NATIONAL QUALITY FORUM

+ + + + +

PULMONARY AND CRITICAL CARE STANDING COMMITTEE

+ + + + +

WEDNESDAY
MARCH 16, 2016

+ + + + +

The Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:00 a.m., Dale Bratzler and David Lang, Co-Chairs, presiding.

PRESENT:

DALE BRATZLER, DO, MPH, Chief Quality Officer, OU Physicians - Oklahoma University Health Sciences Center, Co-Chair

DAVID LANG, MD, Chair, Department of Allergy and Clinical Immunology, Respiratory Institute, Cleveland Clinic, Co-Chair

GERENE BAULDOFF, PhD, RN, FAAN, Professor of Clinical Nursing, The Ohio State University College of Nursing

KENNETH BENSON, Board Member, US COPD Coalition

CURTIS COLLINS, PharmD, MS, Specialty Pharmacist, Infectious Diseases, St. Joseph Mercy Health System

BRUNO DiGIOVINE, MD, MPH, Division Head, Pulmonary, Critical Care and Sleep Medicine, Henry Ford Medication Group

TODD DORMAN, MD, FCCM, Professor and Vice Chair for Critical Care Services, The Johns Hopkins Hospital

KIM ELLIOTT, Q, PhD, CPH, Administrator, Clinical Quality Management, AHCCCS

WILLIAM BRENDLE GLOMB, MD, FCCP, FAAP, Senior Medical Director, Superior HealthPlan

STEPHEN GROSSBART, PhD, Senior Vice President and Chief Quality Officer, Mercy Health
JAMES "MITCH" HARRIS, PhD, Director, Research & Statistics, Children's Hospital Association *
EDGAR JIMENEZ, MD, FCCM, Vice President, Critical Care Integration and eICU - System, Baylor Scott & White Health
ELLA KAZEROONI, MD, MS, Professor, Department of Radiology Cardiothoracic Radiology Division, University of Michigan Health System *
THOMAS LAMPONE, MD, Senior Medical Director, Florida Blue
RICHARD MURRAY, MD, Vice President and Deputy Chief Medical Officer, Merck and Co., Inc.
JAMES O'BRIEN, MD, MS, VP, Quality and Patient Safety, OhioHealth Riverside Methodist Hospital
PATRICIA J. OHTAKE, PT, PhD, Associate Professor, University of Buffalo
SUSAN POLLART, MD, Associate Tenured Professor; Senior Associate Dean for Faculty Affairs and Faculty Development, University of Virginia, Department of Family Medicine *
CRYSTAL RILEY, PharmD, MHA, MBA, CPHQ, CHPIT, Manager, Healthcare Policy, Baxter Healthcare
CHRISTINE SCHINDLER, PhD, RN, CPNP-AC/PC, WCC, Clinical Assistant Professor of Pediatrics, Marquette University College of Nursing
DAVID STOCKWELL, MD, MBA, Associate Professor of Pediatrics, Children's National Medical Center *
CHANA WEST, RN, MSN, Associate, Booz Allen Hamilton
DONALD YEALY, MD, FACEP, Professor and Chair, University of Pittsburgh-Department of Emergency Medicine

NQF STAFF:

ELISA MUNTHALI, Vice President, Quality Measurement
JANINE AMIRAULT, Project Analyst
POONAM BAL, Project Manager
SHACONNA GORHAM, Senior Project Manager
KAREN JOHNSON, MS, Senior Director *
ROBYN Y. NISHIMI, PhD, Consultant

ALSO PRESENT:

SUSANNAH BERNHEIM, MD, MHS, Yale University School of Medicine,
SHERYL M. DAVIES, MA, Stanford University *
FRANCOIS DE BRANTES, MS, MBA, Health Care Incentives Improvement Institute *
KAREN DORSEY, MD, PhD, Yale University School of Medicine
R. ADAMS DUDLEY, MD, MBA, Philip R. Lee Institute for Health Policy Studies *
MARK METERSKY, MD *
AMITA RASTOGI, MD, MHA, Health Care Incentives Improvement Institute *
MATT SCANLON, MD, Virtual PICU Systems, LLC *
JONATHAN SHAW, MD, MS, Stanford University *
CAROL STOCKS, PhD, MHSA, RN, Agency for Healthcare Research and Quality *

ANDREW WILSON, MPH, MA, Health Care Incentives Improvement Institute *

*** present by teleconference**

TABLE OF CONTENTS

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

8:05 a.m.

DR. NISHIMI: In the meantime, let me go over what happened on day one just to refresh the committee's memory.

We had one, two -- 11 measures yesterday. Two of the measures, the Minnesota Community Health measure and Dr. Kleinman's rate measure, not the appropriateness measure, rate measure were consensus not reached.

So, those two measures, well, all the measures go out for public comment. But, those two measures in particular we'll pay special attention to the comments and then you will review them and we vote because consensus was not reached.

And then, they'll either go forward to the NQF members for a vote as recommended, not recommended or, if you still don't reach consensus, they're marked consensus not reached.

For one measure, Larry Kleinman's ED Appropriate Asthma measure, that measure did not

advance, was not recommended. It still goes out for comment.

The committee may choose, you know, based on, if there's a groundswell of positive comments or something, may choose to take it up and discuss it again.

But, by and large, those measures tend not to come back for further committee discussion.

So, this is the report that Poonam spoke to yesterday, goes out for NQF member and public comment for 30 days, comes back.

The staff -- we all compile the comments, proposed response for you to look at. Some of them are just thanking you for your work. You know, some of them are agreeing with you.

And then, there will be some where you might need to discuss or the developer will need to address.

So, that was yesterday's work. And then, today's work, we'll do the balance.

As I indicated yesterday, we do hard

stop and if it looks like we can't even finish the measure by this stopping time, obviously, we won't start it.

If we don't finish all the measures today, excuse me, hay fever here, we don't finish all the measures today, then we have the post-comment, not post-comment, the post-meeting call set for next week for us to finish the balance of those.

We do have electronic software so, you know, we vote. It's probably easier than the little waving of the wands anyway. So, but that's what will happen if we don't finish.

So, I don't want you to worry about, you know, us running over. We're cognizant that folks have time and so we allow for that.

Anything else, Poonam, Shaconna?

Okay, we're waiting for AHRQ and we'll just have to launch if, obviously, she can't get a hold of her. She'll have to take over.

Operator, is she back, Sheryl and Carol?

OPERATOR: Carol has just joined.

DR. NISHIMI: Okay, great.

DR. STOCKS: Hi, this is Carol Stocks.

CO-CHAIR LANG: Very good. So, we will --

DR. STOCKS: My apologies.

CO-CHAIR LANG: Thank you.

We will proceed with discussion of three measures this morning which are population health measures pertaining to asthma, COPD and pneumonia, respectively.

The first, our leadoff hitter for this morning is 0283: Asthma in Younger Adults Admission Rate. Developer is the Agency for Healthcare Research and Quality.

Carol, would you like to briefly discuss the measure, briefly for about two minutes?

DR. STOCKS: Sure, thank you for giving me a couple minutes. I'm sorry, I was a little late. I was confused about the time.

Work that our program does involves

the maintenance of about 100 different indicators. And, by that, I mean the continual process of gathering evidence and getting input from clinicians and empirical testing. Most of the heavy lifting is done by our contractors.

And, Sheryl Davies is on the phone, she'll be available to answer more technical questions. She's at Stanford University, our primary contractor.

In addition to the validation and gathering evidence, we create software that can be applied to user's data so that they can more easily use the exact specifications for these types of measures.

The three measures being reviewed today we call the Prevention Quality Indicators, the PQIs and they've been developed for use with hospital administrative data or billing data type of information that's readily available and routinely created for every patient encounter.

The data have been created in electronic format for decades. So, they offer an

important reliable source of information on certain aspects of hospital care.

The PQI measures are a little bit different from other measures that we have and maybe others that are reviewed for the most part at NQF because they take advantage of inpatient hospital data, not to measure quality of inpatient care but to gain insight into health of the community.

So, we view it as a window into the community, one way to look at it. In some respects, what goes on in community hospitals could be seen as a microcosm of what's going on in the community.

So, the PQIs can be used as a screening tool to help flag potential healthcare quality problem areas at the population level and that need further investigation.

They're based on evidence that hospitalizations for ambulatory care sensitive conditions are potentially preventable given adequate outpatient care.

We believe that there is significant room for improvement in this area based on -- mainly based on the persistent disparities that we have seen for years now. And, most frequently, those disparities or higher rates are seen in populations that can be described as disadvantaged or vulnerable.

The idea behind these measures is not to close the hospital doors or, likewise, the PQIs are not modeled to measure whether appropriate decisions are being made about to admit or not to admit.

