALL: And keep in mind that as specified, transfers directly into care are excluded. You have to be sent home, discharged first, and then readmitted. So Tony was going to, I think, respond. No?

DR. GRIGONIS: I was just going to ask if you knew what the relative proportion was of patients who readmitted back to the LTCH, versus short-term acute case hospital?

DR. INGBER: I must confess ignorance as to the actual number. However, as I speculated before, if you're having a problem and people send you to the acute hospital off the street or wherever you are, they're not normally going to send you to the LTCH directly. But no, I can't give you the number in actuality.

CO-CHAIR HALL: Kathy and Ron, was that your question? Yes, okay.

MS. HALL: I just have a follow-up to that.

CO-CHAIR HALL: Leslie.

MS. HALL: So if the patient is sent to the emergency room for observation and then discharged to the long-term post acute care hospital again, is that in this group?

DR. INGBER: Yeah. The

observation stay, if it's a pure observation, would not be detected. But the readmission to the LTCH would be.

DR. GRIGONIS: I just had one question about that lookback period. I think you did address it, but if you could just comment on the fact that prior to the long-term acute-care hospital stay, you're considering the comorbidities associated with the short-term acute care stay up to 30 days prior to that LTCH admission. Now the fact is, that only represents about five percent of the patients. I was just curious why that was left into your model?

DR. INGBER: It's certainly possible to lop it off and make it one day. I mean it's not a technical issue here. It's certainly doable. We were trying to be more inclusive of the population of LTCH patients, and the question was: do we have reasonably good information about the patient from a stay which would normally have a lot of --

nowadays, you can have 25 diagnoses and all of that. So that lookback of 30 days was to be more inclusive, rather than less inclusive. It isn't a huge number, and if the world were to fall on us, we could chop off the five percent.

But we don't have a lot of LTCH patients to begin with, so we try to retain them.

CO-CHAIR KAPLAN: Can I ask a question about missing data? What did you do with missing data?

DR. INGBER: We were having trouble figuring out what data would be missing, in the sense that either we detect these stays in the administrative data or we don't. The data problems we had were: some of the stays we had turned out to be for managed care people and we excluded them, or some of the people did not have Part A coverage for the period we needed to collect it.

So that's the nature of the

missing data we have. It's not a systematic kind of missing data.

CO-CHAIR KAPLAN: So in your adjusters model, some of those data obviously were not, you know, you couldn't find all those. You'd just assume that they --

DR. INGBER: We would eliminate those patients who didn't have good data, and good data meant, you know, useful data.

CO-CHAIR KAPLAN: Right, and what -- do you have a sense of how many folks that was?

DR. INGBER: Oh yeah. We actually have a chart. We had a sort of flow chart in here. Let's see. I mean it's a small number. Let me put it that way. It's not a major issue.

CO-CHAIR KAPLAN: It's under five percent, yes.

DR. INGBER: The thing is that for some kinds of measures -- oh, there it is. Page 12 of something. The nature of the LTCH

patients, you don't get a lot of new Medicare patients who are missing data because they just came in. So it tends to be a relatively small number. There are also -- we find people who sometimes who change sex in the Medicare data, and there are some things we just have to lose, because it's unreliable and we exclude them.

CO-CHAIR HALL: Larry.

DR. GLANCE: One of the other cross-cutting things that I think we could maybe spend about two minutes with, and that I find that is particularly different about this, our meeting today as it was two years ago, is in discussing the scientific validity of these measures and looking at the statistical performance, in the past we've focused primarily on looking at discrimination, how well these measures discriminate between high and low quality performers, calibration on how well the model fits the actual data.

But one thing that we didn't really look at in the past and that we're looking at now, and I think it's actually very important, is the ability of a quality signal to predict future performance. This is something that all of our measure developers are being asked to look at.

So essentially, what they're doing is they're saying okay, we're going to look at the performance of our hospitals in one time period, and see how well that predicts future performance.

We're measuring that using the inter-class correlation coefficient, and what we're finding is that, at least for this measure, there's actually a very, very high correlation between past performance and future performance. I think the correlation coefficient was .08.

And that's something to bring out, and that's something that we should look at in all the measures that we're evaluating, in

terms of examining the scientific validity of these measures. I think that's very important.

DR. INGBER: I think you wouldn't want the correlation to be too high, because that means that nobody ever changed. But yes, it's -- when we look from one set of year's data to the next, it showed a reasonably high correlation of the facility measure.

CO-CHAIR HALL: Karen, and then Tom.

MS. PACE: Yeah. I just want to clarify that NQF does not really ask the correlation from one time period to another as a reliability test, often because the fact that we are looking at these in the context of performance improvement, as what you just heard.

I mean it's -- I'm not sure how to interpret that when you have high correlation from one time period to the next. Does it mean that the measure's reliable, or does it

mean that no one's changing? So, typically, what our testing task force had recommended is signal-to-noise reliability, or the split half reliability.

So I just want to put that in context, that it's not something specific that NQF asks for, because of the kind of contextual ideas about improvement over time.

CO-CHAIR HALL: Larry, do you want to respond before we let Tom respond?

DR. GLANCE: Thank you very much for that clarification. I guess to me when a quality signal predicts future performance in the next, the following year, that just has a tremendous amount of face validity, as opposed to if the quality measurement was just picking up a lot of noise. Then you would expect very little correlation year to year.

CO-CHAIR HALL: Tom.

DR. SMITH: Just, I'm not as familiar with Medicare claims data, so just going backwards, and this may be an easy

question that's addressed. I don't remember. A lot of people leave long-term care facilities when they die, and are you confident that those people are accurately identified and managed in this measure?

DR. INGBER: Measuring death, we actually are using the discharge status from the LTCH itself. There's a potential discrepancy when you use either Social Security or Railroad Board death dates. So we think the hospital knows when somebody has expired, and we trust them for that.

I can -- at the risk of expatiating here, the Railroad Board death dates are almost, that I've seen in the data, are almost the last day of a month, which is quite a coincidence. So you know, depending on the source. We're using the hospital, in this case the LTCH, as the source of that.

CO-CHAIR HALL: Thanks. Can you clarify? The 30-day, the short stay prior lookback is 30 days prior to the LTCH

admission?

DR. INGBER: Yes.

DR. HEIDENREICH: Is there any evidence on how many discharges are influenced strongly by patient and family preference, as opposed to clinical decisions?

DR. INGBER: Using the admission -- I mean using the claims data only, we really don't have any idea.

DR. HEIDENREICH: I think that was for -- I don't know if there's any clinical experts in the room who deal in these centers.

DR. GRIGONIS: I should just speak to that a little bit. I think most of the decisions are made by physicians. The family members may choose, for example, if they're sending them to a facility, that they might choose different facilities. But it's not usually the case that the patient's family would make a decision, unless it's to hospice, whether not to go back, for example, to an acute care hospital.

CO-CHAIR HALL: Leslie.

MS. HALL: Maybe it's the unintended consequences section, but it's a follow-up to this theme, that we have such disparity.

Oregon has more people dying at home, twice as many as they do in Massachusetts, and we do not want the unintended consequences of this measure reducing the ability for the family to make decisions with more autonomy, because there's pressure to perform in any particular way.

CO-CHAIR HALL: Leslie, are you saying you see a potential here, the way this is specified?

MS. HALL: I think the question comes up is if we have no way to determine whether this was at the will and desire. I mean, the family's decision-making vacillates often, and so coming home and then not having follow-up care, because we're now -- we're in a comfort care only, and then changes of mind.

Now readmitted back through the emergency room, then back through the whole cycle again.

So we want to encourage shared decision-making and encourage patient discussion, and not end up creating unintended consequences of this cycle. So I don't know whether this is the appropriate time to talk about it, or whether this is the appropriate measure, or how that reflects it.

I just think this is a social issue that's emerging more and more, as we get more aware of polls and more aware of advance directives, and we would not want to reward the wrong thing.

CO-CHAIR HALL: Kathy.

DR. AUGER: The follow-up on that is do you have information of whether any of these patients are discharged to a hospice service? Is that in there compared to home, and if so, then maybe that's one way to say like, potentially, we shouldn't be counting readmissions from hospice, because maybe it

was the family just changing their mind. You don't know.

DR. INGBER: We don't take count of where the patient actually goes. The notion is you'll send them to the best place they should go, if you're doing this right. The actual bias, if you want to call it that, if we're going to discuss this, is in the favor of Oregon, because if you go into hospice or palliative care and what-not, you're less likely to get admitted, and therefore your rates will go down.

So even though you're being predicted at a national level to be having readmissions, if you have more patients dying at home, your scores will improve, compared to everybody else. So you're safe.

CO-CHAIR KAPLAN: So when we talk about reliability, I want us to kind of separate, and it comes back to Larry's point, inter-class correlations tell you the ratio of between facility variation over between, plus

within facility variability. So in terms of single item measures, it's often used as a reliability estimate. So is there more, you know, between facility differences in these measures than there is a thumbprint across patients within a facility?

But in terms of validity testing, what Larry was talking about is, is there discriminate validity? Are we looking at the ability of these institutions? Can we distinguish one institution from another, and it gets a little confusing when you're talking about inter-class correlation.

I think where Larry was getting at is, when you shift over into a validity, kind of are we right, do we have these measures right, then the question is: can you pick another measure that should give you more or less the same information, that either you can predict a state in the future, or for example, with mortality rates or other kinds of things, can you find another measure that institutions

that deliver good care should also be showing positive results on?

That gets a little more dodgy when we're talking about these kind of facility issues, until we've had a chance to use the measure. So I just wanted to sort of clarify when you're voting on the reliability versus validity what you're talking about, what we as a group are talking about.

CO-CHAIR HALL: So within reliability, just to beat a dead horse, and not to offend any of you who know this inside and out, within the topic of reliability, with the guidance that NQF has provided in prior white papers, there are really two concepts.

One is that signal to noise concept, right, how much error is there in measuring this provider, versus the variation we see across providers. That's the signal to noise concept. That's an acceptable thing to report upon by NQF guidance.

The other acceptable thing to

report upon is really reproducibility, consistency of getting a particular result, which that again, as Sherry has indicated, that result may be wrong. But is it consistently obtained, and that's really consistency of testing, split sample testing, reproducibility of a finding over time or using different testing modes.

So either of those two larger topics, either the signal to noise area or the reproducibility area, NQF has said are acceptable under this topic of reliability.

DR. GRIGONIS: And also --

CO-CHAIR HALL: Yes.

DR. GRIGONIS: I was just going to add, as far as the validity, a more sensitive test of validity would be the c-statistic. Could you just comment on the fact that it was fairly low? I know health care models typically don't have really high c-statistics, but this was, I believe, .63. So there's a probability of predicting an outcome above

chance?

DR. INGBER: Right. The c-statistic is fairly low on this, and it was disappointing, because you always like to see a higher number. The IRF number was much higher, relatively speaking in this world.

This number is indicative to me of the fact that we have a population of multiply comorbid, sick people, and with the randomness factor added in all health care predictions, this is a tough group to tease apart.

The model itself actually has a good range of predicting people, from over 40 percent on the average on the high end, and I think it was 13 percent or something like that at the low end. It was actually able to have a range, when you look at the people.

So it is teasing them apart, and the predictive ratios indicating over and under-prediction were quite good. But the C statistic is the one statistic that is not so beautiful, indicating that specificity and

sensitivity don't give you the kind of precision you'd really like.

I find that the fact that it still seems to be able to tease apart the facilities pretty well, and that it has a good range of prediction for individuals, that it's not all squinched in and everybody's got a 24 percent probability, means there's something good happening in here, and the c-statistic, however, is just not backing that up, you know. There's no saying that it's bigger that it is. It's better than chance, but not huge.

CO-CHAIR HALL: Thank you. So we're at our time limit. I'll ask Karen to make whatever comment she's burning to make, and then we'll move to vote.

MS. PACE: I was just gong to say on the c-statistic, the other thing to keep in mind is for risk modeling, you don't expect it to be one, because you're purposely leaving out some of the explanatory variables, meaning the actual care provided.

So if it were approaching one, it would really be saying that the outcome is only predicted by patient characteristics. So just to put that in context, even though that range can go up to one, when we're only including the patient factors available at the start of care, we don't expect it to be that high. But obviously somewhere higher is nice.

CO-CHAIR HALL: So Cristie, quick comment.

MS. TRAVIS: It would help me to have an understanding of what's considered good with some of these statistical, you know, testing, both on the reliability side and on the validity side. So if there's somebody who could quickly tell us that, that would be very helpful to me.

CO-CHAIR HALL: Well, I wish there were an answer to that, Cristie.

MS. TRAVIS: Well then I don't know what it's low.

CO-CHAIR HALL: I'll tell you.

With respect to C statistic, you're asking whether you can tell an event from a non-event, and anything better than .05 means you've improved over a random guess.

MS. TRAVIS: Right.

CO-CHAIR HALL: So the economist would say if you can explain two percent of a phenomenon, you know, you're going to win a Nobel Prize. In our lives, we say we want, you know, 30 percent of the residual explained, so that your c-statistic is, you know, remarkably high.

But the real problem is that when you have a homogeneous patient population, your c-statistic's going to look horrible, and as Karen said, when you're expecting there to be an event from the therapy applied, that's going to make your c-statistic look kind of odd as well. So there's not one answer to c-statistic. Usually, we like to be dealing in the .07 range. But in this case, we're in the .06 range and so be it.

CO-CHAIR KAPLAN: The same thing is true by the way, not to interrupt you, with inter-class correlation coefficients, because it totally depends on the sample you're looking at. For some of these things, where there's not very much variation to work with, they tend to be in the range of .025, .22, which is not great.

I mean some of us would not consider that very reasonable at all. On the other hand, that's what you get in these kinds of -- many of these kinds of comparisons, especially when the variation is tight.

CO-CHAIR HALL: And for the ICC type of reliability measurements, that is currently controversial, but many people would say .04 is a minimum. We call it moderate, but it's really considered minimum. Many people are saying it should be .07 if you're paying people or not paying people, but there are very, very few models that reach that level of reliability in practice. So

unfortunately, there's not one answer.

I was about to say that Sherry and I are in this unfortunate position of needing to move the voting along. At the same time, as a scientist myself and as a provider myself, I don't like the notion that we would vote before people have a comfort level to vote.

So we will move a vote along, and you all have to raise your hand and say I'm not comfortable with that, if that's the case, and if we don't get our work done in the next two days, we'll make up for it. So I'll call for a vote, unless people raise their hand and say we are not comfortable. Larry.

DR. GLANCE: One ten second comment. When you're considering the statistical performance of all these models, you have to consider it within the framework of all the other models. For readmission measures, c-statistics of .06 to .065 extremely common. There is nothing unusual

about this one.

CO-CHAIR KAPLAN: Exactly, thank you. And the same thing is true, by the way, with inter-class correlation coefficients.

MS. SHAHAB: So we're going to vote on 2a, Reliability, which includes precise specifications and testing, appropriate method and scope with adequate results. 1 is high, 2 is moderate, 3 is low, 4 is insufficient, and the time starts now.

MS. SHAHAB: We have all 24 votes. 4 is high, 19 moderate, 1 low and 0 insufficient. So we can go ahead and vote on 2B, validity, which includes specifications consistent with evidence, testing, exclusions, risk adjustment, stratification, meaningful differences and comparability, multiple specifications, missing data.

1 is high, 2 is moderate, 3 is low, 4 is insufficient. The time starts now.

MS. SHAHAB: We have all 24 votes. 0 high, 17 moderate, 7 low and 0 insufficient.

CO-CHAIR HALL: Any specific comments on the area of Feasibility. Karen, is your card up? I got you. I don't see any comments. Wait, Wes.

DR. FIELDS: Just real quick. I think in terms of the CMS, the Medicare population, it would be highly desirable to know what the outcome following discharge from LT facilities is. I think it's very much about palliative care, hospice care and really above average community care, or family care for that matter. So although it wouldn't be -- it may or may not reflect directly on the facility, but it would say a lot about the community's ability to really do what's often necessary for these very complex, chronically ill patients. So speaking in favor of looking at disposition, at least at some point in the future on this measure, if approved.

CO-CHAIR HALL: Thank you. Any other comments? Not seeing any.

MS. SHAHAB: We can vote on

Feasibility, which is 3A, data generated during care, 3B, electronic sources and 3C, data collection could be implemented, e-measure feasibility, assessment of data elements and logic. 1 is high, 2 is moderate, 3 is low, 4 is insufficient, and the time starts now.

MS. SHAHAB: We have all 24 votes. 13 high, 10 moderate, 1 low and 0 insufficient. We can go ahead and vote on Usability and Use now.

CO-CHAIR HALL: One second, Zehra. Any comments on Usability? Leslie.

MS. HALL: Back to the flow of care, going into the observation area and then back out to the facility. My only concern would be is that there are so many limited numbers of beds, and is this another opportunity to redirect a patient, because of the potential negative consequences of readmitting to my facility. So my concern is, as a consumer advocate, are we just making it

harder to get people back to the right care as a result, as an unintended consequence?

CO-CHAIR HALL: Other comments or concerns? Yes, Tony.

DR. GRIGONIS: I just want to reiterate the same comment I made before about adapting this for an actual improvement measure, instead of a static two-year time period. Also, I just want to add in terms of Usability and Use, and that is it will be important, I think, when this becomes a public measure that's reported, that some kind of education is added to this measure. Because you'll have facilities, for example, that had very low readmission, who would now look like they're higher because of the movement toward the mean. So I would strongly encourage that that would become part of this measure, so that the public can interpret it correctly.

CO-CHAIR HALL: Helen.

MS. SHIPPY: Just really quickly. So the LTCH is actually the responsible party

here, right? So in terms of Leslie's comment, you know, someone who goes back for an ob stay, who has been discharged from LTCH, the LTCH has no ability to tell the -- or when they go to the ER and the acute care, they have no ability to tell the acute care whether or not to admit this person or not.

It's very rare, actually, when people go out to the community. I think someone made this point before, that patients would be directly admitted to an LTCH from the community. It's actually incredibly -- it's very rare.

MS. HALL: So the observations still go back to the LTCH, versus an inpatient? I'm talking about going back to the long-term plus acute readmission, versus their going into the hospital, direct from observation.

Find a bed, send the patient. Are we negatively reincenting that readmission when we have such a sparse amount of beds

available, especially in a highly complex patient? At my facility, we don't have ventilators, except in one facility. So this is a big problem for us.

CO-CHAIR HALL: So there are two aspects, Leslie. One is you might feel so strongly about this that you would be advising the developer to respecify, and the other is that you would just be voicing to all of us that you think there's a potential unintended consequence that you want all of us factoring into our decision.

MS. HALL: Uh-huh.

CO-CHAIR HALL: Russ.

DR. EDMUNDSON: Just a question. Is it possible that the same patient would be discharged from an acute care hospital into a long-term acute care, get discharged, get readmitted back to an acute care and that same patient gets counted as a readmission both in the hospital and the LTCH counts?

CO-CHAIR HALL: Yes. You mean in

separate measures, could a hospital be double jeopardy, separate measures? Absolutely, yeah. Not within ths measure, no. I understand, yeah. But across measures, double jeopardy can happen. If the developers want to express anything contrary to that?

DR. INGBER: Not exactly contrary, but yeah, it's certainly possible. But given the lengths of stay in LTCHs before they tend to discharge, the fact of it being within 30 days of that initial hospital discharge is a little bit unusual, but possible.

CO-CHAIR HALL: Paula.

MS. MINTON-FOLTZ: My question is does the three-day rule apply to LTCH readmissions or admission to LTCH? If anybody knows, because I guess you could come back as an observation patient and be missed in this readmission, but really don't meet acute care criteria. But if the three-day rule is in place, then you artificially have to be readmitted in order to gain access. Does that

make sense? Observation, no. That's whole another task force.

CO-CHAIR HALL: It's a good question.

DR. INGBER: Yeah. I'm not positive myself about whether it applies. I thought it applied to all of the inpatient stays. But that may not be the case. I'm sorry, yeah, you're right. I don't know for sure.

CO-CHAIR HALL: Wes?

DR. FIELDS: I just want to support Paula's question, because it sort of gets back to the one I asked about the science of this.

The paradox is a lot of this isn't science; it's CMS policy, rules and regs, admission criteria. They're quite different for acute care facilities and for long-term care facilities, and I believe that that's probably the primary driver of where these patients get readmitted.

The other is whether or not they have any remaining eligibility in -- I guess it's Part A, for long-term care. I think it's very likely that there's actually a substantial number of patients that do get readmitted, because they clearly are failing in the community after a recent long-term care stay, and readmitted to the hospital sort of by default, so that you can restart the clock on the three-day rule.

I could be wrong about that, but if so, somebody would need -- you know, from CMS would need to tell me which part of it I'm getting wrong. But I think that's really the practice in community settings wherever CMS patients are served.

CO-CHAIR HALL: So again, Wes, what I'm hearing is either an advisement to the developer, that you feel this measure is spec'd inappropriately because of that concern, or at a minimum, you're raising your own concern for unintended biased behaviors,

and thus unintended consequences.

DR. FIELDS: Yeah. I think the way I would phrase that is that back to the science of this. I think that there's plenty of reasons to distinguish between these outcomes, because I don't think they're the same. I think a patient being readmitted to the hospital is a fundamentally different patient.

Even though they may share a lot of comorbidity. But the reasons they get readmitted to the hospital for acute care are categorically different under CMS criteria, as well as clinical, you know, cognitive stuff, versus the beneficiaries being readmitted to an LTCH. So I just -- to me, this is another example of us lumping what I think we need to be splitting. I'd be much more comfortable with this as two separate measures, looking at each as separate outcomes.

CO-CHAIR HALL: Or if we had a number about the frequency of this readmission

to the LTCH, if it's really a contaminant. I mean if it's a tiny number, it's really not much of a contaminant, right, and not only -- again, that readmission to the LTCH also, no, it does not have to have a 30-day lookback short stay, no. So if we knew something about how big of a contaminant this is, we'd be able to make a better judgment. Do we have the ability to ask the developers to provide that information tomorrow?

MR. AMIN: That's a question for the developer, as to whether that could really realistically be produced.

DR. INGBER: Or reproduced within how many days? No, we can come up with a number. We know where people are readmitted in our data. We just haven't bothered to isolate that. So in a few days, we could pull that out.

MR. AMIN: But not by tomorrow.

CO-CHAIR HALL: So do we have a comfort level to move to voting on Usability?

Oh, I'm sorry, Pam.

DR. ROBERTS: I just have a question for the developers. How are interrupted stays dealt with within the data? Are they -- if interrupted stay is within the LTCH?

DR. INGBER: The interruptions don't really count in here because they have to be discharged discharged. You know, when you have the interruption and they're effectively readmitted and the discharge is effectively cancelled because it's one stay they're going to get paid for.

So we're not looking until after that discharge is a discharge, as opposed to a temporary discharge.

DR. ROBERTS: Okay. So they're excluded?

DR. INGBER: I'm sorry?

DR. ROBERTS: So they're excluded from your analysis?

DR. INGBER: Yes, right.

CO-CHAIR HALL: They're ignored as much as excluded, yes. Frank.

DR. BRIGGS: I was just going to comment. It's my understanding that the three-day rule is applicable to skilled nursing facilities, not for inpatient rehabs or LTACHs.

CO-CHAIR HALL: Maybe I misunderstood the question. Paula, wasn't your question when a patient goes back to a short stay hospital, they have to be there three days now to be considered a readmission, and so therefore it's not considered an event in this model?

MS. MINTON-FOLTZ: That was my question. I know for a fact that applies to SNFs. I don't know whether it's LTACH or inpatient rehab.

DR. INGBER: What, the three-day rule?

MS. MINTON-FOLTZ: It's a three day --

DR. INGBER: The three-day rule that I've seen --

MS. MINTON-FOLTZ: Yes. It's a mandatory three day rule before you can gain admission. You can't just go right into a nursing home from -- skilled nursing right from home or from the clinic. So but I don't know whether -- so it will apply in the next measure for sure, but I don't know if it applies on this one.

DR. INGBER: I don't think it does. I don't think it applies to anything but nursing homes. I'm sorry, I was a little confused the last time around. But the three-day rule is only a SNF rule, and that's it.

MS. MINTON-FOLTZ: I think that when I Googled it, it looks like that's considered an interrupted stay and not --

(Off microphone comment.)

DR. INGBER: I mean, you can --

MS. MINTON-FOLTZ: Yes, which is different.

DR. INGBER: You can walk in off the street to any of the post-acute cares, and there are LTACH admissions off the street and there are ERF admissions off the street. It's just that they're not a high proportion of them. So the three days does not get mixed up in this measure.

CO-CHAIR HALL: So I think we have to move forward. Obviously, some of you may have some level of uncertainty or discomfort around various aspects, and I think you have to express that when you vote. So unless anyone wants to throw up a stop sign, I think we'll move to a vote.

MS. SHAHAB: We're going to vote on Usability and Use. 4A, Accountability and Transparency. It's used in accountability within three years, public reporting within six years, or if it's new, credible plan. 4B, Improvement. Progress demonstrated and if new, credible rationale, and 4C, benefits outweigh evidence of unintended negative

consequences to patients and populations.

1 is high, 2 is moderate, 3 is low, 4 is insufficient information. Time begins now.

We have all 24 votes. 0 high, 9 moderate, 10 low, 5 insufficient information. So we can go ahead and vote for overall suitability for endorsement. Does this measure meet NQF criteria for endorsement? 1 is yes, 2 is no. Time begins now.

CO-CHAIR KAPLAN: Can everyone do it one more time, please?

MS. SHAHAB: We have all 24 votes. 10 yes, 14 no. For Measure No. 2512, All-Cause Unplanned Readmission for 30 days post-discharge from long-term care hospitals.

MR. AMIN: So this is an example of a measure that's sort of fallen into our gray zone, where consensus hasn't been reached. So this measure will continue to move forward in terms of the comment period, and we'll revisit this during the comment

call, and yes.

DR. INGBER: Just to ask a question. If that number of readmissions to LTACHs was a key to any of this, we can call the programmer and have a number for you either this afternoon or tomorrow.

CO-CHAIR HALL: Well, I think what you've heard now is that that expresses our group's opinion for now. The measure is not prevented from moving into the next phases, so if I were you, I would be prepared to comment on those figures in the next phases. Thus, I don't think it's necessary to comment on it tomorrow.

MS. KHAN: Yes. We'll follow up with you with more steps after the meeting.

CO-CHAIR HALL: And I believe we're about a half hour behind schedule now, so we succeeded in reaching that goal pretty quickly.

(Laughter.)

CO-CHAIR HALL: We will move on to

2375, and our invited discussant to kick off a brief description of their measure will be the American HealthCare Association, and then I'll ask Helen and Frank to kick off on the group side. We thank RTI so far. We think we'll have them back shortly.

So when ready, we'll ask American HealthCare Association to introduce themselves and briefly describe their measure.

MS. KHAN: Do you have anyone on the phone who needs an open line?

DR. GIFFORD: My name is David Gifford. I'm a geriatrician and the senior vice president for Quality and Regulatory Affairs at the American HealthCare Association, NCAL. We represent about 10,000 of the 15,000 nursing homes across the country.

MS. SHAW: I'm Urvi Shaw, and I work for the American HealthCare Association as their senior manager of Quality Improvement.

CO-CHAIR HALL: So please, briefly describe your measure if you would.

DR. GIFFORD: So our measure is an all-cause rehospitalization measure for individuals admitted from a hospital to a SNF, regardless of payer status, regardless of condition, regardless of issue. It looks at any readmission that occurs within 30 days during their SNF stay.

If they leave the SNF before 30 days and are then hospitalized after discharge, they're not counted. We utilize the MDS record. Therefore, we capture any stay in the hospital, whether it be an inpatient Medicare stay or an observation stay, whatever other insurers might include out there for the overall measure.

The measure is a rolling 12-month measure. It has a minimum denominator size of 30. It has no exclusions and is calculated for all the nursing homes in the country. Currently, four ACOs are using this measure on

a regular basis. An MCO is using this data and we are in the process of sending the entire state of New Jersey, all the nursing homes in New Jersey's data to a couple of ACOs in New Jersey for their use as well.

And probably since you're behind schedule and the more important thing is to answer your questions, I'll stop there, or I could keep describing in detail.

CO-CHAIR HALL: All right, thank you very much. So I think the group is getting used to the sense that we're going to first discuss the evidence, and I'll ask Helen and Frank to kick off the discussion.