In this day and age, we -- I guess we assume that under the types of pressures faced by hospitals and physicians, we believe there's a pretty high probability that when individuals are admitted to the hospital, they need to be there.

So, the notion behind PQIs is that appropriate quality healthcare in the outpatient or community setting can prevent the need to be hospitalized.

Appropriate care is still centered on

primary care in the case of some of these pulmonary indicators, have specialty care is playing a very important role as well.

But, over the years, I think our understanding of what is being measured through use of ACSC hospitalization rates has become a little broader.

And, in essence, depending on how you use them, the PQIs are measuring the quality of the local healthcare system. And that encompasses many things to achieve the management of the population's health needs.

To put it another way, we're measuring whether the healthcare system is adequate or has the capacity for meeting population health needs.

We do include age and gender as risk adjustment factors when comparing counties or other regions.

And, in the more recent software that's being released this spring, we've added percent of persons living in poverty. However, depending on how you're using these measures, the

idea of risk adjustment goes a little bit against the concept of what the PQIs are measuring.

And, by that, I mean ideally a healthcare system should be able to meet population needs, whether those needs are greater or not.

So, I'm going to stop right there. I don't want to take up any more time.

Thank you very much for letting me introduce them.

CO-CHAIR LANG: Thank you, Carol.

We have Bruno and Susan with us also.

DR. POLLART: I am, yes.

CO-CHAIR LANG: Great, thank you, Susan.

DR. POLLART: Sure. Just a note, I'm on the webinar but I'm not seeing any slides and I'm not sure you're going to be able to see my vote.

So I have notes, I can participate without seeing the slides, but if my votes don't show up, I think it's an issue with the webinar.

MS. BAL: Susan, can you go ahead and just email Poonam. That's P-B-A-L@qualityforum.org with your vote so we can make sure we get those in?

Also, if you want to try to refresh your screen, that sometimes relieves that issue.

And then, also, before we start, I just wanted to mention that for all three of the AHRQ measures, Mitch Harris is conflicted so he will not be able to vote or discuss these measures.

Thank you.

DR. POLLART: All right. All right, I've refreshed my screen and nothing's changed. But, I will send my vote.

MS. BAL: Thank you.

DR. DIGIOVINE: Susan, are you going to start or do you want me to start?

DR. POLLART: Yes, I am absolutely -- I'm happy to start.

So, as was mentioned, this is an outcome measure type, looking at administrative

claims, the level of analysis is population through your county.

First question is, looking at the evidence and the question for the group is, there's some updated evidence provided related to aspects of hospitalization for pneumonia.

But I think our group agreed on our phone call that the underlying rationale for the measure remained reasonable and there was no evidence to repeat discussion or vote on evidence.

So, unless there's disagreement, can we move on to gaps in care?

DR. NISHIMI: Let's first -- yes -- let's just hear from the committee if --

CO-CHAIR LANG: Yes, and that -- so, the -- I don't -- before we get to that, just the only issue I have, and I have no one else to bring it up, so I'll bring up now.

Is just the age, we've gone running around now. I think we have three asthma measures with three different ages. It certainly

would be nice to have asthma measures all have the same age range.

DR. NISHIMI: Okay, so --

DR. POLLART: Age range here is for discharges for patients 18 and older.

DR. NISHIMI: Right. So --

CO-CHAIR LANG: It's 18 to 39, it's not 18 and over. At least that's the data that's in front of me.

DR. POLLART: The numerator statement I see says 18 and older. Is there new information?

CO-CHAIR LANG: Are we looking at 0283?

DR. POLLART: No, you're right, I'm sorry. I picked up the wrong one. I'm looking at -- that was 0279.

DR. NISHIMI: Okay, so --

DR. POLLART: Yes, 18 to 39.

DR. NISHIMI: We'll emphasize that, mention it again when we hit the specifications, but Bruno, was there anything on evidence you

needed to add?

DR. DIGIOVINE: No, I agree with Susan. I don't think we need to review it, but leave it to the committee.

DR. POLLART: All right.

DR. NISHIMI: Does anyone object? Okay.

DR. POLLART: All right. So, we're ready to move on in gap in care?

Our group discussed the question of whether there was opportunity for improvement and looked like there was data around gap in care, especially as it related to disparities.

And, it was looked at from 2009 to 2013. And, it appears there were gaps in care. And also, there's some gaps related to community income level.

So, our sense was there was opportunity for improvement.

CO-CHAIR LANG: Great.

DR. DIGIOVINE: Nothing to add.

CO-CHAIR LANG: Okay, proceed.

DR. POLLART: All right. Bruno, do you want to move on to the next section on --

DR. DIGIOVINE: I think we have to vote here, Susan. That's okay, thanks.

DR. POLLART: Oh, I'm sorry.

CO-CHAIR LANG: It's hard when you're not in the room.

DR. POLLART: Yes.

CO-CHAIR LANG: I think Jim had a question, though.

DR. O'BRIEN: Yes. The question I had I guess is for the developers.

I noticed in Table 1 that now we have an average rate of .28 which is now cut in half relative to where it was 2009.

At what point do the developers feel like this measure's going to be topped out?

DR. STOCKS: I think I'm going to let Sheryl Davies at Stanford answer that.

MS. DAVIES: Sure.

So, the measure, it's off, you're right. And, in fact, hospitalizations in general

have been decreasing over time.

One thing that we look at here, instead of just the overall population is we look at variation in hospitalization rates. Because we're looking at population healthcare, we do take into account issues such as disease prevalence or other community health factors that do impact the measures.

Those are part of the measure concept that we're measuring.

And so, as long as there are disparities that are visible, as there, you know, still are with low income populations.

And, our data and certainly, you know, in the literature, we continue to observe disparities with certain minority groups.

I would say that we're not there yet.

DR. O'BRIEN: Thank you.

And then, also, just to clarify, because I think there's a lot of -- it appears to be that some of the analysis for this measure crosses over 275.

It, I assume, but please clarify if not, that the risk adjusting, the reliability and the validity was all done on the entire population and not broken up by age groups for those two groups? Does that make -- does that question make sense and is that accurate?

MS. DAVIES: Okay. So, the reliability testing, you know, all the testing that's done within the measure testing forum and those are all done using the age groups that apply to that measure.

And so, in this case, it would just be the younger adult population.

DR. O'BRIEN: But, the disparities point that you made, it looks like that includes the two different age groups in the model. So, looking at the disparities, that's all asthma, all this group together, is that correct?

MS. DAVIES: Yes, yes, and the -- so in the disparities table that you received, we stratified that population.

DR. NISHIMI: Anything else? Ready to

vote?

CO-CHAIR LANG: Yes, voting on performance gap.

MS. AMIRAULT: Performance gap for measure 0283, your options are, one high, two moderate, three low and four insufficient.

MS. GORHAM: For the people on the phone, we're having a bit of a technical difficulty moment. Give us one minute. As you can still email or send your votes in via chat.

DR. KAZEROONI: Who do you want us to email the votes to?

MS. BAL: That's only for Susan to send her vote to P-B-A-L@qualityforum.org. But, everyone else who has access to the chat should vote that method.

Thank you.

DR. KAZEROONI: Thank you.

DR. POLLART: So, I tried to send mine through the chat, will you let me know if it arrived otherwise I'm teed up to send the email. But, that's just a slower process.

MS. BAL: We did not receive it.

DR. STOCKWELL: Did you all receive other votes via chat or should we vote again?

MS. BAL: We received all the chat votes, thank you.

And Susan, we received your email vote as well.

DR. NISHIMI: Okay. Go ahead and, Janine, reannounce.

MS. AMIRAULT: So, performance gap for 0283, one for high, two moderate, three low and four insufficient.

MS. BAL: Could everyone vote one more time? We're missing two votes. Thank you.

And, for the people online, that's only for people in person. Thank you.

MS. AMIRAULT: Okay. Four high, 17 moderate, one low and zero insufficient.

And based on the percentage, we can move forward.

CO-CHAIR LANG: Reliability?

DR. DIGIOVINE: So, you wanted me to

do this part?

DR. POLLART: Yes, that'd be great, Bruno.

DR. DIGIOVINE: No, that's fine, perfect.

So, in terms of reliability, the developer noted that they did signal-to-noise ratios as their test of reliability and have a signal-to-noise ratio of .75 or with risk adjustment to .74. So, it would certainly seemed like they had good evidence of high reliability.

DR. POLLART: Yes, and Bruno, I think you pointed out when we did the discussion that there some concern about low populations that the reliability doesn't meet the thresholds for counties with eligible populations under approximately 3,800 individuals.

And, the developer spoke to that in their phone call. Do we need to discuss that again?

CO-CHAIR LANG: Is there any further discussion?

All right, we'll proceed then to vote on reliability.

MS. AMIRAULT: Reliability for measure 0283, one high, two moderate, three low or four insufficient.

Again, reliability for 0283.

Five high, 16 moderate, one low and zero insufficient.

And, based on the percentage, we can move along.

CO-CHAIR LANG: Proceeding to validity.

DR. DIGIOVINE: All right. So, in terms of validity, again, the developer did a lot of nice sort of statistical testing around predictors of outcomes in the population.

I think here, the one thing that did come up in our discussion, probably the main thing that came up in the discussion is that as Carol said at the beginning, I mean, this is supposed to be a measure of health system quality.

And, in their model, access to care is not a significant factor. So, when they look at what predicts hospitalizations, it's prevalence of risk factors, health behaviors and socioeconomic status.

So, although these clearly are important measures for a community, I think there is some question around whether they are valid if the question is, are they -- if access to care is not a significant factor, whether you're really measuring health -- the quality of the healthcare.

I think that would be the one question I'd have for the developer. I think, otherwise, everything they've done shows that they've looked at the question of validity in pretty good depth.

Susan, if you have anything to add?

DR. POLLART: I do not, thanks.

MS. DAVIES: So, this is Sheryl Davies.

I can address -- oh, go ahead. Go ahead, Carol.

DR. STOCKS: I was just going to say that we didn't mean to imply that access to care was not a part of what's driving the hospitalization rates for ACSC conditions.

It's a number of things that are driving it. I think the concept is that the healthcare system should be able to meet the population needs and, clearly, access to care is a big part of that.

It's just that over time, I think our concept has expanded a little bit to recognize that putting -- having the right physicians in place and even having access to insurance may not be all of the answers.

CO-CHAIR LANG: Yes, I think just to amplify this, I just want to sharpen our understanding of the goal of the measure.

The rationale, as stated on the measure says that the measure -- it says with appropriate pharmaceutical and other outpatient management, the risks of hospitalizations decreased.

So, I think what we're -- I think what Bruno was raising is a point that, you know, what is the goal of the measure?

Because I think the rates correlate most closely with socioeconomic factors as opposed to, you know, access or aspects of healthcare.

So, if you could respond to that just so you can sharpen our understanding of the measure, I think that the committee would appreciate that.

DR. STOCKS: Well, the concept is that the healthcare system is not only the physician's office and whether the patient is able to fill the prescription. There's a lot of factors going on.

And, many of those factors can be mediated by a community or a system, ideally, it's a rather lofty goal, but it's one that we believe is very important.

And so, the system should be able to meet the community's healthcare needs.

Does that answer the question?

DR. SHAW: And, this is Jonathan Shaw, I'm a primary care physician. This is very near to my heart.

And, would just say that, currently, looking at things like insurance coverage and ratios of physicians in the area or access to care, but things like vaccination rates, and especially in the next few PQIs, pneumococcal and influenza vaccination rates, tobacco cessation programs which definitely are influenced by primary care as well as public health have a strong influence and correlate strongly with socioeconomic factors.

So, it may not be as narrow as the traditional access to care measures of how many physicians in the area, but those are also access to care issues.

CO-CHAIR LANG: Is there any further discussion regarding the validity of the measure?

And, then we will proceed with the vote on validity.

MS. AMIRAULT: Validity for 0283, one for high, two moderate, three low or four insufficient.

DR. NISHIMI: Ella, can you submit your vote or resubmit it? It's not showing up in chat.

MS. AMIRAULT: Zero high, 17 moderate, five low and zero insufficient.

And, based on the percentage, we can move forward.

DR. DIGIOVINE: So, in terms of --

DR. POLLART: I can talk about --

DR. DIGIOVINE: Oh, go ahead, Susan.

DR. POLLART: Yes, in terms of feasibility, this is a simple one. All the data elements are defined fields in electronic claims and so the measurement's readily available on administrative billing and claims data.

So, the committee felt that it was straightforward and feasible.

CO-CHAIR LANG: Is there any further discussion on feasibility? Bruno, were you going

to say something?

DR. DIGIOVINE: No, that's fine.

CO-CHAIR LANG: Okay.

We'll proceed to vote on feasibility.

MS. AMIRAULT: Feasibility for measure 0283, one high, two moderate, three low or four insufficient.

Nineteen high, two moderate, one low and zero insufficient.

And, based on the percentage, we can move forward.

CO-CHAIR LANG: Usability?

DR. DIGIOVINE: Usability -- I'll go ahead, Susan.

Again, our group thought it clearly very usable being used in public reporting by lots of different states and by CMS.

They were able to show improvement in rates of hospitalization between 2011 to 2013. No, sort of unexpected consequences.

We thought this was very usable.

CO-CHAIR LANG: Any further discussion

by the committee or questions for the developer?

Let's proceed to vote on usability, please?

MS. AMIRAULT: Usability and use for measure 0283, one for high, two moderate, three low or four insufficient.

Thirteen high, nine moderate, zero low and zero insufficient.

And, based on the percentage, we can move forward.

CO-CHAIR LANG: Thank you.

So now, we're going to vote on the overall measure or overall on the measure.

Is there any discussion?

DR. DIGIOVINE: Just remind me, at what point do we talk about other measures, harmonization has, you know, been brought up.

DR. NISHIMI: We have to be through all of them because of the time constraints. We might not get to that. We'll get to it on the phone call, but I have made a note in the record that you feel strongly that this whole age range

needs to be addressed.

CO-CHAIR LANG: On the -- yes, we -- so then, we will proceed to a vote.

MS. AMIRAULT: For overall suitability for measure 0283, one for yes or two for no.

DR. NISHIMI: Ella, can you submit your vote again?

MS. AMIRAULT: 21 yes and one no.

DR. NISHIMI: Okay, so the measure will be reflected as recommended in the report.

CO-CHAIR LANG: Thank you.

So now, we will proceed to measure 0275: Chronic Obstructive Pulmonary Disease or Asthma in Older Adults Admission Rate.

Carol, would you -- is there -- do you wish to make additional comments regarding this measure before we proceed?

DR. STOCKS: No, I don't think so.

CO-CHAIR LANG: All right.

The measure will be discussed by Jim O'Brien and Ken Benson.

Take it away, gentlemen.

MS. BAL: Sorry, just one second.

Cathy, can you please work with Susan so she can get the webinar working for her? Thank you.

Sorry, go ahead, I just wanted to make sure.

MR. BENSON: Okay, this is a previously endorsed outcome measure undergoing annual maintenance review.

It was first released in 2007. New information has been provided which is a review of material and that new evidence pretty much supports the rationale for this.

We see a gap in this in that COPD is the third leading cause of death in the United States and of the leading causes of death, it's the only one that continues to rise.

Performance suggests the gap still exists.

And, any comments at that time?

DR. O'BRIEN: This measure is analogous to the one we just discussed. It's

just the older age group and then COPD winds up being added in. So, you're going to hear a lot of the same analysis and everything else as we go through it.

CO-CHAIR LANG: May I ask about the rationale of including COPD combined with asthma as opposed to making, you know, separate measures since they're different conditions, as we all know?

DR. O'BRIEN: Do you want the developers to comment?

CO-CHAIR LANG: That's a question for the developer.

DR. STOCKS: Jonathan, would you like to answer that?

DR. SHAW: Sure.

There was eight on the expert panel consensus in 2009 and they convened on all the PQIs and, recognizing that there's a lot of diagnostic uncertainty in older adults and uncertainty in coding between asthma and COPD.

The management at the admission level

was often identical or very similar. So, that's the rationale there. Definitely some uncertainty.

But, you know, the patterns are very different among the age 40 and above and goes up strongly with age.

CO-CHAIR LANG: Additional discussion?

Yeah, I would just ask, I guess, I don't know if it's in the data anywhere, but whether there's any data looking at the rates of hospitalization for patients who would fit asthma and those for whom it would fit COPD to see whether that overlap that we know does exist clinically actually is reflected in your data?

MS. DAVIES: So, this is Sheryl Davies.

We don't report that here. In our data, you know, we do see a mix of diagnosis codes. You know, we can't know exactly, you know, what underlies those diagnosis codes, but we do see a mix of diagnosis codes with predominance, and I'll have to look up the number

if the committee is interested, the predominant diagnosis is COPD in this age group.

CO-CHAIR LANG: Additional discussion? Yes?

DR. LAMPONE: I just had a comment, and this would probably cross over to the prior measure that we just reviewed. And, this is for the developer.

Have you seen communities based on the data develop programs or what impacts they're having in the community to help drive this data? I think it's important data to know at the geographic area and the drivers of the exacerbation of COPD and asthma and hospitalization rates.

But, how are the communities and providers in those communities using this data so we can capitalize on the information we have?

MS. DAVIES: So, we don't -- oh, go ahead, Carol.

DR. STOCKS: No, you go ahead.