DR. BRIGGS: So there's a considerable gap in the variation across the country. I think we have some statistics included by state provider, from lows of 13 up to 22 percent by state. I think there's considerable variability there, that points to the overall need. This is again, an outcome measure in terms of readmissions.

DR. CHEN: And just as a comment, you know, it sort of -- the evidence is a little bit of a gamish in terms of what the literature that's reported out there. Some of it is really conflated between long-term care, true long-term care versus people who are in post-acute SNF.

It probably doesn't make a huge amount of difference regarding this measure, but just so you know, the literature out there is very mixed on this topic, although technically I would say that there probably are process measures that do impact whether people do have a higher or lower rate of acute care utilization in nursing home facilities.

CO-CHAIR HALL: Others with comments or concerns on the topic of evidence? If not, we'll move to vote.

MS. SHAHAB: Voting is going to be open for 1A, Evidence. Rationale supports the relationship of health outcomes, at least one health care structure, process, intervention

or service. 1 is yes, 2 is no. Time begins now.

MR. AMIN: Paulette, I think we're waiting on your vote.

MS. SHAHAB: Voting is now closed. 22 yes and 1 no.

So now there's 23 yes and 1 no.

CO-CHAIR HALL: Gap, performance gap. Helen or Frank, anything else to add?

DR. CHEN: No. I think Frank spoke to that.

DR. BRIGGS: I think when you look at the actual readmission rates, the average in this range is 18 as compared to the inpatient rehabs and the long-term acute centers, both having lower. So I think there is a larger issue with the skilled nursing.

MS. SHAHAB: 1B, Performance Gap. 1 is high, 2 is moderate, 3 is low, 4 is insufficient. Time begins now.

MS. SHAHAB: We have all 24 votes, so voting is closed. 15 high, 9 moderate, 0

low, 0 insufficient.

CO-CHAIR HALL: Priority, high priority. Any comments? Seeing none.

MS. SHAHAB: Voting is open for High Priority, 1C. 1 is high, 2 is moderate, 3 is low, 4 is insufficient, and your time begins now. We're missing a few votes. Just one more.

For High Priority, 19 voted high, 4 moderate, 0 low, 0 insufficient.

CO-CHAIR HALL: Let's open the discussion on Scientific Acceptability, Reliability, Validity. We'll again invite Frank and Helen to start.

DR. CHEN: I had a question regarding why planned readmissions were not excluded from this measure because my question would be, how would this be actionable at a facility level, you know, to be held accountable for that?

DR. GIFFORD: So when we developed this measure, it's based off MDS

data. So we don't know and it wasn't a variable in the MDS to collect planned/unplanned. Subsequent to the development of the measure in the last year, CMS has added that as a variable.

It is missing data 82 percent of the time, and so we do not want to use a claims-based component of this because it (a), restricts the population down to only Medicare fee for service, and this is out of the huge, growing population of managed care.

We worked under the assumption, looking at some of the broader data, that other than a handful of facilities, this would not affect the overall rate at a facility level measure, certainly at individual levels. It is probably one of the more common questions we get from many of our members when they're exploring the data that's out there.

We are looking at whether we can utilize that MDS data, but it's been a new item. So the reliability and validity of it

is not tested, so we didn't feel comfortable using that measure overall.

DR. CHEN: Along the lines with missing data, I noticed that in your discussing regarding missing data, you provide some description of the relative low prevalence of missing data, but then you add a statement saying that it would be useful to calculate its effect on the numerator. I wondered if you could flesh that statement out more.

DR. GIFFORD: So we calculate this based on the admission MDS assessment as to whether they were admitted from a hospital or not, and then the requirement is that all discharges, whether they die or not, go home, or any acute care setting, a discharge MDS assessment be completed.

We are able to look then out to see if discharge assessments are being filled out consistently. If overall, 97 percent of discharge assessments are being filled out

that we can assess, if we have less than 95 percent completion rate, we do not report the data for an individual facility that's out there, going forward.

We have not looked specifically at what the impact of say 90 percent or 92 percent would have on the overall, but we felt the greater than five percent missing data for that key element might have -- we would drop them and not report the data.

We do suggest and recommend that a flag be provided back to the providers, that they have high missing data on the district assessment form. It tends to cluster in certain facilities, and therefore they would hopefully improve their data collection piece of that process.

DR. CHEN: There was a request for a clarification during the work group call regarding the ability to stratify based on MA, MCO organizations, and my understanding is that even though they're included in the data

set, that we don't really have a very reliable way of looking at the overall measure based on payer class.

DR. GIFFORD: There is a data element in the admission MDS assessment, where the nursing homes are to indicate what is the payer status. It is very unreliable data. Sometimes the portrayal from families or from the hospital as to what the insurer is is not really true, so you discover it after the fact. It's hard to believe.

The other is actually, payer status actually changes a lot because of eligibility and various issues. You were talking before about a three-day rule. You often don't discover the three-day observation rule until after you fill in your first MDS assessment. So we don't feel comfortable breaking it out by payer status to stratify that. Though when you look at the SNF claims overall, about two-thirds of all the admissions are, on average, are coming from

Medicare fee-for-service.

But that again varies by region of the country and it varies by facility, such at some places have very few Medicare fee-for-service altogether if you use that as a metric. So we do not stratify it by payer status.

DR. BRIGGS: So in terms of the data collection on the MDS, have you looked or do you know the outliers in terms of -- I know on the hospital sides, having looked at thousands of our discharges recently, we certainly get it right in terms of deaths.

But in terms of discharge status to a skilled facility or home, home with home health and things like that, we have certainly a lot of variation between the person who is doing that registration task.

So I was wondering if you had any information in terms of outliers. You have missing data, but then do you have a lot of variation between sites?

DR. GIFFORD: I'm following up to the very last question. Variation on the sites on what, missing data or --

DR. BRIGGS: In terms of their admission sources. Are you seeing a lot more coming from saying that they're being admitted directly from the hospitals, and then on the discharges, are you seeing patients going home, patients being transferred?

DR. BRIGGS: It really varies by region of the country, as far as whether people are coming, say Medicare fee for service and managed care or what issues, and also whether they're being directly admitted from the ER without a stay or coming from home into the facility.

Clearly, those areas with much higher penetration of managed care, those are some areas with more ACOs. We're anecdotally hearing that. When you look at the data overall, like Arizona, which is all managed care, you can really see some differences out

there in Arizona.

It also varies by the observation status issue that's out there going forward on it. We have looked at the sensitivity and specificity of the MDS against Medicare claims for those who have just Medicare claims, and it's got a pretty reasonable sense of specificity, same as most diagnostic tests that we use out there.

84 percent sensitivity, about 97 percent specificity with about -- when you look at the overall measure itself, it's about 96, 97 percent in agreement with the measure itself, because you're aggregating at the facility level. I think I -- did I answer your question? Okay.

CO-CHAIR HALL: Paula.

MS. MINTON-FOLTZ: Thank you. Just a comment regarding the planned readmission question before. It's becoming a very common tactic for the length of stay in hospitals, for orthopedic, trauma and burn

stage procedures, to use the intermittent and planned readmission. So I just -- that might be in the Feasibility, but thought I'd make that comment.

DR. GIFFORD: We would actually -- I mean, if the hospitals are trying to game the system, we wouldn't pick that up. It's really the facility; it's the SNF. They send the person out. They don't know whether, how they're going to be admitted or what they're going to be admitted for.

They will be documented as they went to an acute care hospital and will be admitted, whether it was an observation stay or not.

MS. MINTON-FOLTZ: Well, that's my point.

DR. GIFFORD: If they came back, they would get them there.

MS. MINTON-FOLTZ: I think that's my point. It's a common tactic, in partnership with SNFs, to send people for

well, waiting for the swelling to go down, why should they wait for their orthopedic procedure.

But I think SNFs would be more reluctant to partner with hospitals, and we'd see a shift in length of stay, if we were not able to like somehow weed out that planned admission.

DR. GIFFORD: No, I fully agree that we don't capture planned/unplanned. I would say that hypothesis is very unlikely because the hospitals are so much trying to get people out, and SNFs are so dependent on the volume to come in. They're not willing to criticize hospitals for anything.

I mean, the quality of the transfer information, for example, has been historically abysmal when they show up on the doorstep of a SNF from a hospital. But that has improved dramatically since the hospitals are now being held accountable for the 30-day readmission rate.

Our broad membership has seen that as a huge bonus, and a much more closer partnership. But you're also seeing an alignment, where you have to sort of take everything. So in the areas where there's ACOs that are really actively going on, they're forming tight networks and they're using this data for network selection, but they're also using it to sort of drive all the volume there.

So it would be hard to sort of play, I think, that gaming process that you're describing.

CO-CHAIR HALL: I guess even within an ACO, if you want a tighter coordination between hospital and SNFs so that people could move back and forth as appropriate, this measure would defeat that, or it would be a counter-incentive.

I think in the broader sense, if we as a profession have moved over time toward the notion that certain things that are

planned should be considered separately, which we have for a number of measures in front of us today, and which we have for measures in the past, I'm wondering why we wouldn't ask the same of this measure in front of us today. Leslie.

MS. HALL: I'm sorry. I don't know your data source. I'm not familiar with that, and so I have a question about whether patients' goals or directions are included in your data set, to accommodate for what their desires are for appropriate care, that question.

And then just also understand, in an area where there's a natural decline and a defined scope of practice, what are the alternatives? Are we rewarding or penalizing the wrong thing? So it's my ignorance, if I don't --

DR. GIFFORD: I'm going to answer your first question, and then I'm not sure I understood your second question. The minimum

data set can only be described by the government as minimum. It's been in use for almost, I don't know, 15, 20 years now in the nursing home.

It was required in OBRA-87, so I guess 87 forward. It has over almost 600 data elements in it for standardized data elements. They're on Version 3.0 right now. Every admission to a Medicare or a Medicaid-certified building, which is about 96 percent of all the 15,000 SNFs in the country, are required to collect that at admission, and then every 90 days thereafter significant change in status.

If you are Medicare fee-for-service and most Medicare Advantage plans and a lot of commercial now require you also collect it at admission, and then about every 7 to 14 days thereafter, until they're discharged from that acute care episode.

It collects a robust set of information around ADL function and cognitive

status, pressure ulcers, it's a long list. It goes on and on and on. It's been shown to be pretty reliable and valid.

All other quality measures that are currently used on five star public reported measures and in the payment systems that CMS uses rely on the MDS data collection that's out there. It's collected on everyone, regardless of their payer status. It's a requirement. And the second part of your question?

MS. HALL: And short answers. No, their patient goals and directions are not in that data set?

DR. GIFFORD: There's a couple of questions in there about whether they plan to be discharged home, and what some other goals are. But no, not at a robust level like you might see in a more standard type of rehab center.

MS. HALL: Then my second question is my clinical ignorance, but when is a

decline a readmission, and when is a decline simply a scope of practice that can't be done inside the skilled nursing facility?

DR. GIFFORD: We elected to go with all-cause, to avoid the gameability in that gray zone of when you determine it, and you know, we're using it as a feedback report. So we actually provide it to all 10,000 of our members on a quarterly basis, so they can see the report. Hospitals are using it. At least they had some sense of where they're going on that angle.

I would never want to set a goal of zero for this issue, but I think as pointed out by Frank, 18 percent, I think most of us who practice in the setting would agree it's a little bit high.

CO-CHAIR HALL: Pam and then Karen.

DR. ROBERTS: Could there be any unintended consequence of not having planned events that go back, that could hold the event

until after 30 days so that they're not at risk, especially on like, two-stage procedures?

DR. GIFFORD: I guess that would apply to any -- I mean any potential provider could try to game the system. Whether it's a claims-based measure or MDS measure, whether it's included or not and how it's defined out there. So I guess if it is, they'd have to hold them a long time to get there.

CO-CHAIR HALL: Karen.

DR. GIFFORD: I just can't see it happening a lot, to affect the overall measure. These are definitely, as we know and as we see, we've heard before, gender changes on claims all the time. But it's not at a large enough rate to affect the overall metric.

DR. JOYNT: I just wanted to comment that I think it's interesting that this is -- oh, it's on. Oh, sorry. I just wanted to comment that I think it's actually

interesting that I think the strengths of this measure are almost opposite some of the other ones that we've considered.

I don't think we're going to come up with a perfect measure, but I just think there's a few things worth pointing out about why this is different from some of our claims-based metrics, that I think to me makes it an interesting complement, which is that it's a lot more timely than claims-based measures.

It's all payer, which I think is really important for understanding disparities and for other patterns of care that we may miss, and we're thinking about only folks who have one particular type of insurance coverage.

It doesn't use the shrinkage model from what I can tell from the measure specifications, which may be my own personal bias, which I think is a strength, and it actually -- the data on here would suggest that it runs with other metrics of quality

measured at the nursing home level, which many of the other readmission metrics actually don't. So I don't think that this measure is perfect, but I think it's worth noting the ways in which it differs from claims-based measures that might be important for this group.

CO-CHAIR HALL: Thank you, Karen. Larry.

DR. GLANCE: I think that, I'm probably echoing what a lot of people on this committee feel, that one of the major limitations of this measure is the inability to exclude planned readmissions, and that's -- I think it's a significant limitation.

Most of all of the other measures that we're looking at today, for better or for worse, are based on Medicare data, and I wonder if you would go back and reconsider the decision to link your MDS data with Medicare data and to maybe reconfigure your measure in such a way that at least, for a certain

segment of your patient population, you're able to measure and to exclude planned readmissions.

DR. GIFFORD: It is certainly technically possible to do it. It will only apply for the Medicare fee-for-service, where you have the claims, until we have an all-payer national claims database.

The trade-off for that slight increase in accuracy, because I don't believe it's that significant in accuracy overall, is that we go from having results within three to four months, within the close of a quarter to having results probably 16 months later. Our membership, the number one complaint is that the data is not timely enough.

Basically, they find claims data useless because it comes so late. They don't believe it anymore. They all have their own internal data they're trying to track anyway, and I've seen that with a lot of the hospitals too.

I mean, certainly it's based on payments, so people will pay attention to it. But it's so long in coming that we explicitly elected to lose some of the ability for improved risk adjustment and some of this addition for there. I would rather drive it through the reporting of the MDS data for that.

But if we did that, it would make the measure not useful for our membership or the SNF community. I would hazard to guess that at the -- seeing some of the data at the hospital end, that the percentage of planned versus unplanned and the variation of that across hospitals is so little that I can't imagine it as a significant variation overall for the overall measures, and that that trade-off is not one that I think we would make with the measure.

DR. GLANCE: Really quick follow-up. Do you have empiric data to show that the number of planned admissions is a very, very

small fraction of your overall readmissions?

DR. GIFFORD: Can I turn to the -- can I ask another contractor? So Karen, do you guys wait? You guys exclude planned/unplanned using the hospital stuff in your SNF RM. What is the number of exclusions of planned readmissions overall, on average?

(Off microphone comment.)

DR. GLANCE: So I'm asking RTI, who has developed the other measure after lunch, which is the next measure, which is a SNF 30-day readmission measure that's based off Medicare fee-for-service claims.

DR. SMITH: Can you hear me? Okay. This is Laura Smith. I'm from RTI International. So what I was saying, before you guys could hear me, was that so our unplanned rate for 2011 of readmissions is 21.1 percent.

If we were to have included planned readmissions, that would have added 1.3 percent, two percentage points to that

total, and that 1.3 percent is about five percent of readmissions.

DR. GLANCE: Does the relative ranking -- I mean really, the issue is relative ranking and change and --

CO-CHAIR HALL: Yes, but I'm not sure we can get into that with respect to your measure, since we -- I don't know that we know as a committee whether all the data are comparable and would apply equally to your measure. We certainly can appreciate that we think on the order of five percent of readmissions in this environment are planned, correct?

DR. GIFFORD: If you use the MDS data, as I said out there, right now the ratio is about two percent are labeled as planned versus unplanned, and that's about -- that's being filled out for about 20 percent. So if you assume that 20 percent is represented nationally, it's about two percent, which would correspond approximately to what they

just sort of mentioned.

CO-CHAIR HALL: So a reasonable guesstimate for now is two to five percent?

DR. GIFFORD: Yes. I mean we could certainly try to go back and look at, because we have the claims linked.

CO-CHAIR HALL: Understood. So for what's in front of us now, I think we just have to push forward. We all have some uncertainty again, which I think unless somebody needs to throw up a stop sign, we're pretty far behind. So please do throw up a stop sign if you need to, but otherwise, I think we need to move and vote based on whatever uncertainties we might have.

Again, it's uncomfortable for us to push to a vote because these are great discussions, very insightful comments from all directions. But we don't have the liberty of staying on one topic all afternoon. So all right, we'll push for a vote on Scientific Acceptability and Reliability on this measure.

MS. SHAHAB: We're going to vote on 2A, Reliability. 1 is high, 2 is moderate, 3 is low, 4 is insufficient, and time starts now.

We have all 24 votes. 4 high, 13 moderate, 5 low and 2 insufficient.

CO-CHAIR HALL: We'll go on with Validity.

MS. SHAHAB: 2B, Validity. 1 is high, 2 is moderate, 3 is low, 4 is insufficient. Time begins now.

All votes are in. Voting is now closed. 1 high, 17 moderate, 6 low and 0 insufficient.

CO-CHAIR HALL: Any specific comments on Feasibility?

DR. BRIGGS: Whether it's under Feasibility or not would be -- so this data source actually includes readmissions, both to the inpatient setting and observation setting. The area that it doesn't capture that's very high use by the SNF patients is ER visits,

where the patient goes to the ER, is reassessed, is sent home.

It still places lots of burden on both the patients and the health care setting. Don't know if there's really any data out there perhaps that should become a balancing measure at some point. But do want to point out that this measure does include both inpatient and observation, which had come up in previous discussions.

CO-CHAIR HALL: Thanks, Frank. Helen, anything to add? No. Anyone else. Okay. We'll move to vote.

MS. SHAHAB: 3, Feasibility. 1 is high, 2 is moderate, 3 is low, 4 is insufficient, and time begins now.

We have all 24 votes. 14 voted high, 8 moderate, 2 low, 0 insufficient for Feasibility.

CO-CHAIR HALL: Any specific comments on Usability?

DR. BRIGGS: I think this kind of

cuts across both this measure and the next measure, but the impact of the CMS 2-Midnight rule, making more people observation status, therefore not qualifying for that three-day rule potentially.

I think that has the potential to be not an unintended consequence of this measure, but unintentionally impact the outcome of this measure as we go down the road.

MS. PACE: But they're not differentiating, Frank, you know, for people who go back. Irrespective of whether they meet the 2-Midnight rule or not, they're still captured in this data set. So the op stays are actually counted, you know. The facilities are held accountable for people who go back, even if they don't meet the 2-Midnight rule.

DR. GIFFORD: Yeah. Since it's not Medicare claims, the three-day rule only affects your ability of whether you're going

to qualify to be a Medicare fee-for-service or not. If you don't meet the three-day qualification, you will not have a Medicare fee-for-service.

So even if you had, you know, one or two observation days and then you had two inpatient days and you came to a SNF, you wouldn't qualify for Medicare Part A and you'd be using some other insurance. So you couldn't use a claim for that. But ours is anyone coming in from a hospital.

CO-CHAIR KAPLAN: I have one quick question about it. Different from claims data, this measure may be more vulnerable to things like are you measuring improved documentation? Are you actually measuring improved quality of care?

So could you just give us a little sense of the potential unintended consequences of yes, you've improved documentation and then your quality went in the wrong direction?

DR. GIFFORD: Certainly with any

claims or any measure, you can change how you code anything. So anything's gameable out there. The MDS data is used for payment purposes. It's used for quality measure purposes.

CMS has done some checks in the past but not -- but they just announced last week they're going to start doing more audits of the MDS for the quality measures because they're also about to start using it more for payment and looking at the quality that's out there.

But you know, yes, it's gameable, just like any other measure is gameable.

CO-CHAIR HALL: Paula.

MS. MINTON-FOLTZ: I would just -- sorry. I just need to revisit the planned and the observation stay because I think as we move forward in where people should be in the affordability of care at certain levels, I think observation state is a better option to tune somebody up very quickly, as opposed to

an admission and that again, it's not a game. It's just making sense of where patients should be.

And again, planned readmissions, why should -- patients' recovery between surgical procedures, if they don't need trauma level care, why would we pay for that? So again, I think those are two measures that or two things I just can't get past.

CO-CHAIR HALL: Thank you, Paula. Helen.

DR. CHEN: Two comments, the first one being regarding the MDS coding intensity issues, for lack of a better term. I think -- I'm not sure that this measure actually promotes that per se.

I mean, I think organizations who are using MDS and it's tied to reimbursement have already had a huge push already, in terms of improving the documentation. So I don't think this measure per se would actually facilitate that.

I think frankly, tying it to reimbursement has pushed that for a lot of facilities. But it is certainly a concern, just like in the MA world, coding intensity is certainly a concern. So just to speak to that. And then regarding -- and then I've lost my train of thought, so I'll just stop here.

CO-CHAIR HALL: Yes.

DR. GIFFORD: Can I just respond to that? I'd say the bigger area or the problem is not the admissions, whether they came from the hospital or not.

It's whether they use a discharge assessment, saying they went to the hospital or not. That's why we ended up setting the 95 percent limit, where they're missing the data and they decide not to report that.

Because it's hard -- I mean, to actually say they didn't go to the hospital when they really did, you have an audit trail that's in trouble. More than any other health

care setting, we get audited and in trouble a lot.

CO-CHAIR HALL: The good news is the folks that are not filling out their forms are probably not going to -- are probably going to change that behavior soon.

DR. GIFFORD: Yes.

CO-CHAIR HALL: There's a number of pressures on them to change that behavior.

DR. GIFFORD: And they wouldn't have a data result, and for the observation stays, we would capture them. But for the planned, as we've said, we've beat that dead horse.

CO-CHAIR HALL: Thank you, all right. I don't see any other cards up, so let's move into voting Usability.

MS. SHAHAB: We're going to vote for Usability and Use. 1 is high, 2 is moderate, 3 is low, 4 is insufficient. Time begins now.

Just one more.

We have all 24 votes. 5 voted high, 14 moderate, 5 low and 0 insufficient information.

CO-CHAIR HALL: All right. Any final comments before overall? I do not see any cards raised. Okay.

MS. SHAHAB: Voting is open for overall suitability for endorsement. Does the measure meet NQF criteria for endorsement? 1 is yes, 2 is no, time begins now.

Just one more vote, please. Okay. We have all 24 votes. For Measure 2375, point right on point, 30 SNF rehospitalizations, 22 voted yes and 2 voted no.

CO-CHAIR HALL: Thank you for that conversation. It continues to be an incredibly educational one as we go. We'll now open public and member comment.

MS. KHAN: Operator, can you open the lines for public comment please.

DR. GIFFORD: Thank you all very much.

OPERATOR: Thank you. At this time, if you have a question or a comment, please press star and the number one on your telephone.

(No audible response.)

And there are no public comments at this time.

MS. KHAN: Anyone in the room who would like to make a public comment?

DR. KRISHNAN: Thank you. My name is Dr. Mahesh Krishnan. I am the vice president for Clinical Innovation and Public Policy for DaVita Healthcare Partners. We serve approximately one-third of the U.S. dialysis patient population. We have 170,000 ESRD patients in our 2,220 clinics.

The purpose of my visit this morning is to comment on Measure 2496, which you will comment on soon, standardized readmission ratio for dialysis clinics. We are actually opposed to this all-cause measure for both usability and scientific validity and

acceptability reasons.

We do believe it's important to align incentives to reduce admissions and thus readmissions to the ESRD population, but feel that the measure as constructed has significant issues. There are four specific issues for usability.

With regards to usability, we believe that the first reason is that many of the admissions for all-cause are not actually within control of the dialysis unit. Based on an analysis of 2011 Medicare claims data, ESRD patients have an admission rate of 1.88 admits per patient per year.

The percentage, however, of those admissions that is attributable to factors directly influenceable by the dialysis unit is significantly much lower. Five percent of admissions were due to vascular access infections, and 27 percent were attributable for all cardiovascular disease, which includes fluid overload, which is something that could

be attributable to the dialysis unit.

That 27 percent included things outside of the realm of the dialysis unit, such as coronary artery disease, acute MI and many others. The majority, then, of admissions and presumably readmissions then are due to other effects and organ manifestations of chronic disease, most of which are beyond the ability of the dialysis unit to impact.

This is actually unique, as opposed to the other measures which are being presented today, where hospitals may have 100 percent accountability of admissions and therefore have 100 percent accountability of readmissions. That's different for dialysis units.

Secondly, 17 percent of our patients in our data sets -- and we have significant disease measure experience with a skilled nursing plan for dialysis patients as well as a disease measure group. In that

analysis, 17 percent of our patients had a readmission within three days post-discharge.

That three days represents a time period before the patient has even encountered the outpatient dialysis setting. Again, dialysis is done Monday, Wednesday and Friday. So if you leave the hospital Friday afternoon and you're readmitted over the weekend, there's not a hell of a lot the dialysis unit can actually do to fix that.

Within dialysis units themselves, we do not receive timely data, nor are hospitals required to provide such data to dialysis units to coordinate care. Despite a large program, which we enacted approximately three years ago in the past, where six percent of our payments, six percent of all payments were dependent on finding case mix adjusters through discharge summaries, which means we were highly, highly motivated, at one year after the original discharge, index discharge, less than 50 percent of those discharge

summaries were available to us despite that, making care coordination difficult over that long time period, let alone a shorter time period.

The CMS recently conducted a dry run of the SRR with its dialysis facilities, and the feedback that I received when that closed on Friday was our dialysis units had no idea what to say, because they didn't know what admissions or readmissions they had because of the data gap mentioned above.

Lastly, kind of more related to the MAC. In our special needs plan, as I said, we have actually significantly reduced readmissions across the board for all cause, but that required significant resources. We had to embed case managers within hospitals. We had to expend vast amount of efforts and money on IT systems to coordinate that care.

The current mechanism for enactment, and I know this is sort of out of the scope of this but more in the MAC, is to

include, potentially include this metric for the dialysis unit, to withhold this two percent that would fund the actions to actually improve this, and that is amongst ten other, eight to ten other different initiatives.

With regards to validity and methodological issues, there are two major points. First, the statistical model used to risk-adjust this measure has not been subjected to peer review, which is problematic for us.

But secondly and perhaps most importantly, recently two weeks ago, the NQF noted that socioeconomic status may affect quality outcomes, and socioeconomic status is by definition not taken into account in this model.

Because CMS releases this data on readmission rates, we were able to aggregate all data for dialysis units across the United States, all 5,400 dialysis units, cross-map

those with zip code-based census data for extreme poverty versus not.

We found there's a correlation between socioeconomic status and readmission rates, that patients who are in units with the lower decile of readmission rates tend to always have high socioeconomic status be associated with low readmission rates, and vice-versa. So we believe that this is really important to take into account.

Secondly, there are hospital-based dialysis units, some of which are operated by acute care hospitals, very similar to the discussion you all just had on inpatient rehab facilities. Again, we would assume that readmission rates will be significantly higher than readmission rates in the non-hospital based dialysis units.

In summary, we do believe that it's important to incentivize the alignment between reducing admissions and readmissions for the ESRD population. We believe this

current measure may be more suited for hospital measures, since they are accountable for 100 percent of the discharges, which may then prompt them to give us the data so we could actually coordinate care.

We do believe that if an ESRD-specific measure should be contemplated, that cause-specific hospitalization may be thought about. So it's potentially vascular access-related infections, fluid-related infections, those sort of things, and that that model should be peer-reviewed, risk-adjusted and SES-adjusted as well. Thank you.

CO-CHAIR HALL: Thank you, Dr. Krishnan, for those well-thought, well-spoken comments. I'll remind our group that those comments will apply to the second measure you will hear after lunch. Other public commentary?

(No audible response.)

CO-CHAIR HALL: It's 12:45. Helen has ordered me to allow only 20 minutes for

lunch. We will reconvene at 1:05.

CO-CHAIR KAPLAN: Yes, wait a minute. Helen, I don't know. We need to get the business of this committee, absolutely full-throated discussion, so we need to eat fast.

CO-CHAIR HALL: 1:05 please, 1:05. And just a quick housekeeping note. Not to be draconian about this, but we would ask members of the audience to wait until the committee has eaten or gotten their food before getting food themselves. Thank you.

(Whereupon, the above-entitled matter went off the record at 12:45 p.m. and resumed at 1:40 p.m.)

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

(1:04 p.m.)