MS. DAVIES: Yeah, so, you know, we

don't, especially, you know, AHRQ being part of the federal government, collects systematic data about the use of measures.

So, what we do observe is that these measures are used quite frequently within public health programs and within state reporting.

Within the research, they are used as outcomes measures in this case and with the prior measure. We do see them being used in research and otherwise as they -- in conjunction with other measures such as the prevalence of COPD or asthma and certainly issues such as air quality or, you know, within the research, environmental pollution or job-related pollution and smoking rates.

So, I mean, I'm not sure if I'm answering your question directly, partially because we don't have a systematic way of monitoring use.

But, in general, these measures are very widely used as screening tools for, you know, looking at utilization of population

health.

DR. LAMPONE: Fair enough. Okay.

DR. GROSSBART: With reference to that question, the organization I'm affiliated with has a Medicare Shared Savings Program. So, this is one of the ACO measures that CMS has mandated or is analyzing for each of those covered lives within our and everyone else's Medicare Shared Savings Program.

And, we've hired a hundred care managers who do, among other things, try to manage patients with COPD and keep them out of the hospital. So, this is having an impact, you know, in the trenches here.

CO-CHAIR LANG: Is there further discussion regarding evidence?

Then we will -- oh, Steve, are you okay?

Then we will proceed to the vote.

MS. AMIRAULT: Evidence for 0275, one for yes, two for no. Again, evidence for measure 0275.

Okay, 22 yes and zero no.

So, based on the percentage, we can move forward.

CO-CHAIR LANG: Risk -- gap?

MR. BENSON: In terms of gaps? You know, the -- we think that there is a pretty significant gap in the new evidence. What's introduced supports that.

There are some questions in there of how these gaps arrived. Some apparently contradictory information on the rate of hospitalization for blacks relative to whites.

And then, another items that says there's little variation between blacks and whites. A little confused by that, but overall, the performance data does suggest a gap.

CO-CHAIR LANG: Is there additional discussion or questions for the developer?

Thank you, then we'll proceed to vote on performance gap.

MS. AMIRAULT: Performance gap for measure 0275, one for high, two moderate, three

low or four insufficient. Again, performance gap for 0275.

Six high, 16 moderate, zero low and zero insufficient.

Based on the percentages, we can move on.

CO-CHAIR LANG: Reliability?

DR. O'BRIEN: So, the reliability testing was done similarly to the previous measure, it was signal-to-noise ratio.

The performance was actually better in the deciles than it was for the previous measure with just one -- the smallest that the decile with the counties with the lowest population just came below the specified threshold.

I did have a question for the developers. There's mention of two different risk adjusting models, one that's just age and gender and then the other one that talks about including some measure of socioeconomic status. It mentions that model's available.

Which of the two is actually in use?

DR. STOCKS: Typically, it's the age and gender or no risk adjustment, depending on, you know, what, you know, the users are most interested in measuring.

So, as you know, there's a lot of issues around measuring or risk adjusting for socioeconomic status and indicates of these measures, particularly when you're trying to identify disparities, it can mask disparities.

But, some users do use the poverty adjustment which, you know, is available on the software as an option to compare like communities.

DR. O'BRIEN: So, I guess for the NQF staff, is this committee then voting to improve all different ways that this might be used? So, non-risk adjusted, risk adjusted with socioeconomic status, without socioeconomic status in the model?

DR. NISHIMI: Yes.

CO-CHAIR LANG: Is there further discussion or questions for the developer

concerning reliability?

Then, we will proceed to vote.

MS. AMIRAULT: Reliability for measure 0275, one high, two moderate, three low or four insufficient.

Three high, 19 moderate, zero low and zero insufficient.

Based on the percentage, you can move on.

DR. O'BRIEN: Regarding validity, again, same as the previous measure, two different pieces.

One's face validity in which there's an expert panel convened. This group wound up supporting the measure with some concern which is based on their second level of support.

Some of their concerns were similar to what we talked about.

And, there was also the same empirical validity testing looking at principle components that, again, found an association with health behaviors and socioeconomic status.

CO-CHAIR LANG: I have a question for the developers.

There's a notation in the proposed measure that patients with severe chronic respiratory disease had been excluded because COPD asthma differs in the subgroup from patients with COPD asthma lung.

Could you elaborate on that, please?

DR. STOCKS: Jonathan, do you want to take this as a clinician?

DR. SHAW: Yeah, I think it's in reference -- if you looked at the exclusions, I'm trying to pull them up in front of me, but that's on the Excel appendix.

It's congenital lung diseases, cystic fibrosis, rare, but, you know, definitely of significant prevalence conditions that we didn't feel captured the community health measure of being exceptions.

And, likely be focused and more specialized hospitals, urban regions.

DR. STOCKS: And, we do align our

measures. These specifications, the exclusion, actually originally arose within the pediatric asthma measure during an expert consensus process and -- during our 2009 consensus process.

The panel has recommended that we extend the exclusion, although it doesn't actually exclude that many numerator cases.

They recommended that we include this exclusion because these patients received different types of care and their hospitalizations probably reflect different factors.

CO-CHAIR LANG: All right, fair enough.

Are there any other questions for developers? Any other comments?

All right, seeing no further questions, comments, we'll proceed to the vote.

MR. AMIRAULT: Validity for measure 0275, one for high, two moderate, three low or four insufficient.

Two high, 18 moderate, two low and

zero insufficient.

Based on the percentage, we can move on.

CO-CHAIR LANG: Feasibility?

DR. O'BRIEN: This is based on administrative billing and claims data as well as U.S. Census data and the software is readily available from AHRQ.

CO-CHAIR LANG: Any further discussion questions?

Please, Bruno?

DR. DIGIOVINE: I don't want to go back to validity, but just as you made your point, Dave, I looked at the exclusion criteria and they are all around pediatric diagnoses.

And so, I do have some concerns that they're not excluding adult diseases that would fit into what I think they're trying -- so, they have congenital bronchiectasis, but they don't have bronchiectasis.

So, it would strike me that there probably needs to be some retooling of the

exclusionary criteria to include adults.

CO-CHAIR LANG: Yeah, I haven't been able to pull it up on my screen, either. But, I recall, you know, wondering about conditions like Churg-Strauss and allergic bronchopulmonary aspergillosis and how these other conditions were handled.

Would the developers wish to comment further?

MS. DAVIES: Well, so --

DR. STOCKS: Is that -- go ahead, Sheryl.

MS. DAVIES: Yes, those particular conditions were not brought up by our expert panel in 2009, but, like Carol said, we can certainly revisit those conditions and look at their frequency within the numerator population.

DR. NISHIMI: Okay, so in the comments to the developers and the record of the report, we'll make that indication.

CO-CHAIR LANG: So, we're on feasibility.

So then, where are we with respect to that issue, then? We'll revisit this?

DR. NISHIMI: Well, no. You can revote based on that, otherwise, it'll just be a recommendation to the developer that they take a look at those adult exclusions.

CO-CHAIR LANG: That's fine.

DR. NISHIMI: If you --

CO-CHAIR LANG: Is that okay with the --

DR. NISHIMI: Is that okay with --

CO-CHAIR LANG: Is there anybody wish to revote?

DR. NISHIMI: Okay, so then --

CO-CHAIR LANG: All right. So, then we're --

DR. NISHIMI: -- it'll just be a comment.

CO-CHAIR LANG: Okay, thank you for that clarification.

So then, we'll proceed to vote on feasibility.

MS. AMIRAULT: Feasibility for measure 0275, one for high, two moderate, three low or four insufficient.

14 high, 8 moderate, zero low and zero insufficient.

Based on the percentage, we can move forward.

CO-CHAIR LANG: Excellent. We're proceeding to usability.

DR. O'BRIEN: So, this is in use as mentioned in some of the discussion in a wide variety of ranges including the Medicare Shared Savings Program down in Cincinnati.

There has been improvement from 2011 to 2013 in this rate. The developers suggest over 104,000 fewer hospitalizations. It's not clear what's triggering this improvement, whether it's intentional or it might be also changes in qualifications for hospitalizations or reduction in hospitalizations overall.

CO-CHAIR LANG: Is there further discussion or questions?

Please, Dale?

CO-CHAIR BRATZLER: So, I had, you know, on the previous measure, we talked a little bit about how the measure was being used and, you know, I think it makes sense in large shared savings programs and potentially in community care organizations.

You know, Oregon's been using some of these because they're kind of population-based performance metrics.

I did have a little concern about what's the minimum sample size that this measure works at?

And, the reason I raise this question, I had to go back and look at our QRUR report. But, now, CMS is using PQI 5, so this is 0275 for COPD and asthma at the level of the individual practice as a part of the QRUR, which concerns me a bit that that's really not population level evaluation and it's actually being used as part of the value modifier.

So, I question whether that's an

appropriate use or whether there's the minimum sample size that should be, you know, that you can actually apply this performance metric to?

I'm not sure if my question is completely clear, but I do have concerns about some of the use of the metric.

DR. STOCKS: Can we address that as developers?

DR. NISHIMI: If you can hold off, we have a committee follow-up first.

DR. STOCKS: Okay.

DR. MURRAY: I guess what I'm trying to ask was were we asking for a remedy, you know, like a notification on when you're using this metric, it cannot be used in a size smaller than X. Is that what we're asking?

Because that might apply to some of other metrics that we're looking at, too.

CO-CHAIR BRATZLER: To me, I think it goes to the validity of the measure. The validity may be very good at the population level, but how small does the population get

where the measure is not valid or, you know, I don't know if I'm, again, asking the right question, but I'm interested to see what the developers say.

But, I do have concerns because I went back -- I had to go back real quickly and search the CMS methodology, but they explicitly state that they're using PQI 05 at the practice level as a part of the value modifier now.

So, I'm concerned about the sample size or the population size that this is being applied to.

DR. NISHIMI: Go ahead, AHRQ.

DR. STOCKS: So, the measure that CMS is using, as we know, and the materials with the footnote there, is an adaptation of the PQI. So, the measure before you today is the population health measure.

We provide the information on what CMS uses because it's one of the most visible ways that admission rates themselves are being used.

And, as you heard before, you know,

it's a, you know, within those programs, we can see, you know, the action being taken.

However, the testing that we provide today and all the information is based on the population measure.

We cannot speak to, you know, the minimum sample size with inference. It's a CMS program.

When you change the denominator population, you change rates, you change the distribution across the measured entities. And, those all, all of those factors will actually impact the signal-to-noise ratio.

So, you know, we, as AHRQ, are, and we, as the measure developer under contract with AHRQ, are bringing before you the population health-based measure and we can speak to what we observed.

And that is that, for this particular measure, there are some reliability concerns for small counties. And for those counties, we recommend using the smoothed rates. And, the

smoothed rates actually account, I guess, you'd call it a reliability adjustment.

So, it will account for variation and reliability for those very small counties.

CO-CHAIR LANG: Steve, did you want to say something?

DR. GROSSBART: Just a comment, is it a measurement developer's responsibility for the measure being used liberally and CMS does have a tendency to do that quite a bit among others.

DR. NISHIMI: No, it's -- we don't place the implementation burden on the developer. And, yes, CMS does use this measure and other measures towards its own ends.

DR. O'BRIEN: Although, I would offer that that probably would be considered an unintended consequence that should be identified.

CO-CHAIR BRATZLER: Yes, so I agree. And, I just -- when I -- because I kind of, as we were having the conversation, went and searched the CMS website real quick and they explicitly identified PQI 05 as the methodology behind the

metric that they put in the QRUR that they hold individual practices accountable for.

So, I understand that's not AHRQ's fault and it's not the fault of this measure. I understand that. But, I do -- I think, you know, James's point is a good one, that it is an unintended consequence.

And, if you to the CMS website, it says this is the PQI 05 measure that's being used to hold a practice level, you know, accountable for admission rates for asthma COPD.

CO-CHAIR LANG: Bruno?

DR. DIGIOVINE: Yes, I think you just pointed out that it's not the role of the developer and it's not the -- I guess, it's not AHRQ's role. I guess the question is, is it NQF's role? Is it our role, as a committee, or this organization's role to say, this measure is only approved for measurement at this level?

DR. NISHIMI: And, that's what we do.

DR. DIGIOVINE: Okay.

DR. NISHIMI: You're voting on it and

that's what the report clearly indicates, the level of analysis population.

CO-CHAIR LANG: Is there additional discussion, addition questions for the developers?

As I see no further questions, discussion, we'll proceed to a vote on usability.

MS. AMIRAULT: Usability and use for measure 0275, one for high, two moderate, three low or four insufficient.

DR. NISHIMI: David, can you resend your vote? David?

DR. STOCKWELL: Yes, I'm here. Sorry, I'm having access problems, just a sec.

DR. NISHIMI: If you want to email it to P as in Peter, B as in boy, A as in --

DR. STOCKWELL: Yes, I'll do that. I've got her email.

DR. NISHIMI: Okay.

DR. STOCKWELL: Thank you.

Okay, just sent.

DR. NISHIMI: Okay.

MS. AMIRAULT: Would everyone mind just pointing just pointing one more time? Thank you.

Three high, 15 moderate, four low and zero insufficient.

Based on the percentage, you can move on.

CO-CHAIR LANG: All right, so now, we're going to proceed to vote on the overall measure.

Discussion? Questions for the developer?

Seeing no further questions, discussion, we'll vote on the overall suitability for endorsement of the measure.

MS. AMIRAULT: For overall suitability for Measure 0275, one for yes, two for no.

Eighteen yes, four no.

So, based on the percentage, this will be recommended.

CO-CHAIR LANG: We're now moving on to our third measure, our third population health

measure, 0279: Bacterial Pneumonia Admission Rate.

Carol, would you like to -- or Sheryl, would you like to make a few comments about this briefly before we proceed?

DR. STOCKS: No, I don't think so. Thank you.

CO-CHAIR LANG: Very well, the measure will be reviewed by Susan Pollart and Rich Murray.

Evidence?

DR. MURRAY: This is NQF Measure 0279, Bacterial Pneumonia Admission Rate, PQI 11.

And, this is an outcome measure. The data source is from claims, and the level of analysis is at the population level -- county or city.

I don't know if this one is also being used by CMS at the practice level, but if it is, we should consider that same discussion that Dale just brought up.

The evidence here, you know, includes

providing the rationale that supports the relationship of the health outcome, in this case, admission for pneumonia.

Two processes are structures and care and the developer says that their rationale for the measure is that access to high quality care, early intervention and appropriate treatment including pharmaceutical treatment will minimize the likelihood of mild respiratory conditions progressing to pneumonia reducing the likelihood of hospitalizations.

There is no requirement for a literature review, but they did review some additional literature.

And, the new question for the committee is that, although the developer provides updated evidence related to aspects of hospitalization for pneumonia, does the committee agree that the underlying rationale for the measure remains reasonable? And, there is no need for the discussion or vote of the evidence.

CO-CHAIR LANG: Yes, so the underlying

rationale for the outcome measure hasn't changed since the last endorsement review. So, unless there is an objection from anyone on the committee, there's not any need for any vote.

DR. MURRAY: And the numerator and denominator have not changed.

CO-CHAIR LANG: Okay. Then, we will proceed without voting to performance gap.

DR. POLLART: I can step in if you want me, Richard.

MR. MURRAY: Sure.

DR. POLLART: Looking at gaps, there was evidence of disparity, particularly on older men both socioeconomic status and in rural areas. So, the group felt that there were gaps in care, warranting it as a national performance measure.

DR. MURRAY: And, there were no disparities related to race, right? Just to male patients 65 and over and patients with the lowest income and also patients living in rural locations.

DR. POLLART: Yes, that's what was

reported.

DR. MURRAY: Which further supports the variation in care.

CO-CHAIR LANG: Is there further discussion?

Bruno, yes?

DR. DIGIOVINE: A couple of questions.

One is, the disparities that are reported, are those -- is there data that the developers provided that I'm not seeing right here, or is that just based on a literature review or something else?

And then, I don't know if this is here or in validity. I guess there's still this -- this one, for me, is much harder to understand how much healthcare can intervene in preventing a pneumonia admission.

It just strikes me as not -- the face validity of that, to me, is difficult to understand.

DR. MURRAY: Although the data that they provide in the performance gap of the

overall document, you know, shows sort of progressive decline in the observed rate in the period of 2009 to 2013.

DR. DIGIOVINE: So, does the developer have some health system improvement that they think correlates with that decrease in pneumonia rates?

DR. SHAW: In the evidence -- yes, this is Jonathan Shaw.

There's very strong evidence that pneumococcal vaccinations as well as flu vaccination reduces hospitalization rates in a given area.

DR. MURRAY: And, while we're looking at those data, could we ask the developer to clarify in the table that's under the performance gap section, there's a -- I don't know if it's typo or not, but under the very last row, year 2013, the standard deviation is 2.43, a striking drop from the previous four years. Do you see that?

MS. DAVIES: I can address -- yes, I

can address that.

And, it is actually, we provided some updated tables in the distribution of the PQI is an interesting distribution. I won't go into the details of it.

But, as you know, distributions, particularly outliers can impact the standard deviations.

We provided updated tables taking out those outliers just for clarification.

DR. MURRAY: Okay, thank you.

DR. YEALY: I had one question about the performance gap and the disparities.

Is it -- it looks like male sex and age over 65 is a big drivers, but those are big drivers of mortality prediction when you develop acute pneumonia, at least if you use the pneumonia severity index.