CO-CHAIR HALL: Thank you all for returning to the table. Some of us are still chewing away. That's all right.

We will move on with the agenda, Measure 2510, skilled nursing facility, 30-day all-cause readmission. Developer is RTI. We'll have RTI open the discussion with a brief introduction to this measure, please.

DR. LAURA SMITH: Thank you. My name is Laura Smith. I am from RTI International, and thank you for the opportunity to speak about our measure.

This is a claims-based measure that estimates the 30-day risk standardized rate of all-cause unplanned hospital readmissions for Medicare fee-for-service patients who have been admitted to skilled nursing facilities. The risk window for this measure is similar to the last one that we talked about. It looks at the 30 days after

discharge, but in this case it tracks the patient for that full 30 days because we are using claims to identify the readmissions. This measure is based on 12 months of SNF admissions. It's the calendar year.

I am going to skip through my discussion of the importance of the measure. I think quite a bit of that was covered in the prior discussion. So this measure was designed to harmonize with the CMS hospital-wide readmission measure and the other CMS post-acute care measures to the extent possible to promote shared accountability for improving care transitions. This measure can be used by providers for quality improvement and by patients for decisionmaking.

The statistical methods for the model development are very similar to those that you heard about for the in-patient rehab facility measure and the long-term care hospital measure. It's a random effects risk adjustment model, and it includes

comorbidities and primary diagnoses and demographic information identified on the claims from that prior acute hospital discharge claim that occurred previous to the SNF admission.

We also include prior utilization measures. I just want to note here there was a question that had come from the panel that if people received our written responses there was a question about the ICU days and whether or not we had evaluated days in ICU. And I just wanted to correct the written response. Right now, we only have a yes or no whether or not there were any ICU days in the model.

We do have a set of dummy variables for days in the prior acute hospitalizations or the length of stay, but just an indicator of the days in ICU. Despite not having that -- days in ICU, we do find that our model statistics demonstrate good model fit and discrimination.

Measure reliability and stability

testing showed covariate values remained stable over time, and our split sample test/retest reliability results yielded an overall interclass correlation coefficient of .56.

In regard to validity testing, we did find low correlations with other quality measures as expected, but the higher correlations with the vaccination and RN staffing measures support the validity of the measure. The measure shows variability across facilities nationally and the ability to differentiate facility scores from the national mean.

For a special issue regarding our measure, observation stays are an important issue to monitor, but our analyses suggests that the exclusion of observation stays from the measure numerator have -- will have little impact on the measure right now. We have new results looking at 2011 data where we found that had we included observation stays

occurring in that 30-day window, it would have only added .5 percentage points to the national average, so going from 21.1 percent up to 21.6 percent.

In summary, this measure is designed for quality reporting purposes and monitoring of fee-for-service provided to skilled nursing facility beneficiaries. It focuses on unplanned readmission measures, which are more likely to be attributable to the quality of care being provided in the facility. The measure will provide valuable information to patients and families about the quality of care in the SNF and encourage shared accountability across providers.

Thank you.

CO-CHAIR HALL: Thank you. I'm going to ask Helen and Carol to open group comments. And, again, at the moment we're in the category of evidence.

DR. CHEN: So I think we are actually traversing territory we have

traversed earlier today and probably will continue to traverse over the next day or so regarding the evidence. I think it's clear that this is -- there is some degree of performance gap here. And also, in terms of the evidence, a number of the workgroup members were concerned about the inference that some of the evidence presented was related to studies done about acute care transfers, not NSF care.

But, nonetheless, it is pretty clear that there are processes that would improve transitions, communication, and actual SNF care, for example, nurse staffing ratios, as the developers have mentioned in their measure report.

MS. RAPHAEL: Well, there are different studies. I think the consensus is that there are a significant number of unplanned admissions to hospitals coming from nursing homes and that we can make headway in this area. So I think our group decided to

forge ahead here.

I think there is one thing I wanted to raise in sort of looking this over. We have an overall exclusion rate of 20 percent. That seems like a high exclusion rate, but I just would like you to comment on that exclusion rate.

DR. LAURA SMITH: So we have -- we do have multiple exclusion criteria for this. I believe that the major exclusion criteria that impacts this measure is the requirement for having 12 months of claims in the prior -- prior to the acute hospitalization.

And so this was one of the decisions that we had made in terms of identifying comorbidities that you -- there is prior literature showing that you get a far more effective prediction using 12 months of data compared to just the most current, most recent hospitalization. So that was the -- I believe that's the criteria that drops the most. Yes.

CO-CHAIR HALL: I don't see any other cards up yet, so I'll start with a question myself. The time horizon is 30 days from the hospital discharge prior to SNF, and so I'm not expert enough to know whether all SNFs happen immediately, all SNF admissions happen immediately. But you could have the sense that there is a different number of days at risk for different patients because the time horizon does end at 30 days from hospital discharge. Does the model account for that concern?

DR. LAURA SMITH: So because we track every individual who has a qualifying SNF stay for that full 30-day period, because we are using the claims to identify readmissions, we don't have the same issue of there being variable time at risk that everyone has 30 days at risk.

CO-CHAIR HALL: I understand. So the assumption, then, is that everyone moves from their acute care hospital discharge

directly to a SNF.

DR. LAURA SMITH: Oh. Yes. And so it's not an assumption as -- because that's one of our other exclusion criteria, which is that we require that the admission to the SNF be within one day.

CO-CHAIR HALL: Right. Okay. Great. Thank you.

Any other concerns or questions around the group?

(No response.)

Not seeing anything, we'll move to vote on evidence.

MS. SHAHAB: Voting is open for 1a, evidence. One is yes; two is no. And the time begins now.

(Pause.)

We are just waiting for one more response.

(Pause.)

Voting is now closed for 1a. It was 23 yes, zero no.

DR. CHEN: So moving on to the performance gap, it did seem that there was a fairly measurable performance gap in the data that was presented with a standardized readmission rate range for the 2011 data set of 11.9 to 41.9 percent, which is a pretty big swing in comparison to some of the other measures that we have talked about today and an opportunity for quality improvement across facilities.

CO-CHAIR KAPLAN: So based on Vince Moore's study, the estimate is that 20 percent of readmissions -- 78 percent, so 20 percent overall readmissions are preventable of the 26 percent. So that seems like a very, very -- I mean, 78 percent of these readmissions are preventable. That seems like a fairly incredible statement. Is there any -- do you have any evidence from the data that -- or other evidence that you would cite that supports that number? Thank you.

DR. LAURA SMITH: So my

recollection of that paper has a lot to do with the fact that there are a lot of sort of chronic conditions like CHF and COPD that are showing up as reasons for admission. That was -- in truth, that is our major citation that we were using. I think that we do see variation in readmission rates by facility characteristics that have been identified in other studies as being associated with quality of care like staffing ratios and other sort of managerial characteristics.

So I -- it is not as direct as sort of giving you a percent that are avoidable, but I think that we do see variation from other -- in other studies that are associated with things that are associated with quality. So --

DR. CHEN: I also found those statistics in Vince Moore's study kind of astonishing as well, anecdotally, as an n of 1 experience. For whatever it's worth, we do 100 percent case review of all readmissions

from our post-acute care facilities, and I would say our readmission rates are fairly low. But, on the other hand, the percentage of potentially avoidable readmissions probably runs in the 25 percent range.

MS. RAPHAEL: I was just going to say that anecdotally what you observe, at least at acute care hospitals, is that you have high readmission rates for UTIs and pneumonia, and sort of the hypothesis being that with different staffing that nursing homes could handle those cases at their facility and not send them to the ED or to be readmitted. That's more anecdotal than scientific evidence.

CO-CHAIR KAPLAN: Other comments?

(No response.)

Okay. We are ready to vote.

MS. SHAHAB: For performance gap, one is high, two is moderate, three is low, four is insufficient. And your time begins now.

(Pause.)

We have all 24 votes. For 1b, performance gap, 18 voted high, six voted moderate, zero low, and zero insufficient.

CO-CHAIR KAPLAN: Great. So now is this a high priority? Any comments before we vote? Oh, yes. Sorry.

DR. FIELDS: This is maybe more of a point of information. I would just be curious about whether we would assign this measure a different priority than the one we did before lunch, and maybe talk a little bit about process. I am kind of assuming we are going to stick to our knitting and not compare the two measures or think about how they would both be implemented or whether one would be implemented. But can you help me out a little bit about how process would work if we wind up approving both?

CO-CHAIR KAPLAN: Can you give him the harmonization, homogenization, whatever it is?

DR. AMIN: Yes. Thanks, Sherrie. We have a harmonization and competings measures discussion. So NQF has an algorithm. We, obviously, need to be respectful of the concerns around measurement burden. If there are measures that address the same measure focus and the same target population, we will have to go through an evaluation of whether or not the measures have been harmonized to the extent possible.

This measure and the measure before it would obviously fall in that category, and we have alerted the developers of this prior to -- prior to this meeting. Obviously, they have differences. They have differences in data source and other elements to it. So we would have more of a conversation around what elements of the measures can be harmonized going forward.

Karen, did you have something?

MS. PACE: But it would also be a competing measures issue, right? Not just

harmonized?

MR. AMIN: We could talk about that. But we don't get to that discussion until the measures are recommended for endorsement. And, also, I would just caveat as, you know, we might have to -- that might be a conversation for after the in-person meeting, depending on how much we're actually able to get through during this discussion.

CO-CHAIR KAPLAN: Leslie?

MS. HALL: Just follow up on that and maybe have the same -- added to that discussion is, what is the total number that constitutes a burden? If you have five that are on one item, or two on one item, or 20 overall, what are the definitions of burden? So that we don't just keep adding and adding because individually they are sound, but collectively they create burdens. So in the harmonization process, do we have a calling process as well?

MR. AMIN: Not necessarily. I

mean, the way we're thinking about it is over -- the overarching question is, if there are measures that do address high priorities that they still have a performance gap, that these are still areas of measurement that are important.

So we don't have an overall target of what the overall portfolio should be in terms of terms, but -- you know, so what we're really trying to do is when there's cases that the measured focus and the target population are similar, we -- that's the area where we are going to really have more of a head-to-head discussion.

MS. HALL: But you do have a best in class process, right?

MR. AMIN: Yes.

CO-CHAIR HALL: Yes. So if I may, I mean, two measures could be too many if you decide that that's the case and you think that the two measures are competing and there is only a need for one.

CO-CHAIR KAPLAN: Other conversation? Alison?

MS. SHIPPY: I just wanted to piggyback on the comments about measured burden. I just would also want to throw out the consumer interpretation is something to take into account.

CO-CHAIR KAPLAN: Good point. Anything else?

(No response.)

Okay. We are going to vote on high priority.

MS. SHAHAB: For high priority, one is high, two is moderate, three is low, four is insufficient. And the time starts now.

(Pause.)

We have all 24 votes. For 1c, high priority, 19 high, five moderate, zero low, zero insufficient.

CO-CHAIR KAPLAN: Excellent. Scientific acceptability for reliability.

Helen?

DR. CHEN: So as mentioned in the presentation, the overall ICC was 0.56, which I believe is a little low-ish even for readmissions, measures, as we have discussed earlier today. But what concerned me was the range. So the range of the ICCs that you reported was 0.3 to 0.7. So some worse than chance or some -- you know, not even chance. And I just wondered if you could speak to that a little bit.

DR. LAURA SMITH: Sure. So the results that we are talking about right now are of the split sample test/retest where we combine the 2009 and 2010 data, split it randomly, and then look to add the agreement between the facility scores. So the .56 is for the overall ICC for the measure, and then the range that was reported, what we did was we stratified by facility size and so we were seeing ICCs of .7 for our largest category. I am trying to pull out my table here, so I

can give you a little more specifics.

The one thing that we did do when we did that analysis -- I just want to see whether -- for some of our analyses we did not exclude small facilities. We actually included all facilities, regardless of how many -- how many stays were included. And so I believe that that is part of why you're seeing -- okay.

So there were -- for that smallest range, the .30 was for SNFs that had one to 44 stays in the denominator. So a certain amount of those are going to end up most likely not reported because there tends to be a 25 stays cutoff for the size that would be reported publicly.

And so the -- for SNFs that are 45 to 91 stays, the ICC was .45, to give you the next level up, and then 92 to 171 is .53. And SNFs with 172 to over 1,000 was .70.

DR. INGBER: Let me just add that when you talk about chance this isn't the same

kind of a measure. What we're talking about is when you have a relatively small population and you just randomize it into two buckets. The chances are those buckets are going to look pretty different. So you're not going to get the same picture, even if you risk adjust it really well. Some of them just randomly will have gone to the hospital and others not and, you know, so the numbers, especially in the smaller facilities, do look funny in all the measures.

CO-CHAIR KAPLAN: So with respect to reliability, just to clarify, as expected, your reliability of your estimates for low volume hospitals is crummy, right? Not to put too fine a statistical point on it, but that's what the -- that's what that means.

DR. INGBER: Yes. Because reliability, in the sense of every time you shake up the patient pool and look at it again you're likely to get a somewhat different measure. Yeah.

DR. CHEN: I think you -- for the IRF measure, you talked some about the effect of shrinkage, and I know you sent the additional information to the group regarding that, the effect on this particular measure. Can you delineate that some more for the discussion?

DR. INGBER: The shrinkage effect in terms of the SNF measure?

DR. CHEN: Yes.

DR. LAURA SMITH: So for the SNF measure, our results were similar, I believe, to what we saw with the IRF and LTCH, except for the impact of the shrinkage in the smaller stay sizes. So we have stratified it by decile of the number of stays. We had a fairly small decile, one to 21 stays, and so we were definitely seeing a fair amount of the score that was being explained by shrinkage.

And we did see some shrinkage also at the highest end of the distribution as well, but otherwise fairly consistent with

what we had seen in other -- for the other measures. Mel, did you want to add anything?

DR. INGBER: Well, what I was just going to add is that there is some shrinkage even at the high end, but it isn't a lot. It's relatively small compared to what the risk adjustment is doing. The risk adjustment is doing its job irrespective of size. But, yes, as soon as you get to the small ones, risk adjustment moves it, and then shrinkage says no -- not a lot of data here.

One to 21 is much lower than we even had in these other measures that we presented earlier. That's really pretty tiny and very unlikely to show up on anybody's measure list of reporting.

CO-CHAIR KAPLAN: Frank?

DR. BRIGGS: So in terms of the obs patients, and the shift to obs versus in-patient, say, with the two-minute rule, I think that has the potential to have a big impact on your data set. One thing I'd be

concerned about would be having more hospitals now falling into that smaller range, because you're excluding these patients who aren't being admitted back into observation, or being admitted into observation rather than in-patient status.

Any sense -- is there any data around the use of observation status from SNF as opposed to the general population? You know, is it an issue, or is it not an issue?

DR. LAURA SMITH: So we have -- in the measure submission form, we do cite some somewhat older analyses where what was seen in the 2009 data was that the vast majority of observation stays were coming from the community and also not being discharged to SNF, that that was -- I can't remember the number off the top of my head, but discharge to SNF was not one of the major destinations now. Of course, that is a bit older data, and we know from the GAO report that that -- the patterns have been changing.

The analysis that I was just referencing, we are really just looking at our samples that are in the 2011 SNF RM measure sample. We are not seeing a large proportion of that set of individuals getting sent back to the hospital for observation stays. So it still seems relatively small at the moment, but it certainly seems like something that would bear ongoing monitoring, and certainly with the change in the policy forthcoming.

CO-CHAIR KAPLAN: Thank you. Other comments? Wes?

DR. FIELDS: Well, there are some recent data, Frank. I'm not sure if it's somewhere in the packet. It actually came from CMS from recent morbidity and mortality reports, kind of intriguing really. So it's much more recent claims data, and it suggests that readmission rates are falling, as you would expect, as hospitals change behavior.

What is kind of interesting about it is that the rate of observation services,

according to the CMS analysis, isn't rapidly rising. But the other thing that's sort of curious is that emergency department visits aren't rising either. So the question becomes, how much of that improvement and readmission rate is happening in community settings, primary care, case management? And how much of it is happening around what -- depending on the measure and the context we're calling ED visit that is actually an observation stay that could be 24 or 48 hours long, but less than, you know, to midnight.

So I think this is a huge moving target, and I think the ultimate answer to your question is really important.

CO-CHAIR KAPLAN: Paula?

MS. MINTON-FOLTZ: I'm curious about the inclusion of psychiatric hospitals in the measure. Psych patients are really one of those groups that are really hard to place already in SNFs, and so I'd be interested if other people thought that that might result in

less access for these patients in SNF.

DR. LAURA SMITH: I think we might have to get back to you on the -- it's a small proportion of who we are seeing in the sample, but I don't remember, so I would have to get back to you on that.

CO-CHAIR KAPLAN: Paul, and then Thomas.

DR. HEIDENREICH: In terms of the distribution, the timing of readmission, do you know what percent happens, say, on the first 24 hours? And, you know, what's the thought in terms of the nursing facility being able to influence that as opposed to the transferring facility?

DR. LAURA SMITH: So as you would probably expect, there are -- there is a fairly large proportion that take place within the first 48 hours. I don't have a percent here. There were some supplemental responses that we can get you a copy of.

It looks like the peak is more

around the second or third day. So, I mean, we -- this is something that we talked about with our technical expert panel, that certainly there was some discussion about, should those first 48 hours not be included in the measure.

We got a really strong message from our experts that they should be included because that's really part of this idea of shared accountability in terms of the quality of the transition. I think there could be some debate about sort of how much flexibility do SNFs actually have in choosing to accept or not accept someone into their care. But the message that we got was that we should be holding the SNFs also accountable for the services provided in this first 48 hours.

CO-CHAIR KAPLAN: Okay. We have five minutes left to discuss this measure. So, Thomas, do you want to give us a concise question/comment?

DR. THOMAS SMITH: Uh-oh.

(Laughter.)

Well, just back to the psychiatric admissions. So why not exclude admissions to psychiatric hospitals, given that it's a different population and the circumstances around admissions are often so very different?

DR. LAURA SMITH: Oh. So I should clarify. So the psych admissions -- the admissions to psychiatric hospitals are -- we basically consider them to be planned. They are not counted in our numerator. It was that you could be included in the measure if you had -- your prior hospitalization was from a psych hospital.

DR. THOMAS SMITH: I'm not following.

DR. LAURA SMITH: So --

DR. THOMAS SMITH: You're in the -- when are you in the numerator, and when are you in the denominator if you're a psych patient?

DR. LAURA SMITH: You're in the

denominator if you are a psych patient. But if you are -- if your readmission is to a psych hospital, that is not counted in the numerator.

DR. THOMAS SMITH: At all.

DR. LAURA SMITH: Not for this measure.

CO-CHAIR KAPLAN: All right. Comments on reliability? Because I think some of these issues touch on validity, but we are going to -- first we are going to vote on reliability, unless there is any further discussion. Ready/set?

MS. SHAHAB: So for reliability, one is high, two is moderate, three is low, four is insufficient. And your time begins now.

(Pause.)

We have all 24 votes now. For 2a, reliability, five were high, 18 voted moderate, one low, and zero insufficient.

CO-CHAIR KAPLAN: Thank you.

Now, with respect to validity, Helen, do you want to make some comments?

DR. CHEN: I appreciated that the developers actually tried to do some construct validity using other measures of quality. That was both reassuring and also comforting to me, that there was a reasonable low correlation in the expected directions with the exception of pain management. So that was good. Thank you.

In terms of discrimination calibration, the C statistic was 0.67, which is in range.

And the one concern I had was regarding the exclusions. Although the relative -- the standardized risk readmission measure didn't seem to change much in terms of looking at your exclusions in terms of the absolute change, there were some changes in the quintile ranking in terms of people going up or down, and that was a bit of a concern to me. Can you speak to that, please?

DR. LAURA SMITH: So the quintile ranking I believe changed the most when it came to the gap exclusion criteria. So we talked about that a little bit earlier, that if a patient had a gap of more than a day they were excluded from the measure. And so we were sort of -- we were torn in terms of that issue of there being differences in the risk for readmission based on the time since the prior acute discharge, that you do see a declining risk in readmission over time, and so we went with a day gap for that exclusion.

I think part of the other reason why we -- you potentially see some of the quintile rank changes has to do with there being somewhat of a smaller range in the quality measure. So if you have -- even with a small -- we only had a change of more than one percent absolute change for 30 facilities, that if they were in the middle of that range of the distribution that it would be fairly easy to potentially move in quintiles, if

you're in that mid-range, just because you might have -- there's some clustering in the center of the range for our measure.

CO-CHAIR KAPLAN: Thank you. Other comments?

(No response.)

I have two quick questions, and really quick. One is the -- 90 percent of the facilities were significantly different from the national mean. And, therefore, does it make sense to use the national -- you point out a lot of geographic variability, and then you use the national mean. Can you help us really quickly and concisely understand what the advantages of that would be?

DR. LAURA SMITH: I think that our using that as the national -- as the thing to compare it to isn't necessarily an endorsement of using that. But I think it's a useful way of trying to determine whether or not you can discriminate amongst providers. I think that is a good point, that it becomes less

meaningful if everybody is different from it.

CO-CHAIR KAPLAN: That's more of a use question, so that's not fair. So the largest correlation, however, was with RN staffing, and that was a negative .13. So 1.7 percent of the variability is attributable to -- the largest variation in validity is attributable to staffing of the variables you tested. That's not -- that's pretty weak. That's not -- and that was the strongest variable. Can you help us understand all of these correlations are weak -- or what would you grade your validity evidence as?

DR. LAURA SMITH: So in terms of our validity results, they are consistent with prior studies of these quality measures, that we tend to see low correlation for -- and particularly the MDS-based ones. I do think the best that we can do is sort of look at, did they go in the hypothesized direction?

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

(No response.)

Okay. Can we vote a validity score?

MS. SHAHAB: For criteria 2b, validity, one is high, two is moderate, three is low, four is insufficient. And your time begins now.

(Pause.)

We have all 24 votes. For 2b, validity, zero high, 17 moderate, seven low, and zero insufficient.

CO-CHAIR KAPLAN: Thank you. Now we're on to feasibility. Comments?

(No response.)

Okay. Let's vote.

MS. SHAHAB: Haven't started yet. For feasibility? Okay. Three, feasibility. One is high, two is moderate, three is low, four is insufficient. Time starts now.

(Pause.)

Still need two more. We have all 24 votes. For feasibility, 14 voted high, 10

voted moderate, zero low, and zero insufficient.

CO-CHAIR KAPLAN: Thank you. On to usability and use.

DR. CHEN: This gets at the issue that you were raising, Sherrie, and I think that others have raised in other measures discussions regarding, what does this mean? And what is the consumer going to -- how is the consumer going to interpret this?

And in terms of using the expected, better than expected, worse than expected, and thinking about, you know, benchmarking for this particular measure, I think that's a challenge not just for this measure but for other measures in terms of the overall use for this.

And, again, if people are very different from the mean, what does it mean?

CO-CHAIR KAPLAN: Karen?

DR. JOYNT: I'd like to second that, and also say I think it gets back,

again, to the whole shrinkage issue, which is that if there are hospitals about which you don't have very much information, it seems to me like labeling them that they are the same as everybody else, is actually pretty misleading.

Now, in this case it looks like the small facilities actually had lower readmission rates than the large facilities, unless I'm misreading the supplemental information. So this is not to say that I pretend to understand enough about nursing facility care to know which direction it should go in.

But I do think from a usability standpoint that telling the consumer that we're certain that the -- that a place is not bad is potentially problematic. And so I would just encourage as we think about the usability of these measures to consider really how certain we're willing to say we are about low volume facilities and how we would

communicate that to the consumer.

CO-CHAIR KAPLAN: Larry?

DR. GLANCE: So I agree with Karen's point. I think it's an important point. It's one that we spent a lot of time talking about last time around, the whole issue of shrinkage coefficients and shrinking low volume providers back to average. But having said that, since virtually all -- well, not all, but the vast majority of the measures that we are looking at today are based on hierarchical modeling, I don't think we should -- I hate to use the word, but penalize this particular measure because of that, because most of the other measures also use this same methodology.

CO-CHAIR KAPLAN: Leslie? Thank you.

MS. HALL: I would just echo on the consumer side -- and this relates to all of them, but we don't want to send the message that being hospitalized longer is better, and

that consumers could be confused by all of this pushback against other areas of admitting.

CO-CHAIR KAPLAN: Thank you. Other comments?

(No response.)

Okay. Are we ready to vote? Oh, sorry.

DR. INGBER: Just wondering, some of the comments seem to be confusing me as to what we are looking at. To say something is -- you can't tell if it's different from the average or not different from the average, let's be very careful. Most of these people you can't say are different from any number you would pick out of a hat, because they have a very wide confidence interval.

So the idea is not to tell everybody that they are average. It's to tell them something about them that we can't tell that -- whether they are average or not average or express it in a way that doesn't

say these guys are average because we can't tell, because we could just as well say, well, they're not 20 percent below or 20 percent above because we can't tell.

So it depends on how you use the measure and what benchmark you use. It's not a bad thing about the measure; it's how you use it. I just wanted to make sure that was clear.

CO-CHAIR KAPLAN: Right. The implementation issues, obviously, are -- and the same issue is going to come up over and over and over again with regard to how confident we are in your estimate at a given point in time, and that is true across measures and is partly the job, right, Helen, of the map group?

DR. BURSTIN: Yes.

CO-CHAIR KAPLAN: So for now, as written, we have to vote on use -- I'm sorry. Any other comments?

(No response.)

Okay. We can vote a score.

MS. SHAHAB: For criteria 4, usability and use, one is high, two is moderate, three is low, four is insufficient information. And time starts now.

(Pause.)

We have all 24 votes now for criteria 4, usability and use. One high, 16 moderate, seven low, zero insufficient information.

CO-CHAIR KAPLAN: Overall suitability for endorsement. Discussion?

(No response.)

Hearing none?

MS. SHAHAB: One is yes, two is no. The vote is now open.

(Pause.)

All 24 votes are in. For Measure 2510, skilled nursing facility, 30-day all-cause readmission measure, 19 voted yes, five voted no.

MS. TRAVIS: Can I make a comment

before we get to the next one?

CO-CHAIR KAPLAN: Go ahead.

MS. TRAVIS: In the discussion about use and usability about a number of these measures, we have brought up the map process. And having been part of the map process, I just would like to ask that perhaps one of the improvement initiatives we can put into the map process is to have a better understanding of the use and usability discussion that goes on in these standing committees, because it is an endorsed measure, it's in the set. It is not necessarily transparent and apparent to people that these kinds of conversations have gone on.

So I think as we continue to have this, that's just one of the loops that might be better to pay some more attention to at the map.

MR. AMIN: Thanks, Cristie. Well, this certainly is an area that we are working on internally to be more robust.

For the general committee, just want to make sure we are all sort of on the same page. The measures application partnership, which NQF convenes, makes recommendations to HHS around selection measures for various different federal programs. And we will be bringing the recommendation guidance here on the use and usability of these measures to the appropriate workgroups that are making recommendations on these measures.

And there are many of you around the table who are obviously playing leadership roles on those committees. So we'll make sure that that is connected as best we can going forward. But it certainly is another area that we are focused on. Thanks, Cristie.

CO-CHAIR HALL: Thank you all. Another great discussion. Thank you, our developers from RTI.

So we're a few minutes behind after our first measure. You may think you

are seeing a meaner version of Bruce Hall now that lunch is over, and I keep harping on time. But one of my concerns is that we do have folks in the room who have come into the room to discuss as developers and to contribute. And if we don't stick on our timeframe, we may lose some of their expert input.

So we'll ask right now for our developers from University of Michigan to introduce their measure, Measure 2496, around dialysis. Please introduce yourselves and basically introduce your measure.

DR. MESSANA: My name is Joe Messana. I'm a clinical nephrologist on the faculty at University of Michigan. I work at KECC, U of M KECC.

DR. KALBFLEISCH: My name is Jack Kalbfleisch, or John it says here. And I'm a professor of biostatistics and statistics at the University of Michigan and also work at UM KECC.

So on the -- what we're talking about is a 30-day unplanned readmission measure for dialysis facilities. So it's looking at patients who are discharged from an acute care hospital to dialysis facilities and the rate at which they are readmitted to the hospital within 30 days. It is a measure of dialysis facilities and not of hospitals.

The measure is in the form of a ratio, so it's a bit different than some of the other measures. It's looking at -- it is relating the given dialysis facility to the national norm basically, to a national average and ratio, so the numerator of the ratio is the number of readmissions and the denominator is the expected number of readmissions, which comes from a logistic model. So that based on the number of discharges that there are and the type of discharges and the patients involved, patient characteristics and the hospital discharging are taken into account in getting that expectation.

Modality variation radiation is substantial among dialysis facilities with about 10 percent having a readmission ratio 30 percent above the national norm and about an equal number 30 percent less than the national norm, for example.

Patients at dialysis facilities have on average about two hospitalizations per year. And following a discharge from a hospital, about 35 percent are readmitted within 30 days. So it is a population where the burden of hospitalization is quite high, and so it's a population where there is potential for substantial gain, both in terms of patient quality of life and also in terms of cost.

The measure has been underdeveloped for some time, and we are currently just finishing a dry run of all -- of the measure in all dialysis facilities in the country, and it has been subject to feedback of various sorts over that period.

So I guess, like other measures, the SRR is really motivated in part by the aim to -- aim to encourage health care providers to work together to coordinate care. And in some settings there is substantial evidence that that works, and the general concept carries a great deal of face validity I think.

There is, however, relatively little direct evidence for dialysis facilities as for some other facilities that we have been discussing today. There is some. There is a paper looking directly at that question and an observational study, and looking at interventions within the first week following discharge and showing that those interventions are useful.

There is also a fair bit of opinion and qualitative studies looking at the issue, and in particular an excellent editorial in the Clinical Journal of the American Society of Nephrologists by Jay Wish, looking at that question and pointing out the

potential for gain through the involvement of nephrologists in the first week following discharge, talking in particular about getting away from a paradigm of resumed previous quarters at the end of discharge.

So at the preliminary meeting of NQF, and at other times, this measure has received many comments, a number of which were unique to it, although the measure itself is somewhat similar to many other measures that are being discussed. We have submitted through CMS a written response to the main points raised, and of course no doubt some of them will rise again. My time is certainly not sufficient to review them here.

I would like to -- just quickly on two comments, one which was made earlier this morning by Dr. Krishnan which related to peer review of the measure. In fact, it is based on statistical methods which are quite old and there are books on the subject, and it's the same kind of statistical method that is being

used in all of the measures that have been discussed so far today.

The particular measure itself actually was published. There was a report on it published in the -- in a statistical journal, statistics and the biosciences, which details the methods that are being used here and is a peer-reviewed journal.

Second comment I guess is the concern about potential adjustment for physicians, and I think that was something that came up with a great deal of discussion at the last meeting of this group. And so I guess one question being also whether there really was the regulatory framework in which dialysis facilities could actually control or exert influence on the physicians that are working at them.

So we note in the response that the regulatory framework exists, that there is such a framework. That was also noted on the call. And one of the primary purposes of this

measure really is to promote coordination and promote the kind of involvement at dialysis facilities with their physicians in the handling of patients following hospital discharge.

So including an adjustment for physicians we feel would -- may reduce the incentive to coordinate care, and it also would create a very large lack of harmonization with many other measures that are being considered here.

So, anyway, thanks for the opportunity, and look forward to the discussion.

CO-CHAIR HALL: So, again, we are in the category of evidence still. And our discussants Steve and Sherrie. Steve, do you have any specific comments on this category of evidence?

DR. FISHBANE: Yes. You know, if I might, because we are moving into a very different world here, the dialysis unit, I do

want to make some overriding comments which will apply really to every criterion that we look at. And then I promise I will get directly to evidence.

But, you know, I think what the TEP here really struggled with was, where should this reside? Should it be at the hospital, the physician, or the dialysis unit? And they kept coming up with ways of combining them because the ultimate struggle here is that it is just very hard to understand how the dialysis unit could have any process that would be able to influence this specific measure.

So I would like to remind everybody that this is not nursing home care. This is episodic care. Out of the 168 hours in a week, 12 hours are spent in a dialysis unit. The structure of a dialysis unit, there's nurses, there's technicians, there are dieticians and social workers. There is a medical director, and the medical director is

generally a .25 FTE. And that time is not necessarily spent in the dialysis unit. But even in the dialysis unit, they may have a few of their patients that they see in the unit, but they don't take care of other patients in the dialysis unit.

The treating nephrologist here is really key, because that's where the opportunity is to prevent readmissions. And this is really different than some of the processes that we have talked about so far. Yet the treating physician is only required to see the patient once per month, and there is no structure that provides the medical director or the governing body of the dialysis unit to compel nephrologists to see patients immediately after discharge, which is why I think there has been such a struggle with this measure in terms of, how do you put the dialysis unit as the resident that owns this.

Now, CMS has an incredibly positive record in dialysis, both through the

conditions for coverage, the regulatory authority, and through the quality measures that have been put forth to date of really substantively improving the quality of care. So if I'm going to be negative on this measure, which I will be on a number of facets of it, it doesn't reflect what is the overall great record that they have had. And, you know, I certainly have written a number of articles supportive of that.

So the structure of the unit isn't -- and the current regulatory structure, not set up to empower the dialysis unit to really be able to have any control over any processes where we can prevent rehospitalizations.

The hospital dialysis unit relationship is really important to understand here. It is not like other types of relationships. You heard Dr. Krishnan speak about that for DaVita only 50 percent of the time do they have any information that comes to them from the hospital prior to discharge.

You need to I think understand that there is no organic relationship between the dialysis units and the hospitals.

I run a hospital-owned dialysis unit, and we have some relationship, of course, with the hospital because of that, but that's the minority of American dialysis units. It's a many-to-one relationship. So it's not just one hospital that it returns their patients after discharge to the dialysis unit. You may have 10 or 15 different hospitals. And it's impossible -- usually we don't have information that comes from the hospital. When we do have information, it's only because they happen to be at that one specific hospital.

The power to prevent readmissions here is based on processes that the nephrologist controls. So very clearly reassessment of volume status leads to a lot of readmissions here. And if you can get the nephrologist to see the patient within a

couple of days after discharge, I think that would be the most powerful thing that could take place.

Medication reconciliation here is very important. Reassessment of the patient's medical status and reassessment of the patient's dialysis prescription, the developers here that spoke to Jay Wish's editorial -- and Jay's point was exactly this, that if you don't get the nephrologist involved early and not doing what happens mostly in the United States, which is just resume previous orders, you don't have the ability -- here you've got social workers, nurses, technicians. There really is no setup to be able to prevent readmissions.

So, you know, we speak of trying to improve coordination of care, and I agree that that sounds good, but there really is no natural ability that we have here to improve coordination of care. The hospitals are not pushing out information to dialysis units, and

most American dialysis units have no way of pulling the information. You would have to be making calls to a number of different hospitals asking them for information of what took place during the hospital stay.

Sometimes it will happen to be the same treating nephrologist in the hospital as in the dialysis unit, and that will make it easier. But that's very often not the case.

We talk about the interdisciplinary team. In the response of CMS, we certainly thought a lot about that. It gives the impression that there is a team that is waiting at the dialysis unit ready to reassess the patient and work together. In fact, that is not the case.

It is a virtual team, and what current relations are through the conditions for coverage -- that is our regulatory authority -- is that the interdisciplinary team once per year has to assess and develop a plan of care for a patient. For unstable

patients, you compel the nephrologist once a year to work with the interdisciplinary team. For the unstable patient, which means a 17-day hospitalization or three hospitalizations in one month, you compel the nephrologist within one month, not within a couple of days, to be able to see the patient.

So, you know, I think I will get straight to evidence, but we really worry about -- here about where this measure resides. I mean, I would say first there should be a measure for physicians; secondly, for hospitals. And it just doesn't seem like there is any actual process that under current regulation or practical structure that the dialysis unit would be able to change a process here to work.

You first need the right regulations and structure. If not in place, I can't see how you would have a quality measure. But as for evidence, as the way that we are defining it today, it is -- you know,

two issues that I would make here, and I'll keep it, you know, very limited.

There is only one study that has been done in dialysis units. It is pointed out by the developers. It was by Chan et al. from Priscinius. It was an observational study, and really very trivial results. They showed association between, for example, measuring a parathyroid hormone level and a reduced incidence of hospitalizations. I think it is really a surrogate for the fact that if you get the nephrologist in, if you're able to do that, you can reduce the readmissions.

But the way that we're defining evidence today, which is that there is a probable linkage between the ability to adjust to effect processes to improved outcomes, without having the ability, the right people in the unit and the ability to compel the nephrologist to see the patient, it is unclear for me with the evidence how that would

happen.

And I will, as a final point, simply point out here that there are current studies that are going on to try to look at preventing readmissions, including one by our group that is funded by New York State. And if you look at clinicaltrials.gov, all of them would require -- so the interventions that are being tested would require new models of care, which would require a change in regulation and a change in structure.

I'll stop there.

CO-CHAIR HALL: Thank you, Steve.

I'll ask Sherrie to make any specific comments in this category of evidence.

CO-CHAIR KAPLAN: I'm just going to add one comment. This is about where we were with the readmissions for congestive heart failure when we got pushback saying, you know, the hospitals have absolutely no control over primary care providers where, you know,

they are supposed to see patients within a certain amount of time.

So in terms of evidence building and evidence base, yes, there is only one study here, but this is pretty much where we were when we were looking at all costs -- or readmissions for specific conditions -- some of the specific conditions and the issues raised at the time. Now there are tons of studies out there. We are building an evidence based on kind of where we were then.

But I am not in any way defending this measure from that perspective. I am just noting that this is exactly the questions and the concerns that were raised some time ago with respect to readmission for congestive heart failure.

CO-CHAIR HALL: Thank you.

Kathy?

DR. AUGER: Thank you. So with the caveat that this is way outside of my sphere, listening to, you know, your arguments

sort of against this being in the locus of control for the dialysis unit, I hear you, but you are also sort of making an argument that potentially if this went on and became an endorsed metric and then maybe payment was in some way tied to it in the future, that that might be the impetus to actually get a nephrologist to see patients right after discharge or to get that care coordination to be much better.

So I understand that it's not the usual locus of control right now, but it might make it become one. I don't know.

DR. FISHBANE: Yes. I mean, I think that's a fair statement. But, you know, I'd like, you know, to just remind you of one point on this, which is that there is no regulatory structure to do that right now. So as a medical director -- and I've been, you know, in 20 years a medical director of six different dialysis units -- one of the frustrations we have is that if you don't

change what are the conditions for coverage, our regulatory authority, there simply is -- I could put a policy in place in my dialysis unit that says the doctor needs to see the patient within 48 hours.

The problem with that is that they are just not going to admit patients. And we have pretty tough policies in our units. They just won't admit patients to our unit. So without having the right regulation through the conditions for coverage, it would be wonderful to be able to push the doctor into that position.

I think a much more likely outcome here is going to be some type of change in structure where our nurse practitioners are put into dialysis units, and that we have some regulation and perhaps payment in order to have a clinical encounter, because I do think it's an important subject.

CO-CHAIR HALL: Thank you.

Frank?

DR. BRIGGS: Just a question, because I don't know. But having been on the hospital side with medical staff affairs for years, one of the most powerful things that the hospitals have over the head of the individual providers is credentialing privileging.

So are the nephrologists credentialed to practice and admit at the individual dialysis centers? Or is it just strictly a contract of employment of some sort?

DR. FISHBANE: Right. Good question. And, sure, they are privileged and credentialed to work in the unit. But, again, under the -- so we are very highly regulated under the conditions for coverage. We can compel in annual assessments of patients a monthly visit from the doctor.

We don't have any way of compelling a visit right after hospitalization, and that really is what is

needed in order to be able to reduce readmissions. And I think that that is where this will ultimately come from. But you're going to need a change in regulation and structure in order to have a quality measure that there is any action-ability on.

CO-CHAIR HALL: Carol, I'll have you make a quick comment, and then we'll have Dr. Messana from the developer side reply.

MS. RAPHAEL: How much time during a week does a dialysis patient spend in a facility?

DR. FISHBANE: Yes. So it's between nine and 12 hours during the week.

MS. RAPHAEL: That seems to me like a substantial amount of time, more than is usually spent at a doctor's visit or an outpatient clinic. What other levers do you have besides the lever of regulation to try to change the paradigm here during those 10 to 12 hours?

DR. FISHBANE: Yes. So the

conditions for coverage, as written right now, and, you know, I think you will hear some rebuttal on this, but they are very clear. And, you know, it's a frustration of medical directors throughout the United States, 6,000 medical directors. We wish we had the power to compel the physician.

Now, we're paying, so we're paying out of our practice to have a nurse -- not the dialysis unit, but a nurse practitioner in the unit who we're creating a checklist-based encounter immediately post-discharge. There is a number of dialysis units in the United States that have physician practice owned, hired, nurse practitioners in the dialysis units, and that gives you that opportunity.

But, for example, you know, in one of my units I've got 200 patients. We get about six discharges out of the hospital per week. I could get the nurse practitioner to see the patient, but the only leverage I would have is if I created a policy, which I'm

allowed to do, which says that physicians must see patients within 48 hours of discharge.

The physicians would argue that it's just not possible. They go to three different hospitals, three different dialysis units. We have rural units that are 80 miles away from, 160 miles away from a center. So that we really don't have leverage that is specifically on the physician.

We do have a medical director, which is a .25 FTE. The medical director, though, I think it would be inappropriate for them to see well patients who are discharged, because they don't have established therapeutic relationships with most of them.

CO-CHAIR HALL: Dr. Messana, do you want to reply to some of the concerns you've heard?

DR. MESSANA: Well, I just wanted to make three brief comments. I'm cognizant of the time. One, the comments that I have heard from Dr. Krishnan and Dr. Fishbane sound

to me generally like endorsements of the performance gap and opportunities for improvement. So I would echo a comment made earlier.

Secondly, I don't entirely agree with Dr. Fishbane's interpretation of the interpretive guidance. Having 15 to 20 years of experience in medical director roles in multiple institutions myself, I believe there is probably more leverage in those regulations than has been described, but that's an opinion piece.

In our response, we included the specific regulations for your perusal related to governance of the facility, the scope and performance of the interdisciplinary team, which is by regulation supposed to include a physician, not necessarily in the context or using the same paradigm that Dr. Fishbane describes where the physician -- the physician centric perspective that he has described, where the only real change that occurs in a

patient's plan of care is when a physician lays hands on, right?

That seems a little bit of a caricature of how we practice health care nowadays, certainly how I practice health care.

Now, the dialysis unit has to have a charge nurse to assess every patient before every dialysis treatment. The patients come in thrice weekly for an average of nine to 12 hours total per week, but they are seen three times a week in the dialysis facility for an in-center patient.

So there is great opportunity for there to be interaction between the dialysis staff and the patient, and the requirement that a registered nurse performs that assessment is in the regulations. And some states have even higher standards, but we will stick with the federal criteria.

And there is nothing that keeps that charge nurse or another member of the IDT

from communicating with the physician and asking for updated telephone orders until the physician can come in and evaluate the patient. So if you want to think in terms of a physician centric perspective in interpreting those regulations, that until I lay hands on them, lay a stethoscope on them, no changes can be made. Then, that is a huge performance gap and an opportunity for improvement.

If you think that it's up to the team, which is clearly under the control of the medical director and the governing body of the dialysis facility, as stated in the regulations, then you might develop a different caricature.

DR. FISHBANE: Yes. I would just like to respond to that. I would agree that there is improvement, and you'll hear me very strongly support that there is a performance gap and an opportunity for improvement. So I would agree with Dr. Messana's point on that.

Secondly, in terms of the charge nurse's ability to change any of the processes here that would leverage readmission rates, I might, you know, disagree in pretty strong terms with that. That is simply not the responsibility of the charge nurse, and it's not -- these are not processes of care, assessment of stability after hospitalization, a look at the patient's dialysis prescription and readjusting it. And, most importantly, the patient's volume status as it changes dramatically and it leads to rehospitalizations for volume overload at a high rate, that a charge nurse can work.

And, third, in terms of the conditions for coverage -- I mean, I could read them here chapter and verse. They, you know, simply -- they could be changed and they are wonderful conditions. But this is just one area. I mean, CMS has done so much great here that has improved care dramatically. This is just not a measure that the conditions

give us the ability to really move the physician into place. But I appreciate, you know, the very wise comments of Dr. Messana.

CO-CHAIR HALL: Thank you. We've got five minutes to get through this measure. Right now we are on evidence. Larry?

DR. GLANCE: I'll take 60 seconds. If you believe that performance measurement drives performance improvement -- and I think we all do here -- I think that measuring performance at a facility level has a lot of advantages over measuring at the physician level. And the advantage is sample size. If you don't have the sample size, and you won't have it at the physician level, you will not be able to discriminate between high quality and low quality providers. So you won't be able to drive performance improvement.

CO-CHAIR HALL: Okay. Well, we will have more categories that we can raise comments in. Let's move to vote on evidence.

MS. SHAHAB: For 1a, evidence, one

is yes, two is no. And time begins now.

Just one more vote, please. Okay. So for evidence, 17 voted yes, six voted no.

DR. FISHBANE: The second category is the opportunity for improvement, and here, you know, I think there is an opportunity for improvement. The overall readmission rate is 30 percent, 36 percent for hemodialysis patients. And that certainly there is demonstrated by CMS in UMICH that there is variability between units. So, you know, I don't see any problems here.

CO-CHAIR HALL: Any additional comments?

(No response.)

I don't see any immediate ones. We'll vote.

MS. SHAHAB: 1b, performance gap, one is high, two is moderate, three is low, four is insufficient. And time begins now.

We have all the votes. For 1b, performance gap, 15 high, eight moderate, zero

low, zero insufficient.

DR. FISHBANE: And, again, for priority, I do think there is an important priority area. You've got about 596,000 patients, high costs, so that's 38 percent of Medicare payments here are related to hospitalizations in general. So it seems like it's a large enough population, big opportunity, so the priority is probably fairly high.

MS. SHAHAB: Voting for 1c, high priority. One is high, two is moderate, three is low, four is insufficient. Time is on.

We have all 23 votes. For 1c, high priority, 20 voted high, three moderate, zero low, zero insufficient.

DR. FISHBANE: I'm going to handle reliability and validity separately here. Reliability I think is a little bit easier to deal with. The approach that was used was a bootstrap approach with resampling. The inter-unit reliability was measured to be

0.55, which, you know, I understand it's not tremendously high, but it seems like for most of our quality measures it is essentially what we get.

You know, I'm going to have to defer here to Sherrie, not being an expert on psychometrics, but it seemed to me that this was fairly reasonable.

CO-CHAIR KAPLAN: I agree.

MS. SHAHAB: If there are no other comments, we can vote on 2a, reliability. One is high, two is moderate, three is low, four is insufficient. Time begins now.

We have 23 votes. For 2a, reliability, five is high, 17 moderate, one low and zero insufficient.

CO-CHAIR HALL: Any commentary specifically to validity?

DR. FISHBANE: Yes. On validity, and, you know, here I'll have a number of comments, because I think, you know, the threats to validity here are the same kind of

comments I've been speaking about, which relate to first, again, the ability of any action-ability of the dialysis unit under current structure and practice and regulation to be able to actually move this measure.

I do want to remind people here that the early readmissions is about 16 percent. These are not people that are living in a dialysis unit. Of patients discharged Friday, dialysis unit doesn't see the patient until Monday, and there is no way to be able to touch the patient early.

Again, the TEP here spends a lot of time on validity talking about the fact that their dialysis unit shouldn't be holding this measure, that it should, at the very least, be accommodation of the hospital, the M.D., understand some of the practical issues in doing that. And we probably just have to yield to that. But, again, it's the lack of action-ability at the level of the unit.

The TEP voted seven out of eight

low or insufficient on this one. I think there is a lot of issues. The heterogeneity of dialysis units is important. So that in many American dialysis units there are no peritoneal dialysis patients. It is hemodialysis where there is a 36 percent rate of readmissions. In our main unit, we have 25 percent peritoneal dialysis patients, so I shouldn't argue on this because we are going to look very good on this measure, peritoneal dialysis patients having the lower measure for readmissions.

But there would need to be -- and I may just have missed it in the specifications. I'll turn to the developer on this one, if in fact there is adjustment for that. A lot of facility-related issues here. There great heterogeneity in terms of American dialysis units, so you've got units that reside -- that are free-standing units, you've got nurses that are nursing home based units.

The nursing home based units tend

to have very bad quality measures, and for readmissions, where the patients are rehabbing, they are very sick patients, they are going to get readmitted at a very high rate, and the SRR would be very difficult there.

The size of their dialysis units is tremendously different. We've got units that are 200 patient, units that are 20 patients, and it's very hard to account for that. We see very difficult problems in terms of what is going on right now where dialysis units are blocking discharges, more related to financial issues than to quality measurement, but what is happening is if the patient is still on antibiotics or the patient has other medical conditions going on, the hospitals will try to pitch these complex patients to dialysis units where they are blocked. I would really be concerned here about the tradeoff between length of stay.

Fifth, I will, you know, remind

you that in Washington, D.C. or New York there is a lot of dialysis units close to physician providers. However, many American dialysis units are rural, and doctors may travel, you know, once a month, sometimes by plane, 300 miles to get to a dialysis unit and see patients.

And so there are a lot of threats to validity here. I think a lot of it gets back to the same question of, you know, how can this unit -- how could this particular measurement reside where there just isn't the ability to push the physician to be able to see a patient and whether -- is there validity present.

CO-CHAIR HALL: Kathy?

DR. AUGER: One other follow-up question on heterogeneity of dialysis units. Since pediatric patients aren't covered by Medicare, are they included in this? Are pediatric dialysis units included? And can you comment on, should they be compared in the

same metric as the adults, since the reason for dialysis in pediatric patients is just so different than for adults.

CO-CHAIR HALL: I'll ask the developer to make a note of that. Sherrie?

CO-CHAIR KAPLAN: With respect to sort of issues about staffing, et cetera, et cetera, et cetera, I was hearing opportunities for quality improvement as well. And attribution to a physician versus facility is also -- you know, nursing homes, and having physicians visit patients in nursing homes, et cetera, et cetera, one actually could make the same kind of argument for many, many, many of these measures, including -- and I was harkening back to the congestive heart failure readmissions measure, because that was also pushed back on well, as I said, with respect to being able to control the feeders into and who is going to see patients after they leave the hospital.

And yet what we are seeing is a

fair amount of attention to outreach for telemedicine, for nursing calls, for people that the hospital employs that now try and reach out to those patients to make sure they are not gaining weight in that first 48 hours, et cetera, et cetera, that actually have done a lot to help out with reducing the risk for readmission.

So I was -- one concern that we are holding -- I understand the unique issues of the dialysis facilities, but I was concerned that we are holding this measure to a different standard, and may be missing an opportunity to actually start getting the message out there that this could be an opportunity for us to really restructure how we -- those linkages and continuities and everything else that the quality improvement would suggest is -- would really help patients out.

DR. FISHBANE: Yes. And I would only, you know, say to that -- but I do think

that this is critically important, is that when you look at the studies that have been put forth here by the developer, so many of them relate to hospitals calling patients after discharge to CHF centers that have been set up, usually by hospitals in order to be able to prevent CHF discharges and telephone calls that are made.

And I think a lot of this, you know, Sherrie, falls into what you are saying. The dialysis unit having the very different structure of being the receiver on this end, and not -- you know, the skilled nursing facility has an incredibly organic relationship through all that communication that takes place from the hospital to be able to actually get a patient to a skilled nursing facility.

We heard the representative of a company that treats 40 percent of American dialysis patients that they don't get any discharge information in 50 percent of cases.

And where I have a more natural relationship is hospital-owned dialysis units. Because of the many-to-one relationship, we don't really have the ability. So what would be the performance improvement initiative?

And, again, I would just like to point out that what you see being studied in the United States right now are changes to the model of care that would require changes in regulation, because you'd be forcing nephrology practices, for example, to pay for and put a nurse practitioner into the unit to be able to see the patient early on. So I do think it's different than many of the measures. But, you know I appreciate your point.

CO-CHAIR HALL: I'm going to ask Karen and Paul, and then we will have our developer respond quickly to what they've heard.

DR. JOYNT: I just had a couple of quick points. One, I think that it is worth

thinking about how this is different from heart failure and other conditions, because we are talking about putting the onus on, primarily on, the outpatient provider. And for heart failure we still kept you in the hospital.

So this is actually a bigger shift than just shifting conditions. It is actually shifting to a completely different -- it is almost more like shifting to a patient-centered medical home accountability for readmissions, as opposed to shifting to a condition-specific thing, but keeping it with the hospital. That said, I think that is sort of where we're going in general, right?

So I don't know that I think that's a bad thing, but I do think we should be thinking about it in terms of how we are restructuring in terms of, sort of, ACOs and patient-centered medical homes and making a responsibility for our population.

Dialysis may be the best place to

start that experiment. Whether or not it's ready for prime time as a quality measure versus a demonstration project, not being a nephrologist, I can't comment on some of the financial arrangements.

The other sort of validity concern I would just like to bring up, though, is the competing -- two issues. The competing risks issue of things like mortality and the, I guess, competing risks of things like admission. So if you're a dialysis unit that does a great job keeping your patients out of the hospital, your readmission rate may go up as you get better. And that, to me, is a real threat to validity for something like this where you are talking about a population whose denominator will get higher if they do it that way. You know what I mean. If you do worse on admissions, you might look better on readmissions, and so just how you guys have thought about that would be helpful.

DR. HEIDENREICH: As a

cardiologist, I'm a little uncomfortable holding up heart failure as a success for readmissions. I will say, making it a performance measure, and particularly tying reimbursement to it, has made a lot of hospitals do a lot of things we hope -- I think that are good. I think the jury is still out, though, on how effective they were and whether overall health was improved.

And, particularly, since we see hospitals with high readmission rates have a slightly better mortality rate, which is true for -- not true for MI or pneumonia, but seems to be true for heart failure, so I think you definitely can force action by making this a measure. I'm not -- you know, I think it would be nice to be more confident that that action will lead to improvements.

CO-CHAIR HALL: Thank you all. We're 10 minutes behind. We'll have our developers respond to what they've heard.

DR. KALBFLEISCH: Okay. I heard a

lot, but I will try to be brief. I mean, I think we do see the measure, I think, as being one which could promote changes in patterns of care. That is perhaps its largest aim, really, is to try and bring about change. And I think a lot of the comments that related to that, and that have been -- they have been very good comments about the potential for that as well.

Early readmissions, I guess, was one that is measured. That is one that we have wrestled with a bit, and we commented in our response actually about that, because, as dialysis facilities currently are structured, I think it is hard to see how they can -- they can easily address admissions within the first three days of hospitalization. And to some degree, there is sort of a policy issue there I think that -- is the measure trying to change the pattern of care there, and I think that's -- I think that's really the policy question. And I think it's -- we suggested

that we should probably think of it as including that, with the aim of making that kind of change. And I think that's the policy that CMS wishes to pursue.

DR. MESSANA: So the only thing I will add on that point, though, is that, as we have demonstrated in the followup materials from the last call, we can identify those hospital readmissions that occur within that time period, and we have done analyses comparing, calculating the measure both ways.

DR. KALBFLEISCH: Yes. The measure certainly could be defined either way. I mean, we have dealt with that in some -- they are the highest -- days of highest readmissions, actually, are the first three or four days of the period following discharge. Peritoneal dialysis was mentioned. We don't adjust for that in the model. We have looked at it, and overall they have a slightly higher readmission rate than the -- than hemodialysis patients, but not really much difference.

Let's see. Oh, the question of readmissions and admissions is a very interesting point, because I think when one looks at readmissions the denominator really is a random quantity, but it's the number of discharges from the hospital. And if you reduce the number of hospitalizations, you reduce the number of discharges, so you change the denominator.

And that is an issue I think with readmission measures like this. I think that in the context of dialysis facilities, probably in SNFs as well, and other such secondary care facilities, I guess our view is we have as well a measure of hospitalizations, the standard hospitalization rate, which is actually looking at the rate at which admissions occur in dialysis facilities.

And we really think of looking at those two together, so the one is telling us about the overall use of hospitalization by the facility; the other is trying to look more

specifically at the readmission process, so looking at two different aspects of hospitalization. That is -- so taking the two together, I think, is sort of something that is -- we recommend and I think makes a lot of sense.

Pediatric patients are included, and they are adjusted for in an age adjustment in the model. I think that's -- oh, geography was another one that was mentioned I think, and we haven't seen I think specific differences in rural versus urban. But I'm sure that the challenges those facilities face are quite different.