So, this is kind of circular; of course, they would drive this particular event because, in fact, men die more often with the same level of pneumonia than a woman does and

people over 60, the base way you get points in the pneumonia severity index is your age if it's over 50.

So, I wonder about half of the performance opportunity actually being exactly what we're looking for? It's you want those folks admitted. It's not -- or at least more proportionally admitted.

I'm not sure I'm seeing the problem exactly here.

DR. SHAW: These are not -- these are potentially preventable, not unnecessary admissions. We're not saying that these aren't necessary.

So, in terms of the potentially preventable, I'd go back to the effective high rates of pneumococcal vaccinations, early outpatient detection should still influence the population level rates regardless of the fact that, yes, when the elder male or elder pneumonia patient presents, they should be admitted.

DR. O'BRIEN: Is there a population-

based measure of pneumococcal vaccination and flu vaccination rates already available?

DR. NISHIMI: NQF has endorsed pneumococcal vaccination influence and vaccination measures, yes.

DR. MURRAY: It would be interesting to compare the change in that over time with the change in this.

DR. DIGIOVINE: So, also, earlier, we heard about outreach for trying to prevent COPD admissions and asthma admissions. Is there anybody reporting that they're doing outreach to try to prevent pneumonia admissions?

DR. SHAW: So, in the evidence -- so, this is the measure evidence form under the -- there is a reference to a program within the VA that showed success and it was focused on vaccination. And, it showed significant decreases after that, that after -- it's on page 7 of the evidence form.

CO-CHAIR LANG: Okay, we have Crystal and then Todd.

I'm sorry, Don? Don? Oh, you didn't put your -- okay.

Todd?

DR. DORMAN: So, I guess it's less a question and just a comment.

I think what's confusing me a little bit here is the statement that appropriate pharmaceutical treatment with the degree to which the likelihood of milder respiratory conditions progressing to pneumonia, and I think that's the question.

Where is the evidence that there's some intervention for milder respiratory conditions? What's being presented is vaccination, which doesn't fit what that sentence says and I think that's the disconnect for me.

DR. SHAW: This is Jonathan, again.

There is a large truck of pneumonias which can be treated outpatient if caught early, maybe not in the older population per se, but, you know, just looking at all population, all ages.

And so, you know, with early access to care, many of these patients will be treated for "walking pneumonia" and avoid hospitalization.

CO-CHAIR LANG: I also have a question for the developer.

The measure is entitled Bacterial Pneumonia Admission Rate, yet what you're tracking is discharges.

So, I'm just wondering, you're using the discharge diagnosis as a proxy for admissions, so I'm wondering whether you can reconcile this for us? Is there any data you have that implies that it's apples and apples?

DR. STOCKS: In the type of data that we're using, the records are created at the time of discharge and the principle diagnosis is the one that is adjudged to be the major cause of that admissions. Sometimes that, the first impression or the admitting diagnosis is not what it turns out to be.

We believe the discharge diagnosis -- and, actually, these records aren't even created

until the time of discharge.

Do you have anything to add, Sheryl?

CO-CHAIR LANG: I just had a clarification --- so, are you using discharge claims data, or are you doing chart review discharge diagnoses?

DR. STOCKS: It's discharge claims, not claims, but billing data created by the hospitals and, of course, those billing records are based on the coders looking at the medical records.

CO-CHAIR LANG: Okay. So, it's from the hospitals themselves, not the payers or CMS?

DR. STOCKS: That's correct.

CO-CHAIR LANG: Okay.

DR. MURRAY: So, this measure underestimates the actual admission rate by some amount related to say, death or transfer, is that right?

DR. STOCKS: No, we would include cases. So, transfers are only excluded in one direction to avoid double counting of

hospitalizations.

DR. MURRAY: Okay.

DR. STOCKS: And, deaths would be included. It's just when the diagnosis -- so, in this case and the case of a principle diagnosis here, you know, that principle diagnosis is the diagnosis that, quote, occasions, principally occasions, the admission.

And so, you know, we don't observe a huge, you know, difference between, you know, admission and discharge, in this particular case.

You would only observe that, for instance, if somebody thought somebody had pneumonia and then at the end of the hospitalization, they figured out that it wasn't really pneumonia, then it would be what's called a rule-out diagnosis.

Rule-out diagnoses are not codable as a principle diagnosis or a secondary diagnosis. So, we're actually capturing, you know, what diseases that are recorded by the physicians.

CO-CHAIR LANG: Thank you.

Are there any additional questions for the developers or comments? We're on performance gap, and we're going to vote on performance gap.

MS. AMIRAULT: Performance gap for Measure 0279, one for high, two moderate, three low or four insufficient.

One high, 11 moderate, 10 low and zero insufficient.

And, this is grey zone.

CO-CHAIR LANG: Okay. We will move forward with -- we have to discuss reliability.

DR. MURRAY: So, on reliability, the developer indicates that there were some changes made to the measure specifications since the prior endorsement review.

They added several codes including staphylococcal pneumonia, methicillin-susceptible staph pneumonia.

So, a number of codes were added and some additional exclusions. They did repeat the reliability testing and they also, let's see, they did this at the measure score level using

more current data.

And, they did a risk adjustment model as well using data from the Healthcare Cost and Utilization Project state and patient database in 40 states representing 89 percent of the country.

And, as the summary of their testing, they reported a reliability testing, not at the individual element level, but at the measure level with a signal-to-noise ratio of .97 which they say is very high.

And, when they added social demographics statistics to the risk adjustment model, the signal-to-noise ratio was about the same at .96.

So, the question is, are we confident that this is sufficiently reliable to be useful to look at changes in essentially quality of care?

CO-CHAIR LANG: Susan, is there any additional comments you wish to make?

DR. POLLART: Yes, no, I think that summarized it.

CO-CHAIR LANG: Great.

DR. POLLART: I think the other question was were the appropriate codes included? But, as was mentioned, a number were added that I think are sufficient.

CO-CHAIR LANG: Additional comments, other members of the committee, questions to the developers?

Seeing none, we will proceed to vote on reliability.

MS. AMIRAULT: Reliability for Measure 0279, one being high, two moderate, three low or four insufficient.

Six high, 13 moderate, three low and zero insufficient.

And, based on the percentage, you can move along.

CO-CHAIR LANG: Validity? Susan? Rich?

DR. POLLART: Yes, I can start with that.

So, the validity testing level was at

that measure score and face validity only was tested.

It was tested within -- from four clinical expert panels involving 73 panelists and was convened in 2008 to 2009. The panels indicated the measure was useful.

And, I think our group asked whether a panel convened in 2008 to 2009 was -- if those recommendations were still applicable? But the panels did indicate it was useful.

CO-CHAIR LANG: Rich, any additional --

DR. MURRAY: I guess threats to validity, there are, you know, issues of exclusions. There are issues of miscoding. There's the sickle cell disease.

You know, so the question is, is are the exclusions consistent with the evidence and are they sufficient for us to have confidence in the validity in terms of these? Those are the main threats, I think.

DR. POLLART: Yes, the point was made

still about --

DR. MURRAY: And coding.

DR. POLLART: Yes, there were about a little over 69,000 discharges excluded because of the diagnosis in the immunocompromised state which --

DR. MURRAY: Right.

DR. POLLART: -- if you replace those folks, you'd increase your numerator by over ten percent. So, are they appropriately or inappropriately excluded from the measure when you remove the immunocompromised state?

DR. MURRAY: So, the ability to be confident about the presence or absence of immunocompromised state would sort of -- that would probably increase the uncertainty of the measure.

CO-CHAIR LANG: Bruno?

DR. DIGIOVINE: Yes, I had two questions.

One was either for the developer or for others if they understand it. I don't

understand the way the sense of validity is reported by the panel.

There's four levels of support with different median scores and I'm not sure I understand what all of that means.

And, the second is, this is now the third measure, the other two statistical tests were done to assess validity, and I'm wondering why that wasn't done specifically for this measure?

DR. STOCKS: Sure, so when we do the face validity testing, we asked several questions about usability and the different aspects of the measure, very similar to what the NQF panel is doing when we ask, you know, questions of reliability.

In this case, we asked them to rate different aspects that would contribute to the face validity of the measure and the usability.

In that particular study, we were looking at different levels and different applications of the measure. And so, you'll see

the different ratings there. They just reflect the purpose of that study.

In the case of the empirical analysis, the study that we used to assess the empirical validity of the measure was focused on chronic disease measures. And, did not actually include the acute PQI, so that's why they're not included here.

DR. DIGIOVINE: So, just on the face validity piece, you're saying you asked the panel, do they fully support the use and over -- on a 9-point Likert scale? And, their answer was somewhere between seven and nine?

But, then you asked them, again, do they have general support with some concerns, and they had the same degree of agreement?

DR. STOCKS: So, those -- so, we asked them about the overall usefulness of the measure or a different application. In this case, we're reporting on the population health application.

And, for that, in comparison between counties, and for that, they rated on a Likert

scale and we used the RAND Appropriateness Method, the adaptation of that method to then rate the measure. We convened two -- or then categorized the support.

We convened two panels here, I think that's what you're probably noticing. We're testing kind of a different approach because the nominal panel technique, which is typically used as, you know, a group of, you know, 8 to 15, you know, individuals.

There are some chances for bias, depending on the particular folks that you have on that panel. And so, we were testing a different approach where we would combine the nominal group with a delta group.

So, we also convened a much larger group that did not meet in person but they received feedback from the other group, and that study is also published, and it's referenced there.

So, that's the reason you see two ratings there, it's because we had two separate

panels that interacted with each other but they were considered separately.

CO-CHAIR LANG: Is there additional discussion, questions for the developers?

Yes, Don?

DR. YEALY: You know, I noticed that the measure is called a measure but focused on bacterial pneumonia.

But, it actually looks like aside from the coding, what you're really assessing is community acquired pneumonia.

I understand that a ton of these may be coded as bacterial pneumonia. Has there been any validity check on the frequency in which this truly is bacterial?

In other words, there was some microbiologic evidence of it being bacterial?

You know, in my world and most published world, a third would be about the most that you had any hard data on.

And, I suspect that those comments that you got in 2008 were about community

acquired pneumonia, not specifically about bacterial pneumonia. We throw the terms around interchangeably, but they're not exactly the same things.

So, that's my validity question, not are the data sets big enough to be examined again and again. Is this really bacterial pneumonia we're talking about?

DR. STOCKS: That's a fair concern. And, I think we could consider changing the measure. You're absolutely right that this is most reflects community acquired pneumonia.

The name of this measure, you know, a legacy name going back to the early conceptualization of ambulatory care sensitive condition. The idea was to -- that, you know, many community acquired pneumonias are bacterial.

You'll notice in the code list that we do include unspecified pneumonias and this captures exactly what you're speaking about is that, you know, most pneumonias don't have, you know, either it's not done or we cannot actually

obtain a culture -- a positive culture bacterially, but they're treated empirically.

And so, that is something we can certainly consider is, you know, a change to the name to reflect community acquired pneumonia.

DR. MURRAY: Just as a follow up question. If you're going to take bacterial pneumonia out of the name, does that mean that we're comfortable having somebody with a viral respiratory tract infection with infiltrates that, of course, cultures nothing bacterial and that that would now be in the numerator?

DR. STOCKS: I would like to clarify that we don't actually include codes for virals. So, if the physician does make it as a viral pneumonia, it would not be included in here.

DR. MURRAY: Okay.

DR. STOCKS: That's, you know, differs from, you know, some of the CMS measures of, you know, pneumonia mortality, et cetera.

So, if it's specified as such, then it would not be.

DR. MURRAY: It would not be? Okay.

DR. POLLART: The committee said the added codes are discussed adding staph, methicillin-resistant, MRSA staph. But, what would codes -- the previous codes?

DR. STOCKS: I'm sorry, I'm not sure I understand the question.

DR. POLLART: Yes, that's all right. I'm just kind of -- you talk about added codes, the original codes were ICD-9-CM or ICD-10 diagnosis code for bacterial pneumonia.

DR. STOCKS: Yes, all the codes that are included are included in the technical specification that's been provided in the Excel spreadsheet. That has all the codes. These are just new codes that were added, usually because they've been introduced into the coding system.

CO-CHAIR LANG: Okay, further discussion?

All right, we will proceed to vote on validity.

MS. AMIRAULT: The highest eligibility

for validity for 0279 is moderate, so the options are two moderate, three low or four insufficient.

One high, 9 moderate, 12 low or zero insufficient.

And, based on the percentage, this is grey zone and we'll move on.

CO-CHAIR LANG: Okay, feasibility?

DR. MURRAY: Feasibility. So, all the data elements are in defined fields in electronic claims. The measure is based on readily available administrative billing and claims data.

The AHRQ software is publically available and people have over ten years of experience using it with and there are no fees. So, the feeling should be that it's feasible.

CO-CHAIR LANG: Any concerns, members of the committee, regarding feasibility?

Seeing none, we'll proceed to vote, feasibility.

MS. AMIRAULT: Feasibility for Measure 0279, one for high, two moderate, three low or four insufficient.

Seventeen high, 3 moderate, 2 low and zero insufficient.

Based on the percentage, we'll move along.

CO-CHAIR LANG: Feasibility?

DR. POLLART: The current use of the measure is publically reported, and it's used in accountability programs.

Improvement results were reported. There are -- the PQI 11 hospital admission rate decreased by 87,000 fewer hospitalizations from 2011 to 2013.

The committee felt that usability and use were -- could be used to further the goal of high quality efficient healthcare.

CO-CHAIR LANG: Dale?

CO-CHAIR BRATZLER: So, two things, I'll raise my issue about unintended consequence again, because this is one of the measures that, while designed for population level evaluation, it's being used at the practice level, again. And, that's, again, not AHRQ's fault, I

understand that. It's not the measure developer's fault.

My second issue, though is, I know data that's presented that the rate for this measure is going down. I certainly know in our institution, with the availability of a wide variety of panels to identify viral pneumonia in adults that we're seeing a fairly dramatic reduction in the number that have unassigned causes that used to be assumed to be bacterial.

And, now, you know, you know, you're finding metapneumovirus and respiratory syncytial virus and all sorts of viruses in adults. And so, perhaps, they're being coded as viral, they're not including the numerator any more, but that reduction in the numerator over time may simply be that we're much, much better in diagnosing viral pneumonia in adults which didn't used to do very consistently.

So, I have some questions about whether the measure is improving or are our diagnostic ability for viral pneumonia has

improved a lot over time.

Which then raises, for me, the question influenza vaccine clearly an important intervention to improve this metric, pneumococcal vaccine probably important. But, there are whole bunch of viral pneumonias that are being diagnosed now for which we don't see.

CO-CHAIR LANG: Further discussion?

We'll proceed to vote on usability.

MS. AMIRAULT: Usability and use for 0279, one high, two moderate, three low or four insufficient.

Five high, 11 moderate, 6 low and zero insufficient.

Based on the percentage, we'll move forward.

CO-CHAIR LANG: Now, we are considering the overall suitability of this measure for endorsement.

Is there further discussion?

We will proceed to vote.

MS. AMIRAULT: Overall suitability for

0279, one for yes, two for no.

Thirteen yes and 9 no.

Based on the percentage, this is grey zone.

DR. NISHIMI: Okay, are we ready to take up the next one?

CO-CHAIR LANG: Yes, we're proceeding to the next measure, 0708: Proportion of Patients with Pneumonia That Have a Potentially Avoidable Complication During the Episode Time Window.

Would the -- are the developers on the phone? Would they wish to make a comment?

MS. GORHAM: I'm sorry, Karen Johnson?

MS. JOHNSON: Yes, I'm here.

MS. GORHAM: Okay. Are you starting with the intro?

MS. JOHNSON: Yes, if that's how you want me to do it.

DR. NISHIMI: No, Karen's going to address during the reliability. So, Amita and Andrew -- not Andrew -- yes, is it Andrew and Francois, go ahead.

Karen, just jump in when we get to reliability.

MS. JOHNSON: Okay, great. Thanks.

DR. RASTOGI: Thank you.

This is Amita so with ACHC.

We would first like to thank the members of the standing committee for the thoughtful review of this measure and the comments that we received last week during the call.

Before addressing some of the comments, we want to clarify a few points about this measure because this is a resubmission.

In the original measure, which was endorsed a few years ago, we had three categories of avoidable complications. We have removed one and kept the other two.

The one that we removed dealt with complications associated with comorbidities. A patient that may have other similar conditions. Then we thought that those complications would be best addressed with that main condition itself.

So, in other words, the prior endorsed measure had broader complications than this one, you know.

Second, we want to address the reliability piece. We created a very high threshold of reliability prior to use of this measure, asking that the measure users determine the appropriate sample size prior to calculating a risk standardized score and reporting performance.

This sample size is determined by the specific data set being analyzed and significantly reduces the potential of misclassification of a provider.

The goal is to create a fair balance between the need for accountability on outcomes of scales and an assessment of the actions of physicians and facilities in influencing those outcomes.

The pneumonia PAC measure, or the potentially avoid complication measure, is really centered around the patient's experience with an

episode of care.

PACs measure adverse events that can be experienced by a patient and that when they occur lead to poorer overall quality and higher cost of care.