DR. MESSANA: I would add one additional comment, or a little bit of additional detail about the pediatrics. Patients under age 18 account for on the order of 1,400 or 1,500 dialysis patients in the United States out of a denominator of perhaps 450,000. The majority of those patients are dialyzed in adult-predominant facilities.

There are some or a handful, less than 100, probably in the twenties or thirties, facilities that are predominantly or completely pediatric. Generally, they tend to be small units. They tend to take care of a lot more infants. And since there are a minimum number of observations, criteria built in as exclusions, many of those facilities would not be reported on just because of such small numbers. So the pediatric picture is small, no pun intended.

CO-CHAIR HALL: Karen?

DR. JOYNT: Sorry. That was left there from --

CO-CHAIR HALL: Okay. I see no other comments, so let's move to a vote for validity.

MS. SHAHAB: For 2b, validity, one is high, two is moderate, three is low, and four is insufficient. And the time starts now.

(Pause.)

We have all 24 votes. For validity, one high, 16 moderate, seven voted low, and zero voted insufficient.

CO-CHAIR HALL: Any specific comments on feasibility?

(No response.)

Seeing none.

MS. SHAHAB: For criteria 3, feasibility, one is high, two is moderate, three is low, four insufficient. And the time starts now.

(Pause.)

We have all 24 votes. For feasibility, 11 voted high, nine moderate, four low, and zero insufficient.

CO-CHAIR HALL: Specific comments on usability use? Realizing that many comments have overlapped so far.

MS. SHAHAB: For criteria 4, usability and use, one is high, two is moderate, three is low, four insufficient information. And the time starts now.

(Pause.)

We have all 24 votes. Three high, 11 moderate, 10 low, and zero insufficient information. So we can also -- if there is no other comments, we can go ahead and vote on overall suitability for endorsement.

CO-CHAIR HALL: Any other comments?

(No response.)

No.

MS. SHAHAB: One is yes, two is no, and the time starts now.

(Pause.)

Okay. We have all 24 votes now. For Measure 2496, standardized readmission ratio for dialysis facilities, 13 voted yes, and 11 voted no.

CO-CHAIR HALL: So that this particular measure falls, again, into the 40/60 percent category. So it will move forward through the additional phases of this

process.

CO-CHAIR KAPLAN: Just for the record, I don't count CHF a huge success.

(Laughter.)

I just wanted to make sure everybody knew that.

CO-CHAIR HALL: Too late. Thank our colleagues from Michigan, and move on to the next measure, where we are about 20 minutes behind schedule. We are considering rejuggling the end of the schedule to try to make sure that anybody who needs to leave town and catch a flight can do so. So please bear with us as we move forward, and we will continue to try to be as timely as we possibly can with these critical discussions.

The next measure is 2503, hospitalizations per thousand Medicare fee-for-service beneficiaries, developer Colorado Foundation for Medical Care. And our first discussants will be Leslie and Tom. We'll invite the developers to introduce themselves

and briefly introduce the measure.

DR. BROCK: I'm Jane Brock from CFMC, not Teligent. We just changed our corporate identity.

MS. STEVENS: I'm Beth Stevens. I'm a statistician on the project.

DR. BROCK: So I feel like I should thank the university for -- the team from the University of Michigan for already making some of our points. So thank you.

I want to give a little background about both of our measures. So we have a hospitalizations per thousand measure, and a readmissions per thousand measure. And I just want to talk a little bit about the background of both of them together, and then that will be all the background that we need.

In 2008, CMS funded 14 QIOs to do population improvement with regard to K transitions as measured by hospital -- 30-day hospital readmission rates. So we started out working with hospitals but very quickly we

were being asked to reduce the readmission experience of a population. That population was defined by geography, so the population of fee-for-service beneficiaries within a community, and we defined those communities as a series of zip codes.

So this was based on where the Medicare beneficiaries lived, so it didn't take very long to realize that we had to work with hospitals, so this very quickly grew into community collaborative and collective action type interventions, with a wide range of providers, and then pretty quickly went beyond medical service providers to involve home and community service providers.

So it was pretty clear within the first year of the project that the measures that we had available to us at the time, which were based on readmissions per discharges, we were not capturing the improvements that were being made. So the interventions that reduced 30-day readmissions, reduced 31-day

readmissions, and 45-day readmissions, and which go straight into the denominator. So we had communities that were not capturing the benefits of, say, improving referrals to palliative care, or home community services and, even worse, were almost putting themselves at risk of failure by potentially reducing the denominator faster than the numerator.

So we went to measure just admissions per population and readmissions per population, and that is where these measures come from. So they were being used -- when we first started using them, there were 14 communities, now we are working with about 400 communities. This work is going to expand again next year. With QIOs, we will be working with a lot of rural communities. And we think there is urgency around this message, around this type of measure, because it is not just us. So with all the focus on population health, we think that there is an urgent need

for a way of measuring the readmissions attributable to the population, not defined by their relationship to an existing medical service provider.

So, the numerator is just number of admissions, and we'll talk about the hospitalization measure. Number of admissions divided by the number of fee-for-service Medicare beneficiaries in the target geographic region. So this is not risk adjusted. When we start to work with everybody in the community, it's hard to stick with a definition of avoidable or risk that is valid. So, for instance, some of the communities now are working, say, with our housing authority, so it really gets to the question that started out today around social instruments of health. So when you go to a community initiative, potentially the community can solve all of those problems. So we do not risk adjust this measure. Should I go through validity now, or --

CO-CHAIR KAPLAN: Anything relevant to the -- I think we need an overall description, and then we will go through by criterion and ask you, I think is the best way to do it.

DR. BROCK: Okay. Well, so the description is just hospitalizations per thousand beneficiaries. We calculate that per year and per quarter.

CO-CHAIR KAPLAN: Thank you. Discussion of the evidence? Leslie, and then Tom.

DR. THOMAS SMITH: Yes. Leslie and I are the reviewers, and we decided I would start and you'll jump in to help out. So I think in terms of the evidence, the story and the background are compelling. The evidence is drawn largely from the prior work on rehospitalizations, and certainly that is an important area for performance measuring and performance initiatives.

The logic is that through some of

the research we have about, for example, transitional care interventions, and also patient activation, things like that, that what happens in the community often has a more profound impact on whether people are hospitalized than what happens in the hospital.

So, therefore, we need a community-based measure of admissions. And we should take that step from just focusing on readmissions to focusing on all admissions to broaden the denominator and really create a measure that is applicable for community-based initiatives, for example, public health prevention initiatives.

So the evidence, again, is drawn upon a more limited sample of rehospitalizations, and that work -- readmissions. But the story is compelling, and I think the evidence supports moving forward.

CO-CHAIR KAPLAN: Leslie?

MS. HALL: I don't have a comment.

CO-CHAIR KAPLAN: Other comments? Are we ready to vote or -- oops. Okay.

MS. CENTENO: I just need to clarify. Are we holding the hospital accountable for this measure, or the community?

CO-CHAIR KAPLAN: Good question. Who is the accountable entity?

DR. BROCK: So the accountable entity is an initiative that seeks to improve community-based readmission rates. So it is not a provider accountability measure.

DR. THOMAS SMITH: This touches on a fundamental question, which we will get to I think in the validity discussion.

CO-CHAIR KAPLAN: Thank you. Okay. Oh, Karen.

DR. JOYNT: Just a brief comment about this. I think the developer has actually really undersold the evidence behind this. There is a lot of evidence for how

better primary care and other community-based interventions can improve health outcomes. And the background was completely focused on rehospitalizations. So I don't know that we saw the evidence, but I think this is a scenario which, because this is an outcome and not a process, that there I think is fairly good sort of theoretical basis for understanding that if we do better as a community we will reduce people's likelihood of being hospitalized.

CO-CHAIR KAPLAN: Again, we will come back to this discussion in the usability issue, too, in terms of implementation, but for now we are talking about evidence. Are we ready to vote? Any other comments? Going once, going twice.

(No response.)

Okay.

MS. SHAHAB: Voting for 1a, evidence. One is yes, and two is no. And the time begins now.

(Pause.)

Just one more vote. We have all 24 votes. For 1a, evidence, 22 voted yes, and two voted no.

CO-CHAIR KAPLAN: Performance gap? Thomas or Leslie?

DR. THOMAS SMITH: Sure. Yes, I think this is fairly straightforward, too. I think the answer is yes, there is a gap and an opportunity there. If you're looking at admissions per thousand Medicare beneficiaries, and the measure is calculated at the level of states and communities, it is also calculated annually and quarterly with seasonal correction.

So if you look at it -- at annual levels, 270 or so, around there, admissions per thousand beneficiaries, the quarterly metric is 65, 70 per thousand beneficiaries. Standard deviations are around 20 percent. So, really, significant variability and a real opportunity. Again, one of the main reasons

is the denominator is so large.

CO-CHAIR KAPLAN: Other comments? Are you ready to vote?

(No response.)

MS. SHAHAB: For 1b, performance gap, one is high, two is moderate, three is low, four insufficient. And voting is open now.

(Pause.)

We have all 24 votes. For 1b, performance gap, 19 high, three moderate, two low, and zero insufficient.

CO-CHAIR KAPLAN: Okay. High priority. Oops, sorry.

DR. THOMAS SMITH: Yes. I was going to say we can move on to priority. This also I think is straightforward. Admissions of the Medicare population are a priority. That's a substantial population. It is all beneficiaries. Why not, right?

(Laughter.)

Maybe that's not --

CO-CHAIR KAPLAN: Any comments beyond "why not?"

(Laughter.)

Leslie?

MS. HALL: I'd just like to add that this could be an opportunity for something of a baseline for us to use in other measures in the community, and so a beginning of harmonization and a really good building block. So --

CO-CHAIR KAPLAN: Why not second it?

(Laughter.)

Ready to vote?

(No response.)

MS. SHAHAB: 1c, high priority. One is high, two moderate, three low, four insufficient. And voting begins now.

(Pause.)

Two more votes.

(Pause.)

We have all 24 votes. For 1c,

high priority, 20 high, three moderate, one low, and zero insufficient.

CO-CHAIR KAPLAN: Moving on to reliability and validity. Thomas?

DR. THOMAS SMITH: Again, good. I think they presented data from Medicare claims for five or six years, I believe, so a total of 40 million beneficiaries' data were presented. They did use a split sample and their reliability reproducibility was very high. Weighted cap is where -- .8, .9, very, very high. So that was it. I mean, those were your reliability data right there, so I think that's pretty straightforward.

CO-CHAIR KAPLAN: Any comments? Leslie, anything?

(No response.)

Negative. Okay. Ready to vote?

(No response.)

MS. SHAHAB: For 2a, reliability, one is --

CO-CHAIR KAPLAN: Oops. Who said

-- Wes. Sorry.

DR. FIELDS: I just want to send my compliments to the measure developer. This is my favorite measure of both days. And since we are pretty good at mixing science and empiricism, I just want to point out that after a premium per-member, per-month for any prepaid group contracting with Medicare Advantage contractor, this is the single most important metric for the single largest cost center.

And find it, really, you know, entertaining and amusing, that it is coming at us looking like a community initiative. What it really is is your best single benchmark between the difference between the fee-for-service program and the Medicare Advantage program.

So I'm speaking in favor of its reliability as a metric.

CO-CHAIR KAPLAN: Thank you. Other comments?

(No response.)

Okay. Now voting.

MS. SHAHAB: For 2a, reliability, voting is open now.

(Pause.)

One more vote? We have all 24 votes now. For 2a, reliability. 18 high, six moderate, zero low, zero insufficient.

CO-CHAIR KAPLAN: Thank you. Validity?

CO-CHAIR HALL: Wes, are you still up?

CO-CHAIR KAPLAN: Oh.

DR. FIELDS: No.

CO-CHAIR KAPLAN: Sorry. Thomas?

DR. THOMAS SMITH: So I think here is -- with validity and usability is where we have to have some discussion, and there are some fundamental issues and questions that come up. They present limited validity data, face and construct validity.

They mention the Atul Gawande

paper on the hotspotters, right, the communities, the four communities around the country -- Newark, Camden, El Paso, the other two, that are known to have high rates of inpatient use. And they compared -- and they looked at those four cities using this measure, and sure enough found high rates of admissions per thousand beneficiaries, which gives some validity, some face validity.

They also mentioned a Commonwealth Fund study of potentially preventable admissions, and they compared communities -- I think it was community rates of potentially preventable admissions to rates on this measure and did I think quintile or quartile analyses. And the cap is -- were moderate.

So those are sort of decent face construct validity studies, although it does -- a question was noted that both of these other studies, whether it's Gawande's report or the Commonwealth study, also relied on Medicare claims data, which I guess ultimately

is the same data set that the stewards developed the measure off of. So there are questions there.

There is good control for seasonal variation, we should mention that, for the quarterly measure. Other than that, there is no other validity data, and there is no -- one of the bigger issues here is there is no risk adjustment. That was a conscious decision. The idea here is that this is a community-based measure that communities can use, for example, to track process and outcomes in prevention initiatives. But it does raise a question of how the data would be used and comparability across communities or comparability across providers within a community.

And I think the response -- you guys should speak to this -- I think the answer is that this is a measure that would be put out there with the proviso, or with the note, that it's only for communities to

compare to themselves and not to be used for other comparison purposes. So because of the absence of risk adjustment questions about validity, therefore, and given that -- how can NQF control how people use other measures when they're out there, it raises some important questions for discussion.

So I'll stop with that. I don't know if --

DR. BROCK: Yes. So this measure is developed for a community, a specific place to track their own progress over time. We use it in the program to compare sort of relative improvement among different communities, but comparing one community's hospitalization or rehospitalization rate per thousand to another is not what the measure is designed for.

CO-CHAIR KAPLAN: Bruce? Oops. Karen?

DR. JOYNT: Yes. I wanted to speak to the issue of risk adjustment as well. I think it's an interesting question, because

some of the -- some of the high rates of hospitalization are a marker of poor quality care. It is not just that things are, like, somehow different in some part of the country than others, and some of that is propensity to use hospital services, some of it is availability of hospital services. Some of it is just bad outpatient care.

And so I think it is -- this is a tricky space to decide what to adjust and not, and it gets a little bit, I think, probably into the discussions that were happening in the other committee about, what do you want to hide versus adjust for? And is it enough for people just to compare relative to themselves or is there some bar we should be, sort, of holding people to? So this is another place where I think the use and usability will matter.

One set of validity testing that I would have liked to see is, as opposed -- I agree the validity testing was all sort of to

itself, which is fine, but, you know, Atul chose those communities based on their high utilization. So if that didn't match, then there would have been very something very wrong.

But what we don't know is how other constructs might affect this. What about supply of services? What about rurality? I think there is a lot of other stuff. If we wanted to go down the risk adjustment path, we would need to think about more than just sickness. We'd need to think about infrastructure. And if that's not where we want to go with this measure, that's fine. But I do think that it bears considering that it's about more than just whether or not someone has diabetes.

CO-CHAIR KAPLAN: By the way, the independence of purpose of measurement makes a psychometrician insane, because the measures are only good for the purpose you are putting them to. So I have been mute on that topic,

because it comes up beyond the imprimatur of the committee in terms of how measures are actually being used. But, believe me, many of us consider the purpose of measurement an important property of how you interpret its validity and reliability, for example. So thank you, Karen.

Paula?

MS. MINTON-FOLTZ: Yes. Thank you. And I also want to talk about just the academic medical center, because those do -- I mean, you have communities that do have academic medical centers, but the community feeds into the tertiary piece of that. And those usually do have a higher case mix and would be more ill and more apt to readmit. But I -- because they are not always from that community, they may -- but it's a small proportion, I think, of the overall.

CO-CHAIR KAPLAN: Thank you.

DR. BROCK: So I just want to point out, though, that -- so if the

readmission is assigned to the place that the patient lives, so a tertiary referral center that gets a bunch of patients from someplace else, if they were readmitted, that would be assigned to the community that the patient came from.

CO-CHAIR KAPLAN: Thanks for that clarification. We are going to just come right up and go right around. So, Ron?

MR. STETTLER: Yes. So I agree with Wes. This is probably the single most important metric, as a health plan, that we use for measuring our performance. I think my problem with it is not so much that if you could control people from comparing community to community, you could get away with not doing the risk adjustment. But you have to if they are ever going to compare communities. But even within a community the population changes. And this doesn't adjust for that.

So Medicare, there is age in to Medicare. It is changing over time.

Populations are probably getting younger over time. And I honestly think you would have a miss in the metric. You would be, you know, demonstrating something that doesn't exist just because you are not controlling for that case mix over time, without doing the risk adjustment for age, for sex, for other characteristics, even within the community.

MS. STEVENS: We have looked into that some. The change over time for a relatively short period of time is pretty minimal. I think if we were comparing now to five years ago, the change maybe in age would have kind of changed within a community. But within a community over the life of the quality improvement project, the demographics and risk doesn't change very much.

MR. STETTLER: Then I guess my point would be, why not control for it? I mean, I think you are just missing out on an opportunity to eliminate the controversy, and just adjust for it, and be done with it.

CO-CHAIR KAPLAN: Thank you. Paul?

MR. HEIDENREICH: Yes. I have a similar concern with not including the risk adjustment. Without it, you are basically saying, well, all hospitals can improve -- I mean, all communities can improve. They all can improve the same amount, without some standard.

And like the Colorado group -- community I think from that article, they were I think the best, or they are very good. So when they -- next year, when they are not any better, do we say, "Wow, you guys have failed," while some of these other places improved? If we are truly not comparing -- I think in the end we are comparing. I think people are going to compare. So I would urge you to add that.

DR. BROCK: So we found -- certainly, in the initial 14 communities, we found that the capability to improve was not

a reflection of where people started, if you looked at proportion improvement or reduction in failure rate, so we had some very high-performing communities that improved the same amount as some very low-performing communities. So --

CO-CHAIR KAPLAN: Pam?

DR. ROBERTS: Just in addition to looking at the -- adjusting for case mix over time, also looking for hospital services, change over time, too, and I think that that's an important piece that would need to be looked at for adjustment.

CO-CHAIR KAPLAN: Thank you. Bruce?

CO-CHAIR HALL: Well, I had some concerns, many of which have been expressed. It is not clear to me in reading the denominator spec how we're defining the state or community. So, obviously, somebody later on is -- I mean, state, obvious. But somebody later on is going to define the community, and

then -- and I can see it now, we have already probably seen it already, that, you know, this is just going to be front page USA Today, every single community flagged with their rate.

So I think the notion that we're not going to be comparing across facilities is a false one. I also think the attribution of the patients' care to a home zip code, perhaps different than where they're actually treated, to me is a problem, because then you would be arguing not to have that community feed into the treatment center that might be in a different zip code.

I think controlling across communities, controlling across time, are critical aspects. And how do we think of -- it's not clear to me how we think a community makes use of this information, but maybe that's my feeling.

A question I have, though, is what are we really approving? It seems like CMS

could and probably already has published this in USA Today already. What are we actually approving? We are approving calculation of discharge per thousand fee-for-service Medicare without risk adjustment, without some of the other issues. I'm not even sure how much this is really a measure, versus a simpler statistic. And it's not clear to me why we need to approve this.

DR. BROCK: So, let's see, there are several questions in there. So I do think it's worth standardizing approach to defining readmission and admission and measuring that, and assigning that to place. So I don't understand your concern about the role of the place there treated, in that the place that they live generates the admission, and the place they are discharged to, in a sense, generates the readmission.

So it has a lot more overlay with the social determines of health, which is part of why they are useful. So if a community

knows, yes, everything is terrible in this neighborhood, and that's why when you go to this zip code or -- we are increasingly using census tracks, you know, all of us are taking a hit on quality measures, because we need to solve some of the situations that are involved with living in this place.

And that's why, as part of a wider community initiative, it is so useful. And that is also part of why we don't risk adjust it, because you don't want to set parameters around what a community can and can't solve. You know, so if your housing authority provides good housing to people after discharge, then you can stop wringing your hands about, oh, gosh, the problem is housing.

So that's a lot of why we wanted to just display it as a rate that could be reproduced across places, not necessarily to be compared across places but to know that we can calculate it the same everywhere and track it the same everywhere. Does that answer your

question?

CO-CHAIR HALL: I think it gets to some of the aspects, yeah. I think, again, I'm not sure why we need to approve a simpler calculation, but I do still feel that comparing a community to itself over time and -- is important, and inevitably it's going to appear front page newspaper, and communities will be compared with each other.

DR. BROCK: So, in the end, the answer to that is kind of easy, I think. So if you look at, you know, South Chicago, they can say, of course we have a readmission problem. We have terrible, you know, poverty and housing, and, you know, that's probably all true. But what the community needs to think about is what can we do about that, you know, not like wow, things would be better if we were Grand Junction. You know, what we want them to think about is what should we do now, given this situation.

In terms of defining the actual

measure, it is more important for readmissions than for admissions, I think, so -- I know we are not discussing that yet, but there is a number of ways to define what is a readmission and how to calculate that numerator, you know, number, which is -- we think it's important to standardize that.

CO-CHAIR KAPLAN: Jane, could -- it's Jane, right? Could you just clarify for me because I think maybe I misunderstood the answer to Bruce's question based on what you said before. The community the patient lives in is the community that gets attributed the discharge, correct?

DR. BROCK: Yes.

CO-CHAIR KAPLAN: Okay.

DR. BROCK: So if they are discharged from a Denver hospital and they go to Miami and get readmitted in Miami, that goes on their Denver zip code.

CO-CHAIR KAPLAN: Thank you. Larry?

DR. GLANCE: So I agree that there is a tremendous amount of useful information that could be obtained from this particular measure. I am also concerned that there is the potential for some very real, unintended consequences. Namely, I could see a Dartmouth Atlas type of publication where you could very easily look at a map and using color coding, identify areas that, as a Fortune 500 company CEO, I might not want to have my company.

Maybe as an upwardly mobile yuppie or something of that type, I would look at that map and say, I don't want to live there. Now, you might say that some of those things are pretty obvious without that map, but I would say maybe not all of them. And so people might start moving around, companies might start thinking about where they're going to locate their headquarters based on this kind of map.

So although I think there is a lot of utility for -- within community comparisons

for longitudinal-type comparisons, I think that there is some risk with -- between community comparisons with this kind of measure.

CO-CHAIR KAPLAN: Thank you. Now remember, we are still on the topic of validity, not usability yet, although purpose of measurement shapes validity, so it's hard for me also to kind of make that call.

But, Leslie? Thank you.

MS. HALL: And I would encourage that kind of difference that you describe, Larry, because -- and when we do community needs assessments, in many communities reporting back to our governors or our city mayors about what our needs are there, it gives providers and others an opportunity to demonstrate and identify infrastructure issues, access issues, disparity issues, rather than just having single points of failure aimed at hospitals.

So I think it's an important

equalizer, and transparency is a first step in quality improvement. So I think this is an important effort and quite valid, even though the use cases that we might think of are typically outside of the hospital.

CO-CHAIR KAPLAN: Thank you. This has been a very rich discussion. Karen, did you have something to add?

MS. PACE: Yes. I just wanted to make a comment about the use of the measure. I think one of the ways, at least currently, that NQF is thinking about that is that the way the measure is specified puts some parameters around use in that it should identify the patient population, the setting or the accountable entity, the level of analysis, and that should be the use that is represented in the measure testing as well.

Now, the other overlay is that NQF endorses measures that are expected to be used in accountability applications, but at least currently, we don't make a distinction of

whether that's pay-for-performance or public reporting. But I think there is some element of the use case in terms of how the measure is specified, and just wanted to offer that.

DR. BURSTIN: Let me follow up on that. So the measures we saw earlier in the day, the prevention of quality indicators from ARC I think are very analogous. They are also at the community level. Those are endorsed. But they are actually, just to -- you know, for full transparency, they are adjusted for age and gender specifically because of concerns about differences across communities.

Those are displayed across communities and are very much thought to reflect the quality potentially of ambulatory care available to avoid those hospitalizations. So just to put that there.

CO-CHAIR KAPLAN: I think this -- clarification of this use and utility business gets a little bit confusing for people when you start talking about unintended

consequences. And so people are reacting to, well, how could this go wrong? Any time you measure something and put it out there, people can use it for something you didn't plan. So I think that language gets us a little bit confused about, you know, exactly how this is going to be used.

Also, with respect to populations, it gets -- people like me go, population? What are you doing with statistics? That's a population. You just declare it. You know, come on. So it's a little confusing and it brings in a few -- a little bit more discussion.

We have had some rich discussion now about validity, and I think we are -- unless there's any other issues, I know Karen and Wes wanted to bring something up. But unless we -- can we vote?

(No audible response.)

Good.

MS. SHAHAB: So we're going to

vote on 2b, validity. One is high, two moderate, three low, four insufficient. And time begins now.

We have all 24 votes. For 2b, validity, two high, 13 moderate, eight low, one insufficient.

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

DR. THOMAS SMITH: I don't think there are significant concerns here. I don't have anything to say.

CO-CHAIR KAPLAN: Wes?

DR. FIELDS: Yes. I have been flipping that thing open now for about five minutes. This is a situation, I believe, where a really good idea coming from a community trying to do the right thing has potentially allowed us to face one of the most important longer-term issues for the Medicare program.

And if for no other reason than that, and if for no other reason than there's

thousands of communities around this country, you do need to risk adjust this data. Without the risk adjustment, it probably would be tougher for me to vote in favor of it because although I very clearly understand their intent, and it's a valid intent, the larger issues that confront the program and the trust funds are well-understood.

And you would be doing not just communities but Congress a huge favor if you do this with risk adjustment so that it could become more and more clear whether or not there should be a differential in payment per member per month between Medicare Advantage and the traditional program or not.

It would also probably help us if -- in a drilldown, to understand where are the gaps in case management and care coordination where we can make investments that really will make a difference in outcome by looking at what Advantage can do versus the traditional program cannot.

So speaking very strongly in favor of doing the risk adjustment and the validity, I just wanted to point out, is I don't think of us believe that CMS pays exactly the same amount per member per month in every single geo zip in this country. It is already risk-adjusted.

You know, what they pay plans is already a matter of a whole other body of work that goes on all the time. It would be I think a minor tragedy if you didn't use technology and analysis that already exists and apply it to this population. For me, ultimately, even though this is community-driven, I think all of us feel driven to make sure this population can be served well and that the services that we provide are sustainable. And you won't get there without risk adjustment when you look at this measure.

CO-CHAIR KAPLAN: Thank you. You know, I am hearing things that speak more to usability and use and a little bit less to

feasibility because the feasibility I think answers the question, can it be done? And then usability is, how should it be done and can it be used in the form proposed?

But are there any other comments on feasibility? Go ahead.

DR. BULGER: Yes. And it's probably used, but it speaks to what -- this data is already on the Dartmouth Atlas for, you know, it's hospital discharges per thousand Medicare enrollees by gender and type of admission and by hospital referral region, state, and hospital area. So I mean, you know, it's already to some extent there.

Now, it's different than this measure is, but, you know, this type of data is already available, and it's not -- it's risk-adjusted, so it's risk-adjusted by age, sex, and race, but that's it. And I'm just playing on the Dartmouth Atlas right now, pulling it up, so I mean, you can -- and that's exactly what will happen with this. It

will be out there just the same way.

CO-CHAIR KAPLAN: In terms of feasibility, though, because it's already out there, it is doable, right?

(Laughter.)

CO-CHAIR KAPLAN: So it is -- so whether it should be is a whole different story. But for feasibility we need to kind of vote how feasible it is unless there is another issue. Hearing none?

MS. SHAHAB: For criteria number 3, feasibility, one is high, two moderate, three low, four insufficient. And the time begins now.

For feasibility, 22 high and one moderate, zero low, zero insufficient.

CO-CHAIR KAPLAN: Okay. With respect to usability and use, we've already had some discussions about the implications of risk adjustment and so on. Again, you know, that -- that may go on to address the issues that are going to be moved to the MAP group.

But if there is anybody else who wants to make a comment on usability and use? Go ahead, Leslie.

MS. HALL: I would just advocate that with -- if we're thinking about using this to compare for payment and other types of health care-related issues by thinking that the risk adjustment addresses that better, perhaps. But if we are thinking that this might be used to say, do we have deserts of food? Do we have good transportation infrastructure? Do we have other root causes that are not related to this? Those are not risk adjusted.

So I think that it is an apples to apples consideration, and this gives us a chance to look at other causes of admissions in hospitals and an opportunity to work with community partners to advance and improve health care in our community for unrelated issues that -- to our hospitalization and care that also provides other opportunities for

funding sources and community outreach.

So I would advocate that non-risk adjustment gives us better opportunity for usability.

CO-CHAIR KAPLAN: Thank you very much. In terms of just use and usability, I am hearing a fair amount of concern about risk adjustment. And something -- I mean, Karen brought up the issue of rural-urban and limited accessibility, for example. And Leslie brought up the issue of needs assessment. And to the extent that it -- maybe a more helpful/useful measure might be adjusted for need, how many -- you know, how much are we accommodating the need, et cetera.

All interesting and useful points. We are voting on the measure we got under the microscope right now.

DR. BROCK: Can I comment?

CO-CHAIR KAPLAN: Sure.

DR. BROCK: So our measure is intended to be used as you have described. So

the problem with the Dartmouth Atlas HRR work -- and it's excellent work; we don't have a problem with the Dartmouth Atlas. But the hospital referral region really reflects the case patterns of medical service providers.

And what we find consistently in working with readmissions particularly, which is just a subset of admissions, it doesn't reflect 100 percent what medical service providers do. I mean, I'm beginning to wonder, what is the impact of medical service providers at times?

So to me, the measure is -- it's an alternative way to look at what are the roots of unnecessary admissions and readmissions? And those are often rooted in community realities that we would like to not adjust away.

So, for instance, if a rural community knows, well gosh, yeah, we have a problem because we're like rural, well, but our risk adjusted measure is okay. I mean,

that doesn't help them track whether they're doing the right things to improve the outcomes for the folks that live there.

CO-CHAIR KAPLAN: Thank you.

So again, for the purposes declared in the measure as specified currently, we are voting on -- unless there are any other comments?

(No audible response.)

CO-CHAIR KAPLAN: We are voting on usability and use.

MS. SHAHAB: For voting on usability and use, one is high, two moderate, three low, four insufficient information. And time begins now.

One more vote. We have all 24 votes. For usability and use, five high, seven moderate, 12 low, and zero insufficient information.

CO-CHAIR KAPLAN: So now the overall suitability for endorsement. Comments? None? Yes? No?

DR. THOMAS SMITH: No. I think -- I don't have any.

CO-CHAIR KAPLAN: Good.

(Laughter.)

CO-CHAIR KAPLAN: Any other discussion? Karen.

DR. JOYNT: I will just say I think that this metric does not fit the same paradigm as a lot of the other ones that maybe we are more used to looking at. I don't think that's necessarily bad. It's being asked to do something different. It's being asked to put our focus on where people live and not where they are hospitalized. And national data would suggest, even with the Dartmouth Atlas stuff, only about half of the people are really hospitalized in their HSA. So it's not really clear what that tells you about someone's home.

And I totally get all the concerns about risk adjustment, and everyone is certainly entitled to their opinion about each

piece of that. But I don't think this should be about -- this isn't about a hospital, and it is fundamentally supposed to be different because it's about a community. So I would just sort of throw that out there for consideration.

CO-CHAIR KAPLAN: Thank you. Very helpful. Other comments?

(No audible response.)

CO-CHAIR KAPLAN: Excellent. We vote a score.

MS. SHAHAB: For overall suitability for endorsement, one is yes, two is no. And time begins now.

We have all 24 votes. For overall suitability for endorsement, 12 yes, 12 no, for Measure 2503, hospitalizations per thousand Medicare fee-for-service beneficiaries.

MR. AMIN: Again, this measure falls in our gray zone, very clearly. And so the measure will continue to move forward in

the process, and we look forward to getting public comments on this measure. There is a lot of gray here.

MS. KHAN: We have three measures so far where consensus hasn't been reached. That's 2512, the LTACH readmission measure; the dialysis facilities readmission measure; and this one.

MS. SHIPPY: So I think with the split-vote measures, I'm seeing a little bit of a trend, but I would love to hear NQF report back whether the use and usability feels like that's the closest, and that that's leading into the kind of gray zone. Are you able to report back just on those three that are in the gray zone what the use and usability votes were?

CO-CHAIR HALL: You're sort of -- I don't mean to put words in your mouth. You're saying that you think the use and usability and the overall votes are mirroring each other?

MS. SHIPPY: They seem like it.

CO-CHAIR HALL: So is that a problem?

MR. AMIN: I mean, the way that we sort of think about this is the criteria are hierarchical in the fact that importance measure is the most important, scientific acceptability. I don't -- I mean, the votes aren't always this sensitive to predict the outcome, but it sounds like validity and the concern around unintended consequences seem to be what's described in the votes in the majority of these, although I don't want to hypothesize because I think every one of the measures is a bit different.

CO-CHAIR HALL: It does appear that. And looking back, I can -- the usability votes are, in terms of going one way or the other, are reflecting which ones are not having consensus reached.

MS. SHIPPY: You're saying the usability or the feasibility?

CO-CHAIR HALL: Usability.

MS. SHIPPY: Usability.

CO-CHAIR HALL: And, again, just a question for NQF colleagues. We will not be asked to comment on the public comment, right? Our discussions and the public comments together, separately but together, go forward the next --

Dr. BURSTIN: So what will go forward is will this measure out, like the couple of others we have had like this that say, consensus not reached. We'll lay out the discussion, the pros and the cons.

You'll then get public comment. You'll have an opportunity to reassess it. We have now built in any gray zone measures automatically. The committee can reassess post-comment. Again, just an opportunity to gather more input.

And I have to say, in general, over the last few months of steering committees, it has generally been validity

that has tended to be more of a split. That leads to the -- to the split at the end, but also certainly I think unintended consequences is another one in particular, which is part of usability, I think is what you're seeing.

CO-CHAIR HALL: Challenges. So we remain somewhat behind and we will ask CFMC to proceed with introducing the next measure.

DR. BROCK: So the next measure is readmissions per thousand, and it's -- you know the background. So in this one, the numerator is the number of readmissions within 30 days of a previous discharge from acute care hospital divided by the number of people in the geographic region of interest pro-rated by day for their participation in the Medicare program. That most often, in fee-for-service, means censoring for death. So it's an incidence rate. It's occurrence per person-day.

So we developed this measure, like I said before, you know, hospitalization --

rehospitalizations are prevalent and costly. It is very difficult to track community improvement if you are only assessing the denominator of recent discharges because that, in fact, varies a great deal by community capacity to not hospitalize people. So with regard to tracking change, a community can lose on the readmissions per discharge measure by reducing their hospital admissions faster than they reduce their readmissions. In terms of the validity and reliability, we have already described that. It's the same process. We did a split sample reliability, and it was very high. The caps were .8, .9.

We compared to the Commonwealth study of avoidable hospitalizations to compare the quintiles -- I think it was quintiles, and the caps on that were high -- .7.

So we note that we have proposed this measure as both a quarterly and an annual readmissions per thousand. They are also very highly correlated, like .99 or something like

that between quarterly and annual.

CO-CHAIR HALL: Thank you. We will invite Paul and Karen for initial comments. We are back into the category of evidence.

DR. JOYNT: So a lot of this is very similar to the one right before this, so I will be brief. Again, there is not a ton of evidence other than the sort of major study which I think probably drove the development of all these measures. But I think there is a good face validity that much of the promising work around readmissions reduction is taking place out of the hospital, and this is a good way to acknowledge that.

CO-CHAIR HALL: Anyone else, comments on evidence?

(No audible response.)

CO-CHAIR HALL: Seeing none, we will vote on evidence.

MS. SHAHAB: For 1a, evidence, one is yes, two no. And the time starts now.

For 1a, evidence, 21 yes, one no.

CO-CHAIR HALL: Any specific comments on opportunity or performance gap?

(No audible response.)

CO-CHAIR HALL: Not seeing any.

MS. SHAHAB: 1b, performance gap, one high, two moderate, three low, four insufficient. Time starts now.

We have 23 votes. For 1b, performance gap, 17 high, four moderate, two low, zero insufficient.

CO-CHAIR HALL: And priority, specific comments?

(No audible response.)

MS. SHAHAB: 1c, high priority, one high, two moderate, three low, four insufficient. Time begins now.

We have 23 votes. 1c, high priority, 15 high, six moderate, two low, and zero insufficient.

CO-CHAIR HALL: And back to science, reliability and validity? Paul or

Karen, any specific concerns that we didn't touch on last time?

DR. JOYNT: I think the reliability of this one is, again, excellent. But the numbers are huge. So there is really fairly good correlation. I don't think we're measuring things wrong.

CO-CHAIR HALL: Any other comments?

(No audible response.)

CO-CHAIR HALL: Not seeing any.

MS. SHAHAB: Voting is open now for 2a, reliability. One is high, two moderate, three low, and four insufficient.

Two more votes.

We have all 24 votes. For 2a, reliability, 18 high, six moderate, zero low, zero insufficient.

CO-CHAIR HALL: So we're back to validity.

DR. JOYNT: I'll just add one additional thing, which is an interesting

thing about this metric is that the racial and ethnic disparities are particularly high in this measure, which I think is in part because it's not risk-adjusted. It actually shows you the real differences in health outcomes, particularly when you think -- when you realize that the actual prevalence of disease is one big driver of disparities. That sort of leaves in all the disparity, which I found to be a useful thing here.

The other thing that I thought was particularly useful about this metric, which was a fact that was brought up by the developers, which is that it won't -- it will fall if admission rate falls. So if communities -- if we're sticking with the community improvement, it will go in the same direction as admissions if a community intervention is done that improves outpatient care. I think all the other issues around sort of risk adjustment and all that are probably as -- I don't have anything

additional to propose.

CO-CHAIR HALL: Paul?

DR. HEIDENREICH: I think a new issue here is, at least in prior studies I have seen like this, that admission rate drove this measure much more than readmission rate. And I'm not sure if that's -- I'm not sure what the correlation is, but it was incredibly high. And that hospitals would change their admission rate, and because of that there would be some percentage of readmissions; therefore, they would have a change in readmission rate. And I think this study was trying to show that, oh look, readmission rates have dropped with some community intervention. Actually, what they really dropped were admission rates.

So as long as you have them both together, I think they're useful. If you only approve this one, I think it could be very -- I'm not sure it has the validity.

CO-CHAIR HALL: Do the developers

want to respond to that?

DR. BROCK: Yes. So if I can comment, so we -- we actually published this, the results of the 14 community pilots, which defined and used these metrics. So, you know, we found them to correlate almost exactly, but we felt it was the other way around.

So you know, if you think about it, if you go into a hospital where 20 percent of the people in their beds are 30-day readmissions, and you take those people out of the admissions pool for 30 days from now, then they have dropped their admission rate at the same -- you know, they do very much go together at first.

We subsequently have seen admission rates slow down in their decline, while readmission rates continue to decline, which we assume has part to do with influx of new beneficiaries. But we haven't done full testing of that.

So anyway, we did see that

relationship but assumed it was the other way around.

CO-CHAIR HALL: Any other concerns or questions, or does anyone feel the need for clarification around anything we mentioned on the last measure? Tony.

DR. GRIGONIS: Yes. Do you adjust -- or do you have an idea of how many beneficiaries readmit within that 30 days, the same beneficiary? And is that an issue that could influence the result, in addition to the fact that since you're not looking at planned or unplanned, there may be a lot of situations where they have to come back?

DR. BROCK: We did not look at planned or unplanned. So we -- just to be clear, we count every admission within 30 days of a discharge as a readmission event that goes into the numerator as opposed to, you know, what's on the hospital compare measures where they require a 30-day interval free of readmissions to count the next admission as an

index admission that can take a readmission.

So in terms of how many of them were very frequent, like being readmitted within 10 days, 14 times in a row, we did look at that. We did look at that, but in a way the issue is to interrupt that cycle of inappropriate acute care support and see that reflected in the numbers.

In terms of taking out planned readmissions, we did not do that.

CO-CHAIR HALL: Sherrie?

CO-CHAIR KAPLAN: I just wanted a quick clarification from the statistician. So on the readmission side, there is not independence of measurements for the admissions. You have to be admitted to be readmitted.

MS. STEVENS: Correct. So if we had -- if a person had five admissions in a month, in the admission measure that we have already talked about they would have five in the numerator, and four of those would count

as readmissions. So they would have four for the readmission in the numerator. I think that's what you're asking.

CO-CHAIR KAPLAN: Yes. With respect to modeling, you would not -- you know, that's a non-independence issue. You cannot put -- you know, those two measures are not -- on the readmission side, it's not possible to be readmitted without having been admitted.

CO-CHAIR HALL: That's true. But I think -- it's true, but we're saying that that's a favorable aspect/attribute of this measure is that you can reduce your readmission by reducing your admission, as well as by reducing your readmissions. So it's true, they're not independent in any way, shape, or form, and it could be viewed as a favorable attribute.

MS. STEVENS: Right. And a readmission counts as an admission, and then we look 30 days out from that. So that is a

little bit different than the hospital compare measures.

CO-CHAIR HALL: Absolutely. Other comments?

(No audible response.)

CO-CHAIR HALL: Okay. Move to voting?

MS. SHAHAB: For 2b, validity, one is high, two moderate, three low, four insufficient. And time begins now.

We have all 24 votes. For 2b, validity, four high, 12 moderate, seven low, and one insufficient.

CO-CHAIR HALL: Comments on feasibility?

(No audible response.)

CO-CHAIR HALL: Seeing no new comments on feasibility.

MS. SHAHAB: Voting for criteria 3, feasibility, one is high, two moderate, three low, four insufficient. Time starts now.

One more vote.

We have all 24 votes for feasibility, 20 high, two moderate, two low, and zero insufficient.

CO-CHAIR HALL: And usability. And Kathy, do you want to open up?

DR. AUGER: So I think even putting aside the risk adjustments, which I think we have already talked about quite a bit, I think the one part of the readmission metric that bothers me a bit here is the fact that the planned aren't excluded; therefore, one way to potentially game the system is to actually delay, like what would otherwise be necessary care. And so just by putting it beyond 30 days, then you would not have that readmission count against your community.

CO-CHAIR HALL: Other concerns? Paula?

MS. MINTON-FOLTZ: I just want to point out that the -- I think it was 2011, we were here the last time for all-cause admission. The health of the community was a

big topic there. You know, how much is the community responsible versus the hospital? So I think this is a way of really getting at some of that.

CO-CHAIR HALL: Thank you. John?

DR. BULGER: Yes. The only comment I would make is I think it's a good measure, and I think it -- it measures a completely different thing that we are already measuring. But I think if it gets out there it is going to be extremely important, that that's highlighted very well, because I think there is already confusion amongst providers and the public about what the readmission numbers mean.

And I think, you know, this is going to have -- be a completely different number from anything we have, and it will be real easy for people to say -- to throw this number out and try to compare it to this -- the possible compare numbers that you talk about and create more confusion.

So I think, you know, its usability is completely different from anything we have. And making sure that everybody understands that very well is going to be extremely important.

DR. BROCK: So I agree with that. And I know I have spent some time on the CDC's website, and a lot of the statistics they put out. And they always have a link how to use this -- how to use these measures. And I think it should be accompanied by that kind of a thing.

Also, I just -- these numbers are out, and they're mapped to zip codes and they're on our website. They've been being updated quarterly for four or five years, and we haven't had -- we haven't seen people hurt themselves with them yet. So, you know, improvement initiatives use them.

CO-CHAIR HALL: You aren't aware of any people hurting themselves.

(Laughter.)

DR. BROCK: They didn't call me.

CO-CHAIR HALL: Any other -- any other comments on usability?

(No audible response.)

CO-CHAIR HALL: Seeing none.

MS. SHAHAB: For criteria number 4, usability and use, one high, two moderate, three low, and four insufficient information. Voting is open now.

We have all 24 votes. For usability and use, four high, 11 moderate, nine low, and zero insufficient information.

CO-CHAIR HALL: Before the overall vote, any additional comments?

(No audible response.)

CO-CHAIR HALL: It's curious to me that we are -- we are reaching a different conclusion than we did last measure, but that's just a curiosity. There were -- risk adjustment considerations still apply, but any other comments before overall? Wes?

DR. FIELDS: It is really just a

question. It's not clear to me -- if NQF approves this measure on readmission, would the effect be for it to be done in Colorado or in all 50 states and Puerto Rico, et cetera?

CO-CHAIR HALL: So again, what bodies would choose to implement a program based on this? It's not --

DR. FIELDS: Well, I guess maybe, what is the intent of the developers?

DR. BROCK: It's already mapped for every zip code in the U.S. and Puerto Rico and the Virgin Islands.

CO-CHAIR HALL: Does that answer your question?

DR. BROCK: Well, okay. So Puerto Rico and the Virgin Islands are interested in each other, just because they perceive themselves as similar markets. So people do use it to locate who they think has similar problems.

I know in Colorado, our resort communities always want to talk to each other

because they have a lot of temporary residents and substance abuse issues and things like that. So, I mean, people do use it to find where they think are similar places.

CO-CHAIR HALL: Okay. Any other comments before our vote on overall?

(No audible response.)

CO-CHAIR HALL: Seeing no cards raised.

MS. SHAHAB: Voting for overall suitability for endorsement, one is yes, two is no. And time is open.

We have all 24 votes. For overall suitability for endorsement for Measure 2504, 30-day rehospitalizations per 1,000 Medicare fee-for-service beneficiaries, 14 yes, 10 no.

MS. KHAN: So just a process check. This is also in the gray zone. So it will be going out for public/member comment.

CO-CHAIR HALL: Okay. We will break. Thank you very much, CFMC.

And what are we scheduled for, 15

minutes? So, let's say 4:05. We will cut it short to about 12 minutes. 4:05, please.

(Whereupon, the proceedings in the foregoing matter went off the record 3:52 p.m. and went back on the record at 4:04 p.m.)

MS. KHAN: Let's go onto the next measure.

MR. AMIN: We can send out an email with that information as well, just so you have it.

CO-CHAIR HALL: We're going to move to -- slightly out of order -- to Measure 327, Risk Adjusted Average Length of Inpatient Hospital Stay. The developer is Premier. In order to accommodate their schedule, we're going to move them up in order. And I've notified our primary discussants of that change. So our colleague from Premier, Gene, do you want to introduce yourself and briefly introduce your measure.

DR. KROCH: I think it's on now, right? Yes, I'm Gene Kroch. I'm at Premier.

I'm chief scientist and do a lot of different things for Premier in regard to health services research.

I have a secondary appointment at the Wharton school and the University of Pennsylvania. So Philadelphia is my hometown. You can't hear me? That's not usually my problem. Okay, let me know if I have to repeat myself. Okay, sorry about that.

I'm chief scientist at Premier, Inc. Which is located, by the way, in Charlotte, North Carolina. But I live in Philadelphia. I'm health services research. I do a lot of measurement work.

And I'm affiliated with the Leonard Davis Institute of Health Economics. And I have an adjunct over at the Wharton School in Healthcare Systems. So I don't know, there's not much for me to say about this measure, really, because it's kind of been out there for so long. I've been working with this measure personally for over 20 years. But

it's not been something that's been out in the public sector very much.

It was endorsed by NQF back in -- I thought it was 2006, but the notes I saw said 2008. But it, as you probably know, measure stewardship involves updating on a regular basis, the measure in case there are any changes that are relevant to it's properties. And we were doing that on a regular basis.

Risk adjustment, as it says up on the screen, it's risk adjustment, average length of stay using a geometric mean. Given the distribution of length of stay and the properties of length of stay, as you can imagine, it has a very large positive skew. And so using geometric mean allowed us to use the general linear model with a log link, which improved the fit and reliability of the measure. The main thing I want to emphasize is that it is designed to allow a lot of flexibility and sub-setting.

Although it's set up here that it's about measuring at the hospital level, which it certainly does. Our users, that's to say the 850 or so hospitals that subscribe to Premier System, for this and other reasons, they generally are not all that interested in the overall length of stay, because they're doing this for performance improvement. And they're looking for where are they -- where are they insufficient or where do they need to focus more, in terms of making sure that the processes are followed in an efficient way.

So it's designed to allow flexible sub-setting so you can look at it at the diagnosis code level. It's a model that's stratified by diagnosis code. At the procedure code level, at the level of the attending physician's specialty, at the age category, at the DRG, look at hospital lists relative to other attending physicians. It has that attribute.

It is a measure that is very widely

tracked by again, the folks that I know in the Premier Alliance, they're looking at it all the time. In fact, they are not all that keen on the value of whatever costs measures out there.

So they much prefer to look at themselves and benchmark themselves using length of stay. And indeed, length of stay is, as you might expect, highly correlated with patient-hospital costs. The model itself fits very, very well. We have a generalized model that we use for other outcomes, like mortality and morbidity and complications, that doesn't nearly get to this level of power. It explains about 60 percent of the variation based on exogenous patient factors.

It's -- as you might imagine, we found a high correlation between this measure and various different measures of cost of care that we looked at. Procedurally based cost of care, cost accounting based cost of care, using HCRIS reports for the cost to charge

ratio to turn charges into costs.

They all are sort of different variations and have their own peculiarities. And as you might expect, as you drill down they look different, but overall, they're very highly correlated with, as you might expect, length of stay. The other thing that I think this measure is, is it's obviously very easy to understand. I mean how long -- and of course we define a hospital stay based on whether you were admitted or not.

So the shortest possible hospital stay is a day. You can't have a fraction of a day. And -- but it's sort of, kind of a no brainer as to how you would compare your numbers to somebody else's numbers.

We've found that in our validation studies, that when we wanted to explain variations in length of stay, that among the factors that were driving variations in length of stay, we found comorbidities certainly. But complications importantly.

And right now we just put out a paper about the financial consequences of inpatient complications, making use of the information on whether the or not the conditions present at admission are not. Performance gaps, I guess that will get talked about by the committee. But there are clearly effects. And not all of them go in the same -- in the direction you might expect. But there's an income effect, there's a age effect. There's a race effect. And a whole bunch of others that I guess, I don't have it written down in front of me right now.

The -- we did a study quite a while ago. As I said, this measure's over 20 years old. We did a study for the Commonwealth fund in which we found an inexorable decline in risk adjusted length of stay over the study period. Now that ended in the mid-2000s. That trend has begun to look very convoluted, in the sense that it started to flatten out. It actually went the other direction. And now

it's going down again.

So length of stay seems to reflect some of the forces that are going on. And most people sort of say well, what's driving length of stay more or less by hospitals is they're trying to improve their bottom line effectively.

Especially in A DRG environment where it's not really the payer that's benefitting from shorter length of stay, but the provider that does. So let me -- I think I -- oh, the only thing else I should say about it is, this things been around for so long. We have applied to almost every data set you can imagine.

The hospital alliance of Premier, which is now 850 hospitals, MedPar data, the National Inpatient sample that's put out by AHRQ through HCUP. It's -- you know the point I wanted to make about that is just that it's a model or a measure that's easily applicable to almost any kind of data you might have.

Whether or not it's the HR data. Or whether or not it's administrative data. Or however you get the data.

And the other point I would make too is that -- this will be my last point. That we do use the sufficient statistics from the calibration of the model to establish confidence intervals around length of stay, so you can test hypothesis about whether your length of stay's indeed higher or lower than you would expect. The -- I'm forgetting what I was going to say.

CO-CHAIR HALL: Okay, we will probably invite you to respond to some additional comments.

DR. KROCH: Yeah, let me end there.

CO-CHAIR HALL: So no worries.

DR. KROCH: Thank you.

CO-CHAIR HALL: No worries, thank you Gene. We'll turn to Allison and Ron for initial comments. We're on the category of evidence.

MS. SHIPPY: So Ron and I are going to kind of ping pong back and forth. I'll start with evidence. So I think I want to start with a question related to the -- how the forms will filled out. And there seemed to be some inconsistencies as to whether it is defined as an outcome. And there are some times in the forms that it's defined, or selected, as an intermediate outcome. So I would like to first start with Eric, because I think that help take us through the algorithm.

DR. KROCH: Right. And what I didn't mention is that the origin of this measure actually was something called the Corporate Hospital Rating Project, which was conducted at the Wharton School, between about 1988 and 1992 or 93. And the principal investigator on that study was Mark Pauly, who basically said this isn't an outcome, it's an input. And to economists, that's what we think of.

When I was filling out the renewal information for this measure, I tried to slip that one past and I was caught. No, no, no, no. You're not an efficiency measure. You're an outcome measure. That's how we have you listed. So you have to respond to that as you seen it. I think that our -- as I implied in my opening comments, our members used this as an efficiency measure. They don't really use it as an outcome measure. And in fact, when we do profiling, it's a separate dimension. It falls into cost dimension as opposed to the quality dimension. But --

CO-CHAIR HALL: Alison, did you get your questions answered?

MS. SHIPPY: Sure, yeah. So I think it does leave me a little perplexed in that the use -- the evidence base that you're noting within your membership, that has kind of driven why you're using this measure, if they themselves are defining it as an efficiency measure.

I think that is a little perplexing for me. But think the rationale does merit the focus. So --

DR. KROCH: Yeah, and there's also the related issue is implication is that lower risk of adjusted length of stay is better. And we can all think about reasons why that might be necessarily true.

CO-CHAIR HALL: Any additional comments on evidence? Paul?

DR. HEIDENREICH: Okay. So in terms -- since this is a re-endorsement, I assumed we'd be looking to see if there was evidence that -- if things had changed over time. And were the things that were present in 2008 when it was first, or whenever it was endorsed, have we succeeded? And are we now in a place we no longer need it? Because that's from staff. Is that a way to be considering this, or is it --

MS. PACE: That would be definitely under performance gap. So that is a

consideration for endorsement maintenance, if there's really no longer an issue with performance gap. That's the relevant question. Does it need to continue endorsement?

DR. HEIDENREICH: Yeah. It just seems that this should have gone to a group that was considering efficiency or cost measures. Because if we want -- I mean we can kind of put that hat on briefly and do that, if that's what you know.

MS. PACE: Yes, and basically we just tried to bucket these measures as best we thought we could. Since people oftentimes talk about length of stay and readmission together as being sometimes related, we thought it would be logical for a group looking at admissions and readmissions to look at it. It is a bit of a lone wolf, to be honest, in our portfolio. So kind of anywhere we put it, the committee would have gone, why is it with costs measures? Or why is it with

this?

So it landed with you. But there's no expectation that measures necessarily need to outlive their usefulness, if they're proving to be useful and driving improvement, then they should remain endorsed.

DR. KROCH: You could ask something a little different, which is, has the model changed sufficiently to warrant reexamination? And the answer to that is yes. And the main reason it jas changed is because of reporting requirements on the part of CMS. So we now have information that we didn't have before. That's very important when you're trying to control for patient variation and looking at risk adjusted length of stay.

CO-CHAIR HALL: Ron?

MR. STETTLER: Yes, and I'd say based on looking at the submission, there's still a lot of that that is for variation. Right, so the gaps between different hospitals, different regions, as you said,

sex, and age and various other demographic variables that you have in your model are definitely present. So whether or not overall length of stay has changed over time, I would say there's still a big gap and large variation across different groups that you're being measured on.

CO-CHAIR HALL: Leslie?

MS. HALL: I just had questions about -- have things changed? Years ago it was always based upon a midnight census discharge and then risk adjusted. And so now with electronic health records and we can actually get when a bed is vacated. Or that orders are placed, versus orders executed, versus bed vacated. Are we looking to improve the data collection to actually get a more accurate length of stay that might be more helpful?

DR. KROCH: Yes. I mean we started out -- the whole purpose of the measurement was to look at acute care. But actually, as

you can see in the write up, we actually look at other care settings. And it plays a very different role. And as you might imagine in SNFs and in rehab facilities, the whole issue of observation stays is one that we're now, I think, taking time to sort out. And what's kept us from sorting it out is because the government can't make up it's mind what the rule is. So we're going to wait for that to happen and then we're going to try to model it, as another dimension to this measure.

CO-CHAIR HALL: Steve?

DR. FISHBANE: Yes, two questions on definition. One, is it's mean not median, so what do you do with patients that are in the hospital for 150 days? It's very unlikely that it's the process of the hospital. Are outliers excluded?

And second question is just heterogeneity, you know when these global, as opposed to close, specific measures for hospitals, when it's global length of stay for

the hospital, if you have one hospital that ten percent of their cases are OBs with one or two day length of stays for the baby and the mom. Other hospital are doing no OB, mostly cardiac stuff. How do you deal with that?

DR. KROCH: Well, first of all, it's -- the model is stratified, so you're not directly comparing OB patients to cardiac patients. And that gets back to my other point, which is for practical purposes, that comparison doesn't usually help at all in trying to improve your processes. As far as outliers are concerned, this came up, I think, in the group that reviewed it. Their question was why did you choose 100 days, which was what we did choose.

And precisely because we discovered that doing geometric means does help with problems in the details, but not enough in cases where it looks like there are -- the number of days is so large that it is more likely that it's a data coding error than it

is a real number.

And so that's why we said okay. And that we did through validation testing over the years. And 100 turned out to be about the right number. Now in terms of sigma variation, it's well beyond six sigma, so.

CO-CHAIR HALL: We okay to vote on evidence?

MS. SHAHAB: Voting for 1a, evidence. 1 is yes and 2 is no. And the time starts now.

We have all 24 votes. 1a evidence. 23 yes. 1 no.

CO-CHAIR HALL: Opportunity to improve, or performance gap. Comments or questions?

MS. SHIPPY: I can start. So as Ron noted, there is some variation. I think the range is 2.1 to 6.8, and that was observed from 2010 to 2012. But I think the overall aggregate number, we haven't seen that much improvement on. So I do thank that there is

just a larger question that we have, or that I have, about is this the right lever for enforcing any sort of change or improvement?

DR. KROCH: Yes, its value, as I said before, in aggregate is not all that useful, not to the people who have been using the measure. It's being used for quality improvement. And it's very helpful there.

MS. SHIPPY: But just a follow up, I think there's multiple users that we would like to see use this measure information, beyond the internal purposes.

DR. KROCH: Yeah, I mean when you start looking at the aggregate data, and we've done this, done looking at hospitals for regions or counties or states or anything. We start to get into an analysis where you quickly find out that discharge policy plays a big role, and not entirely under the control of the hospital.

Because community resources turns out to be a big issues. And where we found

hospitals that were flagged with excessive lengths of stay, they were in places where they didn't have any place to send the patient. And that showed up.

Now that's again, from their point of view, that's what they wanted to know. You know, and of course I -- you can argue, at least I argue, that the hospital has a role in that. You know, the adequacy of community resources, the hospital should be applying the appropriate pressure.

MS. SHIPPY: Right, and there is question about whether it should be paired with another measure, but there was no recommendation that it should, to better provide a fuller quality picture. And that it was just possible to pair it with another measure and save readmissions and length. So I don't know if you can speak to that?

DR. KROCH: Yeah, it's possible to pair it with, obviously, readmissions.

CO-CHAIR HALL: I think we can

reexamine that question under some of the later categories as well. Any other performance gap questions?

Seeing none.

MS. SHAHAB: Voting for 1b. Performance gap. 1 high, 2 moderate, 3 low, 4 insufficient. Voting begins now.

We have all 24 votes. For 1b performance gap. 10 high. 11 moderate. 3 low and 0 insufficient.

CO-CHAIR HALL: Topic of priority. Allison or Ron?

MS. SHIPPY: Sure. We noted that it does represent a priority as defined by the evaluation criteria. And that it -- high numbers.

CO-CHAIR HALL: Other comments or questions?

Not seeing any.

MS. SHAHAB: Voting for 1c, high priority. 1 high, 2 moderate, 3 low, 4 insufficient. Time begins now.

One more vote.

For 1c, high priority. 12 voted high. 10 voted moderate. 1 low and 0 insufficient.

CO-CHAIR HALL: Scientific reliability and validity. Ron?

MR. STETTLER: So when I looked at the submission, the -- there seemed to be discussion about the reliability, explaining 60 percent of the variability was explained that's a lot. But I didn't see a lot of statistics actually in the document, a lot of statistics over time showing it was consistent. And had you know, time based reliability. But can you just talk a little bit more about your reliability measures.

DR. KROCH: Yeah, I, again, the model -- the measure's been around for a long time. So we didn't -- we don't do split sample analysis for example, which is one way to get to reliability. Because we have a natural split sample.

Which is to say, we have the history with the measure. And we can -- every time we re-calibrate the model, which we do every year, based on two years worth of data -- that's about nine million discharges -- we first look at whether if we run the new data on the old model, what do we get for readmission scores? Then we'd run the new model on the new model and what do we get for readmission scores?

And how much do they agree? And the answer you might not be surprised to hear is that these days, it doesn't change very much. So we don't think we've over fitted the model. That's a concern when you're working with general limiting models. We don't think it's over fitted.

MR. SETTLER: I guess the same question though goes back to you know, there's no c-statistics, there's not R-squared's. There's not a lot of information about the accuracy of the model other than your

statement of 60 percent.

So I don't -- without the evidence and the submission, it's -- I guess I talked to the staff, I don't know exactly how to review this.

DR. KROCH: Well, we're not talking about discrimination here, so c-statistic is not relevant to this particular measure. But when I refer to the 60 percent, what I was referring to was the R-squared.

Now that's the overall R-squared across 142 strata. Which I mean, obviously, we have an R-squared for each of those 142 strata. And that's in the detailed literature. I mean we you know, we have a document that's basically a large spreadsheet which gets updated every year, that shows all the parameter estimates and all of the sufficient statistics, and the quality of the fit. Since it's a general linear model, R-squared is a perfectly descent information criteria.

MR. STETTLER: And that wasn't supplied, so. Even the R-squared wasn't supplied. I guess you're saying the 60 percent was your low -- average R-squared across the 142 strata?

DR. KROCH: Well the way you do it when you have a stratified model is -- the simplest way to think about it is: imagine correlating the actual length of stay with the expected length of stay as predicted by the model. And taking that correlation and squaring it. That's an ordinary, unadjusted R-square. Yeah, I mean one thing that could have been in there would have been some -- I mean there clearly are strata that don't fit very well. And have R-squared's that --

When you -- first of all, let me make this point. The 60 percent is clearly big. And that's because it's taking into account the explanatory power of the principal diagnosis, which it's stratified by.

So if you then go down and look

within, you're going to get a lower R-squared. So none of the R-squared's for the individual, like heart failure or AMI, or whatever it might be, none of them are as high as 60 percent. They're all I would say they're in the 30 to 40 percent range at most.

CO-CHAIR HALL: I'm not sure if Ron wants to follow up, but Taroon wants to make a comment.

MR. AMIN: Ron, one of the questions you asked was you know, one of the questions was for staff effectively, how to evaluate this. I would just continue -- I know it's getting late, but I just want to continue to remind you that we have this guidance document here that has a little bit of an algorithm.

And if I may, you know, NQF, particularly for reliability and validity testing, requires empirical analysis. And you know, as you work through the algorithm down from the boxes, the first question is really,

having some testing of the statistical test, and not just descriptive analysis -- not just descriptive statistics.

And so I would encourage you to use this algorithm in the way you're sort of thinking about this. And there's other components here that might be relevant.

MS. PACE: Right. And the rating scale has a category for insufficient. So if there's insufficient information for you to evaluate according to the criteria, that's the rating you would use.

MR. SETTLER: Thank you.

CO-CHAIR HALL: And right now we're on reliability. Trying our best to stick to our guns on reliability. Any questions? We can vote on reliability and move to validity. Is that a question Sherrie?

CO-CHAIR KAPLAN: Yes, I have a -- so reproducibility, in terms of the way you were discussing it, because R-square wouldn't necessarily guide you along those lines. It

means, we just say that's the amount of variation explained? It doesn't say how much of that variation's reliable, how much -- and what's the error variance, the residuals may tell you something about that. But the explained variance does not.

And that's why I'm a little bit still perplexed. And do we know enough about the precision of the estimates, or exactly what it is you're comparing? Are you comparing models over time? Are you comparing the beta coefficients for the -- for what -- help us understand the reproducibility issue. Not just the explain variance.

DR. KROCH: Okay, so the really issue -- the really -- the issue is what the specification is. And whether or not, if you hold the specification constant and apply it to a new set of observations, do you see things move as a result of the fact that your parameters are shifting? And that's what I was describing.

CO-CHAIR KAPLAN: And you're -- So you are considering that as true score variation, not error? Right? Because error gives you a sense of trying to estimate how much of the score's actually moving in response to change, and how much of the score is actually moving in response to errors in measurement.

DR. KROCH: Right. And errors of measurement, not that depends on what you -- there's -- of course the errors in measurement for the raw rates could -- for the raw length of stay are very minimal. There's very reliable reporting on that. The error associated with the model, that's what I was referring to the sufficient statistics. So when I -- you want to answer the question, is your average length of stay higher than you expect? That's where the error variance comes in. And that comes from the model.

CO-CHAIR HALL: I see no cards raised. Let's vote on reliability.

MS. SHAHAB: For 2a reliability. 1 is high, 2 moderate, 3 low, 4 insufficient. Your time begins now.

One more vote.

We have all 24 votes. For 2a reliability. 1 high. 11 moderate. 6 low. 6 insufficient.

MS. KHAN: So this is in the gray zone. We're going to keep going.

MR. SETTLER: All right, I think we did start -- oh, go ahead.

So I think we did start to cover validity earlier with the fitting. I don't know if this is the right place to bring it up, but it does have socioeconomic in it, as an adjuster. And I think we need to bring that up. From what we heard this morning, I think that's conceptually okay at this point. Even though historically it wasn't. But in some instances we can have it in.

MS. PACE: Right. I think the question would be: what's the justification

for it? And you would want to see how that actually plays out in the analysis, with and without it. And you know, so.

MR. STETTLER: So I guess the question is, can you talk about your socioeconomic adjuster and the genesis of it. I don't think it was in the original model that was originally approved by NQF.

DR. KROCH: Actually it was. And I was part of the process, because I was the one who was submitted the work on it. And the subject of socioeconomic status wasn't discussed at all in 2008 or 2006. This issue has come up since. And as you noted, the -- I don't know how to call it, the guideline that NQF is going by, and other organizations as well, has shifted.

We have been using socioeconomic status for all of our outcome measures since the inception. But it's very straight forward. It's what information we have on the individual patient that we can either get

indirectly through where the patient lives. Or directly, which we can, for our proprietary data. So what's in there? Well, obviously race is in there. Income. We actually use distance as a proxy measure. Relative distance traveled. And --

MR. STETTLER: Payer class?

DR. KROCH: Payer class, yes. Well you know, payer class is yes, it is socioeconomic in many ways, yes. So it's in there.

CO-CHAIR HALL: So I guess, to your question, Ron, it's incumbent on us as a group to develop -- to develop whether we feel that this is justified that this needs to be here for this purpose, versus the rest of the measures we've discussed all day.

MR. STETTLER: Correct. I think the other -- I still am troubled by the 100 length of stay cutoff and, scientifically, how it was determined. So you simply have observed over time that it's about the right

level to exclude?

DR. KROCH: Well, at one point we weren't -- we weren't excluding outliers at all. And the argument was, we're using a semi-logged model, so we don't need to.

Turned out that that led to some false positives and false negatives that we felt was -- and that what was showing up, as we were doing the analysis, was that when we got stays that were that long, as I said before, in most cases the data were incorrect. The misplaced decimal point if you will, I don't know what.

So there was a real gap there. And in fact, there -- if you look at the distribution of raw lengths of stay, it's at -- well at about 50 or 60 days it falls off enough that you don't really have enough data to be able to do much analysis with the other points anyway.

But they do -- they can be very crazy and so they can then sort of wreck the

actual computation. 100 is -- is I agree, is somewhat arbitrary. I mean, it's based on just experience over years. We've played with different ones. At one point we had the outlier, was if you had a length of stay of at least a year, you would be considered an outlier. And we discovered that that led to reports that in some cases, again not usually, but in some cases, that didn't look right.

CO-CHAIR HALL: Thanks. I don't -- oh, I'm sorry, Steve?

DR. FISHBANE: Yes, I'm sorry to harp on the outlier thing, but you know it still doesn't make a whole lot of sense to me. You know, for looking at the efficiency of hospital operations and management of most diagnoses. So for the most typical diagnosis that are treated in the hospital, congestive heart failure, pneumonia, or COPD on the medical side, and the typical surgical diagnoses, you know an inefficient hospital will have length of stays that instead of

being five days will be seven or eight days.

A more efficient hospital will be three or four days. The admissions that go over 20, 30 days, tend to be problems that occur with a patient that usually, sometimes can be under the control of the hospital, but usually aren't. So that for our 17 hospitals, we, you know, usually we'll look at an outlier as being something longer than 30 days. And unfortunately you do have hospitalizations that go into the 50, 60, 70, or 150 day range.

When you're using a mean, it would just seem to me as something that the developer should consider, at least on an ongoing basis for the future, that you're no longer capturing the efficiency of hospital care, when you're allowing outliers at 100 days, or 80 days, or 60 days, I would just encourage something shorter.

CO-CHAIR HALL: Thank you. So this is a geometric mean, right, the comparison?

DR. KROCH: Yes, it's ending.

CO-CHAIR HALL: And it's stated differently in a couple of places in your document. But in one point it says this is reported as the days above the average. Is it those opportunity days you're recording, or the actual average length of stay adjusted?

DR. KROCH: Yes, that was -- that was incorrect. I think I corrected it in the documentation. It's not the number of days above -- excessive days. It is literally -- there's an OE ratio, just you might have for mortality.

CO-CHAIR HALL: Okay. So documents that are posted still state days above average, so we'll tend to that.

Kathy?

DR. AUGER: So just a point of clarification from the intern and the staff. I'm not totally sure how we still figure out the socioeconomics risk adjustment, but even so, like where do we consider that? Do we consider that in validation? Do we, the

validity? Okay.

CO-CHAIR HALL: So I have one other question as well. Intermediate complications, in the category of what you know about that's driving these. Are your intermediate complications in the hospital adjusted for?

DR. KROCH: Yes. They're risk adjusted complications.

CO-CHAIR HALL: All right. Steve do you still have a question? Or are you okay? You're all right.

Larry?

DR. GLANCE: Since we're talking about the scientific validity, we're talking about the validity of the risk adjustment. And so we're looking at the model and the model performance, if the R-squared is .6, it really is pretty phenomenal. And so I think we should go back to that concept, because we are evaluating the validity of the model. So R-squared of .6 is very, very good.

CO-CHAIR HALL: I would like to

hear whether anyone has any thoughts or concerns about the race, income, distance, payer class variables. Do we feel that there's adequate justification that those belong in this model, since this is the first time we've run into this issue. Allison?

MS. SHIPPY: I'm not compelled to have it included. I know that there's been discussion about other risk adjustment issues in previous measures. So it's difficult for me to rectify how this one's going to be different.

CO-CHAIR HALL: Wes?

DR. FIELDS: I think there's a tremendous, you know, volume of literature across the medical specialities that makes it pretty clear that low economic status, low income, and some racial attributes all result in higher -- results in patients presenting to the hospital in a later stage of their illness or injury. And that they have higher rates of comorbidities.

So I think SES is pretty fundamental to length of stay. And I think you know, without having an expert opinion about how it was formulated in this measure, you know, I think it does belong in the conversation on length of stay.

CO-CHAIR HALL: Frank?

DR. BRIGGS: I don't know the data in terms of how it impacts when they present. But certainly from the hospital's perspective, in getting them out and getting them placement, socioeconomic status, distance traveled, certainly impact the discharge timing. And you're more inclined to keep them an extra day, two days, things of that nature.

So it certainly impacts length of stay.

CO-CHAIR HALL: Larry?

DR. GLANCE: So, when I think of SES, I think of it boiling down to two issues. If you include SES in a model, you potentially adjust away racial disparities, which is bad.

On the other hand, if you don't include SES in the model, then you disadvantage hospitals and providers that take care of low SES patients. And with current approach to value based purchasing, those hospitals will get fewer resources.

And those are the hospitals that are in the worst possible positions to get fewer resources, because if your goal is to improve population health, you don't want to take resources away from those hospitals. I don't think there's a good answer to this. I don't think we're going to resolve it today. But I don't think it's unreasonable to include it in the model.

CO-CHAIR HALL: Cristie?

MS. TRAVIS: Well, I think this is what the entire SES project is supposed to be focusing on. And all of the same thoughts that are being shared here, are thoughts that are being shared in that process.

And it would seem to me that the

guidance from NQF at this point is not to use SES in risk adjustment. We've got the project going underway, which is supposed to develop a consensus around that, and therefore, it would be my position that we should wait and see what the project comes out with on the other end. And therefore I would not be supportive of including those in this particular measure.

CO-CHAIR HALL: Karen, are you wanting to make a comment? Your care is up.

MS. PACE: Oh, not about this though. So I don't know Helen what your thoughts are about process? I -- you know the difficult part with this performance measure is that it did originally -- it was originally endorsed prior to the guidance that we're talking about that's under question.

So I don't see any -- I don't know if there's any specific information about how -- what coefficient that gets, or what it's you know, contribution is to the model, that

might be useful information. But I don't think there's a right or a wrong answer at this point, in terms of the committee's action on this.

You know, depending on what happens with the risk adjustment process, we may be looking at other things as well. But you know, I -- I'm sorry I can't give you more guidance. And I don't know, Helen, if you have anything more to add.

DR. BURSTIN: No, it's -- Jean's right. This measure came in before the guidance that said don't do it. So they're in a bit of a funny place, compared to other measures that have come subsequently. So I feel like you know, part of our process is to evaluate measures based on what has been put before you. Guidance has changed since then. So I think -- I think it really is up to you guys, and the Chairs to make a decision. We can always revisit it post-comment. We'll get better clarity as we move forward.

But I you know, I think -- just want to emphasize the fact that this measure was in fact put forward when there was not this guidance by the developer.

MS. PACE: I just had -- wanted to make sure I understood, because I didn't. What is happening with the long episodes? Because it's not listed in your specifications as an exclusion. What are you doing if the episode is over 100 days, what are you doing with them? Are you ex --

DR. KROCH: They're being excluded, yes.

MS. PACE: Okay. Because it's not listed as an exclusion, and there's no analysis as excluded, so.

DR. KROCH: Oh, it should be, okay.

CO-CHAIR KAPLAN: Let me just jump in real quickly and say when you -- anytime you modify a measure in the way it's kind of framed, especially one that's been in the works for a while, then anything you look at

could be the way you modify the measure, as opposed to actual changes. So you face this tension between do we fix the problem now and in advance of this guidance, or do we change it, and then you compromise comparability.

So inevitably there are all of these trade offs between when you change something to make it more reasonable. And then lose all that information that you've been tracking over time, because you've actually change the measure and you can't see differences between actual variability and things that you did with the measure itself.

CO-CHAIR HALL: Paula?

MS. MINTON-FOLTZ: Yes, I was wondering if there was any risk adjustment thought about like obesity, like history of substance abuse? Those are the ones that we can't, as hospitals, they're not attractive to post-acute. So if they need post-acute care, we can't get folks out of the hospital with that.

So I didn't know if there was a way -- 100 days, they could be on variance. I mean we could be ready to discharge them way before 100 days. But there's no access to post-acute care. So again, that kind of goes back to health of the community.

CO-CHAIR HALL: Allison?

DR. KROCH: Is that a question I was supposed to answer? Or --

CO-CHAIR HALL: That's all right.

MS. MINTON-FOLTZ: Well I think it -- yeah, it's just that the risk adjustment needs to be more than just probably CMI.

CO-CHAIR HALL: Okay, we will move to vote on validity.

MS. SHAHAB: Voting for 2b validity. 1 is high, 2 moderate, 3 low, 4 insufficient. Time begins now.

Just one more vote.

We have all 24 votes for 2b validity. 3 high. 16 moderate. 4 low. 1 insufficient.

CO-CHAIR HALL: Feasibility specific comments?

CO-CHAIR KAPLAN: Actually I don't think we had any comments on feasibility.

MR. STETTLER: I mean it seems that it's pretty straight forward. He's been doing it for a long time.

CO-CHAIR HALL: All right, we'll move to vote.

MS. SHAHAB: Voting for 3 feasibility. 1 high, 2 moderate, 3 low, 4 insufficient.

We have all 24 votes for feasability. 22 high. 2 moderate. 0 low. 0 insufficient.

CO-CHAIR HALL: Usability comments?

MR. STETTLER: This is where I think we mostly focused on the proprietary nature of the model. I know it's not conceptually proprietary. But help me understand how others can use it without actually going through the coalition.

DR. KROCH: Without actually going through what?

MR. STETTLER: Whatever you call your 850 hospitals.

DR. KROCH: Oh, yes, right. So the best thing I can say about that is that there have been a number of systems that basically took the measure and fitted it themselves. But these were relatively large hospital systems.

So there is an issue there. If you're small, it's an awful lot of work to actually apply the model in a reliable way to your data.

So it's you know, I think you -- it's a legitimate point. It's a big model. And you have -- there are a lot of parameters to keep track of. Can it be done? Yes it's been done by a number of people.

And it's not conceptually difficult. It's just that's a lot of data.

CO-CHAIR HALL: Leslie?

MS. HALL: I'm sorry, I thought that I read that it was -- although it was proprietary, there would be no licensing fees associated with this for others to use?

DR. KROCH: Yes, well we -- we made -- that's always been understood.

MS. HALL: Okay.

DR. KROCH: If we made this measure that could be used, that we would -- I mean we've applied the -- what we use to do is we used to apply a number of measures, and the length of stay was one of them to the MedPar data. And we just posted it on the website. And anybody who wanted to look up how they were doing could just look it up.

So --

MS. HALL: The instrument itself requires licensing, but the reports are not?

DR. KROCH: Well the -- If you don't -- there's an interface that requires licensing that let's you do all kinds of kind of cool things that you can't do just by

running straight reports.

So it allows you more flexibility in how you subset and that sort of thing.

CO-CHAIR HALL: Allison?

MS. SHIPPY: I think we talked about this on the conference call a couple of weeks ago. I'm harking back to the 4a, the accountability within three year -- yeah, accountability use within three years and public reporting within six.

And this has been endorsed for six, maybe eight years. And there was some mention of a brief conversation with OCSQ about having this publically reported. But if you could speak more to your plan.

I take it seriously that NQF is you know, working as a purveyor of a public good, so.

DR. KROCH: So you mean our plan for making it publically available? Well --

MS. SHIPPY: An accountability program, or any other transparent way.

DR. KROCH: So, I was referring to the actual guts of the model are organized in a multi-tabbed spreadsheet that let's you essentially take some data and insert it in the data provision tab and actually get the risk adjusted length of stay out of it.

I mean so that model has been around for a while. It -- the thing is that you -- well, again, it really depends on what kind of an organization you are in terms of how comfortable you are with the patient-level data.

Because the one aspect of this model is that the unit of analysis is the patient. So you have to be working at that level. And therefore you're working as I said, with it's large not just in the sense that there are a lot of parameters, it's large in the sense that there are a lot of observations.

So -- but it will fit, but --

CO-CHAIR HALL: Larry?

DR. GLANCE: Quick question of clarification. Is the model in the public domain? In other words are the risk factors and the beta values in the public domain?

DR. KROCH: Yes.

DR. GLANCE: Okay.

DR. KROCH: Yes, they are. That's what that spreadsheet is about.

DR. GLANCE: Okay.

CO-CHAIR HALL: Kathy?

DR. AUGER: And is that the NQF, what they -- is that the NQF official, that just the measure, like the betas have to be in public domain, is that what the requirements are?

MR. AMIN: I would say the specifications need to be clear enough for somebody to be able to pick up the measure and be able to run it. Reproduce it, yes.

CO-CHAIR HALL: Along the lines of this topic though is the question of improvement. And you've actually shown some

pretty remarkably stable numbers over time. With as you mentioned previously, maybe some hiccups lately.

But do we meet this criteria that use of this information is enabling organizations to improve over time? Is that demonstrated to us?

Because if it's been in play for six, seven or eight years, it feels like we should know by now.

DR. KROCH: Yes. The numbers are used routinely for looking for opportunities to improve care. Usually from the point of view of streamlining the focus of operations.

So and I don't know if I'm answering your question.

CO-CHAIR HALL: So some of the information you report for instance, over a years and years, shows length of stay and the 3.1, 3.2 region, pretty darn stable. Then you mention that some sub-models demonstrate more improvement over time then others.

That raises the questions should those sub-models be up for approval and should everything else be dropped?

DR. KROCH: Oh, I see, I see. Yes, because where we've seen -- there are definitely areas that you might expect have improved. Orthopedics is one. Cardiovascular treatment is another.

There are a few categories like that where there have been quite a bit of movement in length of stay. But when you look at the whole hospital, the number doesn't change much.

CO-CHAIR HALL: Understood. But one of the criteria in front of us is that when a measure has been in play for a period of time, we are expected to see, as it says, progress on improvement.

Sherrie?

CO-CHAIR KAPLAN: Just to follow up on that. If you did an interrupted time series analysis or just a straight forward

time series analysis over like you know, from 2008 forward, would you see a trend that would be significant? Or would you see a flat line?

DR. KROCH: There -- you're talking about overall at the hospital level? There is a trend. And it's been downward.

But there have been some surprises some years where that trend did not carry through. And we've always -- looking for the fact that is there a natural, if you will, asymptote if you will, is there something that basically you know, you're not going to get below that.

And at some point when you have a length -- an average length of stay, a geometric length of stay -- geometric mean length of stay of three days, you can see why that can be an issue.

CO-CHAIR HALL: Thank you folks. We've passed an hour on this measure.

Cristie?

MS. TRAVIS: Well I guess I just

need another little bit of clarification. Because 4a up there says that it needs to be used in accountability within three years, or public reporting within six years. And so I'm thinking that the results are what we're talking about.

In other words, and I guess I didn't hear that these results were being used in public reporting. Something that the public could go get access to these results.

So I guess I just need a little bit of clarification. Not the model itself being in the public domain, but the results of who's reporting on these results publicly for use by the public?

DR. KROCH: Yes, and we've made it available publicly for public data. That's what I mean by MedPar and NIS.

MS. TRAVIS: Okay, thank you. Thank you.

CO-CHAIR HALL: Frank?

DR. BRIGGS: So, having been in

Premier and having our data in there, Premier is one of multiple organizations that have tools like this. All of them I think, without exception, produce their own modeling around these similar things such as cost of care, length of stay, mortality.

So I'd be very surprised if any one of them would ever get to the point that it would be for accountability or public reporting.

Now I can tell you, like I said, as a previous member in Premier, and our system still is, we do use it and we have done demonstration projects around different disease populations using the data to look at one of our measures.

So if we wanted to look at how we care for COPD, we would use the data set and look over a period of time before and after intervention if their length of stay shortened.

I think there's lots of evidence in

terms of that as presented at their annual conference every year at the Premier Breakthroughs Conference, of people using the data to drive improvements.

CO-CHAIR HALL: Okay. Great discussion. We will vote on usability.

MS. SHAHAB: For usability and use, 1 is high, 2 moderate, 3 low and 4 insufficient. Time begins now.

Just one more.

Now we have 23 responses for usability and use. 1 high. 14 moderate. 6 low. 2 insufficient information.

CO-CHAIR HALL: So we'll take an overall vote. Any additional comments or concerns before the overall? Not seeing any.

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

We have all the votes for overall suitability for endorsement. For measure 0327, Risk-Adjusted Average Length of

Inpatient Hospital Stay, 13 yes and 10 no.

CO-CHAIR HALL: All right, that's in 40/60 zone, so you're all familiar with that. We thank you Gene from Premier.

And we will invite our next developers up from Acumen. So we -- yes, so we need to do a public comment at 5:15. But since it's 5:10, we have to allow for people to be jumping on the phone at exactly 5:15.

So we'll have our developers come to the table for the next two measures, 2505 and 2380. And -- ah, they're on the phone. They're on the phone, okay.

So let's just -- before we embark on that discussion -- all right, we're organizing here. Great.

MS. KHAN: I believe Deborah Dietz is also on the phone.

CO-CHAIR HALL: Operator, could we open Deborah Dietz' line.

MS. DIETZ: Oh, that was me. I was on mute. Deborah Dietz is here, hello.

CO-CHAIR HALL: So can we enter -- can we ask now if there's any public comment in the room? Okay, let's do that.

MS. KHAN: Is there any public comment in the room?

CO-CHAIR HALL: We're not seeing any public comment in the room. So we'll ask our colleagues from Acumen to go ahead and introduce themselves and we may do one more check on public comment in a few minutes.

So please introduce yourselves and a brief introduction of your measure.

DR. COOK: Sure. In the interest of time, would you like me to introduce both of our measures at once?

CO-CHAIR HALL: Let's do the first. Let's stick to our guns here.

DR. COOK: Okay. No problem.

CO-CHAIR HALL: In case there should be any disagreement between the two, let's not muddy the waters. So let's --

DR. COOK: All right. My name's

Keziah Cook. And I'm an Associate Policy Researcher at Acumen, LLC.

We are developing claims-based quality measures for home health agencies under contract with CMS. And these complement some claims based measures of hospitalization that are -- were previously developed and NQF endorsed and publically reported.

So the first measure we're discussing is the emergency department use without hospital readmission during the first 30 days of home health.

This applies to all home health patients who begin home health within five days of hospital discharge. And measures the occurrence of any emergency department use that does not result in an admission to the hospital during the first 30 days of the home health stay.

About nine percent of home health stays involve a emergency department use of this type during the 30 day period. And those

stays -- or those emergency department visits account for about $45 million dollars per year in expenditures.

We propose to publically report this measure using categories. So each home health agency would be categorized as better then expected, same as expected, or worse then expected, using a statistical test.

And based on about three years of data, those results would be publically reported on the Home Health Compare website.

This measure complements the re-hospitalization measure that we'll be discussing in a little bit in that both measure, acute care usage by previously hospitalized home health patients. And it also complements a measure that's currently publically reported called emergency department use without hospitalization, which applies to all home health patients, not just those who were previously hospitalized.

The measure is risk adjusted using

a patient level multinomial logistic model. And that patient level model is what allows us to calculate an expected distribution of rates for each home health agency.

So when we're saying a home health agency is better than expected, we're saying that based on the expected distribution of rates per their patients' individual health characteristics, that agency was in the lower tail. So they had significantly fewer emergency department visits without readmission than anticipated.

So happy to just stop there.

CO-CHAIR HALL: Do we want to invite public comment again at this point? Okay. So we'll break for one second.

MR. AMIN: Operator are there any public comments on the line?

OPERATOR: If you would like to leave ask a comment at this -- question or comment at this time, please press star 1.

CO-CHAIR HALL: Can you clarify

public comment on any item discussed, not just this.

OPERATOR: There are no public comments at this time.

MR. AMIN: We'll take any public comments on any of the measures discussed in the room or on the phone. So are there any public comments on the phone?

OPERATOR: Again, please press star 1. There are none at this time.

MR. AMIN: Okay, thank you.

CO-CHAIR HALL: Okay, thank you.

So thank you for that brief introduction. We'll ask Wes and Pam if they would like to kick off the discussion for evidence.

DR. ROBERTS: Okay. One of the issues that had come up on one of the calls was that the measure developer was going to provide a categorization of agencies by region at the median. So I don't know if you want to --

DR. COOK: Interested, or I can just briefly describe the findings and if anyone wants to see the tables.

DR. ROBERTS: Can you briefly describe it.

DR. COOK: Sure. There were variations across regions. The observed rate across CMS regions, so the 10 CMS administrative regions, range from 7.4 percent in the New York region to 11.1 percent in the Seattle region.

And as one might anticipate based on that, the New York region had the most agencies who are categorized as better than expected, with 27 percent of agencies in the New York region categorized as better than expected. And the Seattle region had the largest number of agencies categorized as worse then expected, with 29.3 percent categorized as worse then expected.

The other variation across regions that stands out here is the fraction of

agencies across regions categorized as same as expected, differs fairly substantially from a low of about 70 percent of agencies who are same as expected in Seattle, New York and a few others. Up to a high of 88 percent of agencies in the Dallas region that are characterized as same as expected.

And what's really going on there is the Dallas region, so Texas and surrounding states have a predominance of very small agencies where it's difficult to make a statistically significant statement about whether your know their performance caring for 25 patients differs from expected.

CO-CHAIR HALL: Okay, thank you. That feels a little bit like evidence and performance gap. So we'll hit both of those very quickly.

Any other specific comments on evidence? Yes, Karen?

DR. JOYNT: I just didn't see much evidence that we have any idea that home

health does anything for readmissions. Or that they have the -- I mean home health at least in my experience discharging patients from the hospital can range from someone going to draw blood next week to someone providing really pretty comprehensive care.

And the evidence -- I mean the data on readmissions is not great in general at showing that any intervention can work. And there's certainly not a particularly robust amount of data suggesting what structure and type, and set up and cost, and arrangement and providers for home health can.

So I'm a little confused about the evidence base.

CO-CHAIR HALL: Great. Thank you for those comments. Anyone else what to chime in? Wes were you -- you were also a primary discussant. Any other concerns you want to -- or points?

DR. FIELDS: Yes, I just want to follow what Karen said. The evidence as I

recall, it really was about innovations in some home health services. With different kinds of home monitoring, home surveillance.

And there's a lot of activity in paramedical care. And different ways of extending care into the home environment or those kinds of things.

But I think that it's really kind of an underwhelming amount of evidence in terms of peer-reviewed literature. And what was there to me didn't speak for the need to have this metric and this particular measure as a way of driving innovation.

The other one, if I heard right, if the total number of ED visits that are in this category cost $45 million, in the Dirksen School of Economics, that you know, applies in this town, that's the -- that's very -- that's a rounding error.

So I just don't see that there's a huge amount of resources to be saved here. And I guess fundamentally I see this as most

likely to be a failure discharge planning, or a case management or a care coordination.

And that's supposed to happen before folks leave the hospital. And in many cases, as Karen said, the services that would make a difference probably aren't services directly provided by home health agency.

So I was a little at a loss for this one.

CO-CHAIR HALL: Anyone else comments about evidence. Yes Paul?

DR. HEIDENREICH: I'm just wondering, technically do we need to have any evidence at all since this is an outcome measure? And we just need a plausible rationale that might be you know, might have some evidence someday?

CO-CHAIR HALL: Yes. That's the specification of the guidance that's in front of us.

Karen?

DR. JOYNT: I'm not sure the -- I

think the thing -- we have plausible evidence that things that take place outside the hospital prevent readmission. But I don't know that we have evidence that home health agencies control what they do that prevent readmission. Right?

I mean I don't know how it works for everybody else's clinical experience. But in my experience, when we send someone home, the inpatient team makes the decision what they're sending them home with.

So I'm just confused about the structure under which a home health agency would be working with an inpatient setting or with -- I just don't totally understand how the home health agency themselves are the ones who are actually making any decisions about what's delivered I guess.

That's the plausibility that I don't quite understand. That may just be my misunderstanding of the system, but --

MS. DIETZ: This is Deborah Dietz

from the developer. Do you want any input n that?

CO-CHAIR HALL: One second Deborah. Let's take one question that's raised here and then we will ask you.

MS. DIETZ: Sure.

CO-CHAIR HALL: Thank you.

MS. DIETZ: Okay.

CO-CHAIR HALL: Kathy?

DR. AUGER: I was just going to say that I think there is a bit of heterogeneity in terms of what home health offers. And then there's some decisions by home health companies about when to escalate versus not.

And so I think that you get some variability there. But then I also think that the fact that we're seeing variability in performance sort of argues that there may be things that home health can do to sort of move this metric because we are seeing variability in performance.

CO-CHAIR HALL: Thank you.

Deborah do you want to respond to some of what you've heard?

MS. DIETZ: Yeah, I just want to, you know bring up that there are a lot of best practice guidelines that are -- have been promulgated by the quality improvement organizations and that agencies use in order to reduce their rehospitalization rate.

And it has to do everything from when they schedule visits in terms of front loading visits. So that when a patient is perceived to be a risk for rehospitalization, that they're out there quite a bit right at the beginning.

Medication reconciliation where they're working with the patient to make sure that the meds that they were prescribed at the hospital are actually in their home. And the patient knows how to take them and isn't taking two of the same thing.

Education, looking to see when the patient -- what the home environment is like

so that if the patient perhaps is at risk for falls, that they're doing things to try to minimize that. Getting PT in there when it's appropriate.

So agencies in general recognize that there are things that they can do and should do in order to minimize rehospitalization.

CO-CHAIR HALL: Okay, thank you. I don't see any additional cards raised. Let's --

DR. BURSTIN: Quickly, I believe there's actually been a long standing measure, correct me if I'm wrong, of ED's for home health patients.

DR. COOK: There is. There is a current measure that's publically reported that applies to all home health patients. This measure applies specifically to those patients who come to home health from an inpatient setting, who are discharged from a hospital and then begin home health within

five days of that discharge.

So the existing measure applies to the entire home health population and this applies to a more narrow population that may be at particular risk.

DR. BURSTIN: Okay, I understand there's been a very long standing focus on reducing ED use broadly among the home health community. I wanted to emphasize, and there had been.

I mean again as primary care doc, perhaps different, lots of interventions that home health takes to in fact try to reduce that bumping back into the hospital.

DR. JOYNT: It think that would be a really helpful thing to read. I just didn't -- I was not even aware that existed in the evidence or I would be happy to read it.

CO-CHAIR HALL: All right, we'll move forward with a vote on evidence.

MS. SHAHAB: Voting for 1a, evidence. 1 yes, 2 no. Time begins now.

We have all the votes for 1a evidence. 16 voted yes. 6 voted no.

CO-CHAIR HALL: Well move on with performance gap or opportunity to improve. Pam or Wes did you have specific concerns?

DR. FIELDS: I'll defer.

CO-CHAIR HALL: Can you hit your mic Pam.

DR. ROBERTS: There was a 9.1 percent rehospitalization to ED, which was one of the priorities and leads to homes gaps. And that there was -- there was concern and I think we've already talked about with the care coordination with the hospital of follow up and access.

CO-CHAIR KAPLAN: Can I just comment on something. This is the biggest difference I've seen I think in that worse then expected variability by geographic region. It grows from 3.9 percent to 29.3 percent.

So in terms of a performance gap,

you know, that is rather major. I was -- again that's -- I think amongst the measures that we've considered so far, the biggest gap.

And the question is what percent of that is mutable, blah, blah, blah, blah, blah. But at least for the gap, looks to me really impressive. I don't know if anybody else had gotten that.

CO-CHAIR HALL: Any other comments? We'll move to vote on this opportunity.

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

We just need one more vote.

All votes are in for 1b performance gap. 12 high. 8 moderate. 1 low. 1 insufficient.

CO-CHAIR HALL: Priority. High priority area.

DR. ROBERTS: There was a study that was provided that ED visits and inpatient hospitalization showed an importance for the

coordination of care. 54 percent of all ED use after inpatient say was utilized. And then we've already mentioned the 12 percent readmission. And the geographic disparity.

MS. SHAHAB: Are there any additional comments?

CO-CHAIR HALL: Wes?

DR. FIELDS: Yes, I think really for me, the priority you place on this really depends a whole lot on the specifics about how you define the emergency department visit.

And this is where if we really are talking about bad discharge planning or inadequate discharge planning, or inadequate community resources, to me there's a big difference between a you know, a home care patient coming back for a skin tear and getting a Steri-Strip and going home. And one that's going to be in the hospital for two midnights because of something like COPD or diabetes, or you know pick your cost, your Medicare cost center.

So one of the things that I was not comfortable with was I didn't see anything in the structure, the measure, that would distinguish between really, what would amount to a trivial ED encounter with a patient rapidly being returned back to the community. And one that was probably unstable for discharge home in the first place, and who's bouncing back because they're clearly beyond the scope of the home care agency to deal with.

And I just didn't see in the definition of this measure that we would be able to identify the difference between those two encounters in the emergency department.

CO-CHAIR HALL: And that's in the context of their not being a readmission, right?

DR. FIELDS: Yes, so I think this one will be entangled in whatever the outcome is from the two midnight rule, and the future definition of observation services. And it

relates a little bit I think to that M&M report I mentioned earlier, which is also the leading edge of the health service research.

So it's tough for me to give you an up or down vote on this because if I really am not sure if we're talking about Steri-Strips or acute coronary symptoms. It's tough.

CO-CHAIR HALL: Other thoughts? Larry?

DR. GLANCE: So, I think that's a really great point. But I think that same limitation that you're describing for this particular measure is probably shared by a lot of the other measures as well.

And clearly if somebody is readmitted within 48 hours of discharge, whether it be because of home health err -- whether it be the home health care measure, the dialysis measure, a lot of other measures that we've considered.

Clearly a readmission with a very short -- within a very short time frame is

probably I think something to do with what happened during the hospitalization. Versus if it's seven or ten days out, it is more likely related to what's happened subsequently.

I don't think it's really been dealt very well with any of the measures that we've considered today. But I think it's a very interesting point that you make.

CO-CHAIR HALL: Other thoughts or comments? Well vote on priority.

MS. SHAHAB: Voting for 1c, high priority. 1 is high, 2 moderate, 3 low, 4 insufficient. Time begins now.

Two more votes.

We have all of the votes for 1c, high priority. 2 high. 14 moderate. 5 low. 1 insufficient.

CO-CHAIR HALL: Reliability, validity. Opening comments? Microphone.

DR. ROBERTS: There was -- they used the -- conducted the split half

reliability test that we discussed before. They did have, the result showed a high level of internal consistency.

And they did find that the transitions between better then expected and worse than expected are extremely rare, which shows that categorizations seem to be robust.

And I'll open. That was reliability.

CO-CHAIR KAPLAN: Pam, can you move a little closer to your mic because we're having trouble.

DR. ROBERTS: Right. Do I need to repeat what I said?

And then for validity. They looked at --

CO-CHAIR HALL: Pam, let's hang on reliability just one sec. Any other comment on reliability? Let's vote reliability.

MS. SHAHAB: For 2a reliability. 1 is high, 2 moderate, 3 low, 4 insufficient. Time begins now.

We need one more vote.

We have all 22 votes for 2a reliability. 4 high. 14 moderate. 3 low. 1 insufficient.

CO-CHAIR HALL: Pam why don't you pick up with validity.

DR. ROBERTS: So validity, they looked at Medicare certified agencies with at least 20 home health stays from July of 2010 to June of 2013. They did use a risk adjustment model and developed this using 80 percent random sample.

And they had some -- they did have exclusions on their risk adjustment analysis. And they used OASIS measures, which is their measure used in home health. It was used in validity testing. And also without home health readmissions, were associated with underlying quality for readmissions.

So I'll open it up from there.

CO-CHAIR HALL: Other group members with comments? Seeing none -- oh, wait,

Karen.

DR. JOYNT: A couple of things. One, I don't totally understand the exclusion of cancer. And the second thing is I don't totally understand how we can account for the selection into this exposure.

Which is true for any of our metrics. There's a selection into a hospitalization, I get that. But this seems to be one in which the use of home health is so different that the selection into the exposure actually makes a big difference in what your rates -- your expected rates are going to be.

And unless the risk adjustment model here is a lot better than our other ones, I'm not sure that we can really account for sort of that selection bias into the metric. But I'd be interested in hearing the thoughts.

I'm trying to read this, but I haven't gotten to that yet at this end.

CO-CHAIR HALL: Any group member comments? Does our developer want to reply to that?

DR. COOK: Sure. So I mean I think just -- you know home health is a different population. You know different types of patients are discharged home with home health care then those who are discharged to a SNF for instance.

In terms for how we account for those differences. First of all you know this is a measure that specific home health agencies, we're not trying to compare home health agencies to SNFs or to other care settings.

We also account in our risk adjustment model for the Medicare HCCs that the patient experienced in the six months prior to home health. And also we use the DRG from the hospital discharge as one of our risk factors.

And then finally, we use the length

of stay of the proceeding hospital stay.

To answer your question about why cancer's excluded, that's actually for consistency with the hospital wide readmission measure that's calculated as part of the hospital value based purchasing program. That measure excludes certain kinds of index hospitalizations and also certain kinds of rehospitalizations.

So we mimic their exclusions of index hospitalizations for this measure.

CO-CHAIR HALL: Wes?

DR. FIELDS: Yes, here too, I do have qualms about validity. I think my clinical experience and my 2000 partners is that often people are referred to home health because it's the only post-discharge service they're eligible for in terms of criteria.

In many cases these are low income persons, families that have other issues or struggles and as often as not, having done this 30 years and 90 thousand times and across

the street from Leisure World, these patients can come back with non-trivial issues that are a reflection really an inadequate community care plan.

And I guess what I'm troubled by is that it winds up looking like a measure of our performance, the emergency department, for referrals we didn't make, for discharge decisions we didn't make. And for resources we can't provide.

CO-CHAIR HALL: Question Helen?

DR. BURSTIN: I guess my interpretation of this is it actually reflects the quality of the home health agency, not the ED. Did they do everything they could to avoid having the patient go to an ED?

Obviously when patients are sick enough, they're going to go anyway.

DR. FIELDS: We'll wait for the headline in USA Today.

DR. BURSTIN: It's very clearly and if you just look at -- I mean again, we're

just being, does the care, you know, the level analysis does -- when you say facility here, do you mean home health agency just to clarify?

DR. COOK: I mean home health agency. It's a little bit of a misnomer with home health because they don't have an actual building or anything.

DR. BURSTIN: So again if endorsed measures are used that the level of analysis that have been tested and endorsed, so I think that's confusing because it says facility and so your immediate mind goes to hospital. It's in fact the home health agency if that helps any.

CO-CHAIR HALL: Sherrie?

CO-CHAIR KAPLAN: Can you -- can you help me clarify a little bit. Because when I read this application, it said the percent differences between better and same or worse agencies was like 3.5 to 6.5 percent depending on which validity variable you use.

And on one hand everybody was about as expected, so the model did a good job. On the other hand, that's not a real good you know, evidence of discriminate validity for example.

But then when I looked at the data you provided in table three, it looks like on average there's a lot more better and worse then expected, which would suggest that there is some discriminate validity.

So can you help us understand what those data are about.

DR. COOK: Sure. So I think biggest difference is that in our original submission, we also included analysis of the agencies who have between 1 and 19 stays. In this regional breakdown we already excluded those because we're not intending to publicly report that information.

I think the -- you know the kind of most interesting discrimination that this measure can do is actually for the larger

agencies. So the agencies with 200 plus or even 1000 plus patients. Which is certainly only a subset of agencies, but they do account for treatment of the majority of home health patients.

CO-CHAIR HALL: Other comments? Validity. Seeing none.

MS. SHAHAB: Voting for 2b, validity. 1 is high, 2 moderate, 3 low, 4 insufficient. Time begins now.

Still waiting for two votes.

For 2b validity, 0 high. 17 moderate. 4 low and 0 insufficient.

CO-CHAIR HALL: Specific comments on feasibility? Going once, going twice.

MS. SHAHAB: Voting for feasibility, 1 high, 2 moderate, 3 low, 4 insufficient. Time begins now.

Just one more.

Voting for number 3, feasibility. 10 high. 10 moderate. 1 low. 0 insufficient.

CO-CHAIR HALL: Usability. Specific comments? Karen?

DR. JOYNT: Sorry, but once again I just think the best way to get better at this is to select lower risk people into home health.

And we see that, if you look at the differences in utilization of home health in Florida versus here. They send people home with a bag of IV fluids you know, because they're not quite hydrated yet late in the stay.

And those people are never going to come back. They're healthy people who had their appendix out or something like that.

At least from what I understand of the home health data, there's such a big difference in utilization that I think it's a really tricky denominator and a really easy way to improve this is to send more people home with pretty minimal services.

CO-CHAIR HALL: And your fear is

that that's not captured in the diagnosis that are adjusted?

DR. JOYNT: I can't tell. It doesn't look like it. I mean I might just be reading it wrong.

But it seems to me like home health is just -- it's a very selected service that can be very different in it's intensity. So maybe I'm just misunderstanding how well the model captures the intensity of service delivered.

I certainly happy to be pushed back on if I'm getting it wrong.

CO-CHAIR HALL: Fair enough. Wes?

DR. FIELDS: Really nothing else.

CO-CHAIR HALL: Pam?

DR. ROBERTS: If I remember correctly, weren't the low level LUPAs excluded? So that would take care of the real short stay ones?

DR. COOK: That's right. LUPAs are sort of home health encounters that involve

four or fewer visits.

So we did exclude those. It actually only has a very minimal effect on the agency level outcomes. But it sort of seems to get at this point, that if someone's only providing very minimal services, perhaps it's not appropriate to include them in this measure.

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

MS. SHAHAB: Voting for criteria 4, usability and use. 1 high, 2 moderate, 3 low, 4 insufficient information.

We have all 22 votes for usability and use. 1 high. 13 moderate. 6 low and 2 insufficient information.

CO-CHAIR HALL: Any comments before the overall vote? All right, we'll vote overall.

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

We have all the votes for overall suitability for endorsement for measure 2505, Emergency Department Use without Hospital Readmission During the First 30 Days of Home Health. 15 Yes. 7 No.

CO-CHAIR HALL: Thank you. So we are potentially going to lose our conference lines and our sound at 6:00 o'clock.

Sherrie and I would like to argue that we should push forward with the next measure because they're parallel in some sense. And our developer is here. We don't want to make them come back again tomorrow morning.

Is everybody -- or are most people okay with that? Okay, all right. We're not going to --

CO-CHAIR KAPLAN: Yes. Let me make a bid for -- yes.

CO-CHAIR HALL: No, 1 high, 2 low. No, no. For the sake of time then, can we ask our developer to call our attention to the

highlighted differences between what we just discussed and the next measure.

CO-CHAIR KAPLAN: And let me make a bid please for really concise, sharp -- I know everyone's worked really hard all day long. So it's really a tribute to you that you're all still firing on all your cognitive cylinders.

But if you can just sharpen up the you know, comments. And really make them concise. And the response from the developer --

CO-CHAIR HALL: And where were you like eight hours ago?

Okay, developer.

DR. COOK: Sure. So in a lot of ways this measure is actually very similar to the measure we just finished discussing. The rate of rehospitalization among home health patients is a bit higher. It's about 13.5 percent within the 30 day time frame.

The denominator of this measure and

of the measures we just discussed are identical. So that's identical.

The numerator is capturing you know, those patients who are rehospitalized either through the emergency room or direct hospital admission.

The interventions that we anticipate the agencies can take to prevent rehospitalization are parallel to those that they take to prevent emergency department use without readmission. Although again, they may be targeting those more severe kinds of conditions.

So for instance a patient who's surgical wounds are deteriorating, they require hospitalization. Where as a patient who you know, skins their knee, might only need emergency department visit.

But overall, we're thinking that the home health agency is working to prevent the escalation of care to the next level. And that once the care is escalated to the next

level, in some sense it's out of their hands, whether the doctors in the emergency room feel that it's best to admit the patient or whether they can stabilize the patient and then return the patient home.

Again the proposed public reporting is categorical. You know overall, I think we're a little less good at categorizing on this measure then the other measure. Although in large part that's because our risk adjustment model does a little bit better on this measure than on the ED.

So we have more same as expected's because our risk model is capturing a little more of that -- of those nuanced differences between patients.

But still among the largest agencies, we are able to categorize a substantial faction of those largest agencies as better then average or worse then average. And we do see the variation across regions in the fraction of agencies that are categorized

as better or worse then expected as we did on the previous measure.

CO-CHAIR HALL: Pam you're listed as discussant. Do you have any specific comments on the evidence?

DR. ROBERTS: For sake of time, it's very similar to the other measure.

CO-CHAIR HALL: Okay, any other -- yes, Wes?

DR. FIELDS: Yes, I actually -- I think the rationale here is stronger. I mean if this encounter results in admission to the hospital, I think that speaks for itself.

The other thing that I find curious, but it maybe supports my earlier contention, is that the rate of readmission is actually higher than the rate of ED visits without hospitalization.

So I find that somewhat counterintuitive. If you think about admission ratios in general -- I'm running out of time. But I -- so in general I have fewer

concerns and problems with this then the other.

CO-CHAIR HALL: Okay, we'll vote on evidence.

MS. SHAHAB: 1a, evidence. 1 yes, 2 no. Time starts now.

We have all 22 votes. 1a evidence. 18 yes. 4 no.

CO-CHAIR HALL: Performance gap, opportunity to improve. Any specific questions or comments? None.

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

Just one more.

We have all the votes for 1b, performance gap. 7 high. 13 moderate. 0 low. 2 insufficient.

CO-CHAIR HALL: Priority. Comments above and beyond what we've heard. Not seeing any.

MS. SHAHAB: 1c, high priority. 1

high, 2 moderate, 3 low, 4 insufficient. Time begins now.

One. Just one more.

We have all 22 votes for 1c, high priority. 8 high. 14 moderate. 0 low. 0 insufficient.

CO-CHAIR HALL: Scientific, reliability, validity. Pam?

DR. ROBERTS: There was a comment that came up during the calls regarding interclass correlation testing not being done. And the developer did respond regarding why that was not appropriate.

And that it was more -- and they did use a split half test to assess measure reliability.

CO-CHAIR HALL: Wes?

DR. FIELDS: This would actually fit better in the next category. But I'll just briefly suggest that I think that there's a pretty decent chance that the readmissions probably are not always related to the reason

they were referred to home health services.

So I -- just again I think the scope of services available for most types of agencies probably is not adequate if the assumption is they can deal with all those chronic conditions that those patients typically have.

So that's the part of this that's not intuitive for me.

CO-CHAIR HALL: Thank you. Any other comments? Let's vote on reliability.

MS. SHAHAB: For 2a, reliability. 1 high, 2 moderate, 3 low, 4 insufficient. The time begins now.

We have all 22 votes for 2a, reliability. 2 high. 17 moderate. 3 low. 0 insufficient.

CO-CHAIR HALL: On validity we just heard Wes express a concern that bleeds over into validity. Are there any others? Above and beyond what we've discussed? I'm not seeing any.

MS. SHAHAB: For 2b, validity. 1 high, 2 moderate, 3 low, 4 insufficient. Time begins now.

We have all 22 responses for 2b, validity. 0 high. 18 moderate. 4 low. 0 insufficient.

DR. FIELDS: Chairpersons, point of personal privilege. Is there -- there's no filibuster privilege at NQF? I just wanted to make sure. Because I can keep you here all night if you know --

CO-CHAIR HALL: There actually is, but no one else has to stay.

CO-CHAIR KAPLAN: And the lines over are off, okay.

DR. FIELDS: Never mind, never mind.

CO-CHAIR KAPLAN: Feasibility. Specific comments? Feasibility, specific comments? Not seeing anything.

MS. SHAHAB: Voting for number 3, feasibility. 1 high, 2 moderate, 3 low, 4

insufficient. Time begins now.

We have all the votes for feasibility. 10 high. 10 moderate. 1 low. 1 insufficient.

CO-CHAIR HALL: And usability. Karen?

DR. JOYNT: I do worry a little bit with this one on unintended consequences. We already have a really hard time doing home health agencies to go into some of our most troubled neighborhoods.

And I think this is -- if we don't deal with some of that stuff, again, awaiting socioeconomic decisions from the other panel and all that stuff, I just worry a about it a little.

CO-CHAIR HALL: Good sentiment. Anyone else? Yes, Helen?

DR. CHEN: Just keep in mind we're already reporting this. So home health already reports this. This is the only difference is this is within 30 days.

So you know we're already reporting acute utilization.

CO-CHAIR HALL: I don't see any other comments.

MS. SHAHAB: For criteria number 4, usability and use. 1 high, 2 moderate, 3 low, 4 insufficient information. And the time begins now.

2 high. 15 moderate. 4 low and 0 insufficient information.

CO-CHAIR HALL: And overall. Any additional comments? None.

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

Just one more vote.

All votes are in for overall suitability for endorsement for measure 2380, Rehospitalization During the First 30 Days of Home Health. 16 yes and 6 no.

CO-CHAIR KAPLAN: I think in general the cognitive fatigue sort of is

sharpening up the gray area.

CO-CHAIR HALL: So we thank you for your efforts so far. And we look forward to reconvening these discussions tomorrow.

We ended a half hour late, which is somewhat on the back of our last developer. We pushed them and cut their discussions short. Hopefully they will not prosecute us for that.

So NQF colleagues any additional remarks?

MS. KHAN: Operator -- oh, did you want? Oh, we can end the call now. Thank you.

MR. AMIN: So beginning tomorrow we have 7:30 breakfast. We'll get started promptly at 8:00 o'clock tomorrow morning.

Hopefully we'll see some of you in a half an hour at Neo for dinner.

CO-CHAIR KAPLAN: I'd just like to add a great thank you. Because I know that this is arduous. You know it is really tough

to get through as many measures as you all did.

So to Bruce's thank you, I would like to add my thank you. Because I really do know how hard this is.

Thanks again.

CO-CHAIR HALL: Just leave everything. Leave the clickers and whatnot on the tables.

CO-CHAIR KAPLAN: Can they leave their personal stuff?

MS. KHAN: Yes, you can leave your clickers and your name tags at your tables. Please don't leave anything valuable in ths room. I cant' guarantee it's safety. But if you have papers, you can leave those.

(Whereupon, the above-entitled proceeding was concluded at 6:01 p.m.)

B

C

F

L

Q

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: All Cause Admissions and Readmissions Steering Committee

Before: NQF

Date: 05-05-14

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