As such, they are very useful for physicians and facilities in the move to alternative payment models because of reduction in PAC's reduces cost of care while improving overall patient outcomes, helping to meet two of the three goals in the tripling.

So, up front, we would like to offer a few clarifications from the discussion from last time.

Very briefly, the first one are patients are included in the denominator as being hospitalized with pneumonia as the discharge diagnosis on the inpatient claim specifies pneumonia.

However, if the discharge diagnosis is amended to be septic shock, the patient is no longer included.

If the patient has a pre-existing condition such as bronchiectasis that can be identified in claims prior to the pneumonia trigger date, that pre-existing condition gets flagged as comorbidity and is used a risk factor to adjust on.

So, we do not count this in the numerator for the measure in the diagnosis even if it appears in subsequent claims because now, it's considered as a risk factor and a pre-existing condition.

The originally endorsed measure included all types of pneumonia in its definition, making it rather heterogeneous population in the denominator.

So, we have modified the measure to separate out well defined diagnosis of healthcare acquired pneumonia such as ventilator-associated pneumonia and aspiration pneumonia.

However, there are a few diagnosis codes that are less well defined or specific and can be used to identify either a community

acquired or the healthcare acquired.

In certain instances, the diagnosis codes were left in, but then they apply uniformly in the measurement process.

And finally, the workbook that is attached within the submission addresses several details including the risk factors that were flagged when present and included in the severity models, they include PAC rates for each provider and their risk standardized rates.

Importantly, this measure is a reflection of relative performance, not absolute performance. The extent of which this imprecision in a specific administrative claims data set exists, that imprecision should be the same for all providers and not affect the relative performance rate.

Thank you.

CO-CHAIR LANG: Thank you.

James, Chana? Evidence?

DR. O'BRIEN: So, as mentioned, this is an outcome measure. The level of analysis is

actually specified as potentially being the clinician, the facility or even population.

This is adult patients age 18 or older who have an encounter, either inpatient or outpatient, that's associated with pneumonia who then are followed for one month and looked at to see if they develop one of these what are called PACs, the potentially avoidable complications.

As the developer mentioned, there are two different types of these.

The first one is one that is thought to be directly related to the pneumonia itself.

The second one are ones that the developers suggest are patients safety failures.

This is an outcome measure and so, obviously, it is one up for maintenance, so the evidence discussion is a little bit different than if we were looking at maybe a process measure.

And, we'll get to a little bit of the specification. I think that on the face it, reducing preventable avoidable complications

obviously is very much directed towards patients and would be appropriate.

I think the question for the committee that we'll have to decide is whether or not as specified this measure actually accomplishes that.

CO-CHAIR LANG: According to the NQF algorithm for evidence, this is eligible for a pass rating, as the underlying rationale is the same since the last NQF endorsement review.

So, we don't have to vote, if the committee agrees.

Okay, seeing no objection, we'll proceed to performance gap.

Oh, please?

DR. YEALY: I guess I'm still a little bit troubled by on the evidentiary basis. It looks to me the PACs would be largely dominated by inpatient care, yet, what's included is both inpatient and outpatient.

And, there's got to be an overwhelming ascertainment bias because we just simply don't

look, you know, people don't get pressure sore evaluations at home every day. They don't get the pressure sores either, but the truth is, we don't look anywhere near as closely as say with DVT and phlebitis, whatever else you pick up.

And so, I'm struggling here to see, this does -- this looks like a hospital-based measure wrapped in another set of clothing. Not that there's anything wrong with trying to cut back on complications.

The other problem I have is that the PACs aren't weighted, so respiratory failure and phlebitis count the same. They strike me as dramatically different preventable events. One is, you know, unfortunate, but not a particularly big deal. The other one is, you know, awful and potentially life-threatening.

And so, that's why I have concerns. I wasn't part of the previous evaluation of this, it's just, on face value, it looks -- it doesn't seem right to me.

CO-CHAIR LANG: Would the developers

wish to elaborate further on the measure?

DR. RASTOGI: So, regarding the rating of the PACs, we agree that all PACs are equally rated and the whole idea here is that if a patient develops one complication, it quite often sets off a roller coaster and has many, many complications.

So then, that particular patient will still be counted as a yes or a no, so the complication is there.

When you're measuring performance of providers, for example, say physicians managing the pneumonia condition, then we will hold them accountable. But then, what percentage of their patients had these complications?

If it is one patient who went bad and had a lot of bad complication, then it would not adversely affect them.

However, if they are consistently bad and many of their patients or a majority of them had potentially avoidable complications, then they will stick out as an outlier.

So, it's a relative performance measure and that's everybody would be measured according to the same yardstick.

DR. LAMPONE: I just wanted to add a comment on that.

As you look at our elder population, more and more of those patients are being treated in the home setting and there may be less opportunity to impact, intervene or prevent some of these complications in the home.

When you get into a lot of assisted living facilities where you have large populations of these, I think those same problems exist, less so in skilled nursing facilities.

But, I think you'll have a difficulty and you may find there are cohorts of a medical practice that treat many of these patients in an assisted living or in the home.

So, you may have some skewed data in that agreement.

DR. YEALY: You know, again, that reply strengthens my concerns.

So, I'll give you two physicians, one who surveils every day for phlebitis and provides wonderful care, the other one does nothing and has they both have four events happen, one has four respiratory failures, the other one has four phlebitis, they get patted on the head and they look the same by this measure.

And, yet, they're wildly different levels of care and quality. And I'm just not certain that, as constructed, this dichotomous approach can give us the feedback that we need.

DR. O'BRIEN: I suspect that under the scientific acceptability, that will be the opportunity for talking about that.

There are other concerns, I think, also around the way this is specified.

CO-CHAIR LANG: Todd?

DR. DORMAN: Just a simple question of clarity.

Is mortality considered a PAC?

DR. RASTOGI: No, we don't specifically look for that. Mortality is a

separate measure.

DR. DORMAN: Although the list of PACs that are included, I'm dubious that you can die without having one of them.

DR. RASTOGI: That's right. We don't count it separately, but if the patient had, say, respiratory failure and died, then the respiratory failure would be picked up as a PAC.

DR. DORMAN: I asked related to the home mortality. So, we just discussed the attempt to improve quality by outpatient home management of pneumonia in which case I wonder how all of these would be picked up, coded and that's why I asked.

DR. RASTOGI: Right, and our measure is tested at the below 65 population. So, yes, these are valid concerns for the entity.

DR. NISHIMI: Is there anything else specifically related to evidence? We've kind of skipped ahead a little, so I'm trying to bring us back a little bit.

And, there didn't seem to be any

indication to vote. Is that the Chair's assessment on evidence?

CO-CHAIR LANG: My understanding is that, based on the NQF algorithms, this is eligible for a pass rating.

DR. NISHIMI: Right.

CO-CHAIR LANG: But, does the committee believe that it's appropriate to vote on the evidence? Yes? No? Maybe? No? No.

We'll proceed then to performance gap.

DR. O'BRIEN: The developers don't present actual gap. They do present some performance scores for 82 facilities and 170 providers that show a median performance for -- in the 60 percent range, with an interquartile range that varies from in the 40s up to the 70s to 80s.

There's not that I could find any analysis related to gender, socioeconomic status, race or ethnicity or geographic differences.

MS. WEST: A majority of the data that were provided for gap were related to the

pneumonia itself and not specific to the PACs. That was one of the big discussion points that we had during our work group meeting.

DR. SCHINDLER: And, Chana, I don't know if this gets at the same point, but is there or are there data -- I guess I'm a little unclear of why we're looking at PACs only in the context of pneumonia. And, maybe I missed that in the beginning.

But do patients with pneumonia look wildly different from other acutely ill patients in terms of their PAC rate?

MS. WEST: Not that we saw in any of the data that were provided. So, that might be a question for the developer and anything that they saw that the did not provide.

DR. O'BRIEN: And, it might come to usability, what I read, the majority of the PACs that are reported are actually PACs associated things like bypass surgery, so aren't even specific to this diagnosis.

DR. NISHIMI: Does the developer want

to address those comments?

DR. RASTOGI: So, yes, maybe I'm unclear about the question. Is that related to the performance issues?

DR. O'BRIEN: So, I think the question was related to why PACs associated with pneumonia as opposed to PACs in general or other diagnoses? What's the reason to tie it to pneumonia?

DR. RASTOGI: Right, right. So, our potentially avoidable complications go beyond just direct pneumonia PACs.

So, we mention that, you know, appropriate use antibiotics and up-front care can reduce some of the respiratory kind of complications that are related to pneumonia.

And then, the type 2 PACs which are often are seen because of, you know, say, the pressure sores, line sepsis and all that could be avoided by better processes and care.

DR. DIGIOVINE: I think that --

DR. DE BRANTES: This is Francois de Brantes.

If the question is about whether or not there are other measures for similar pulmonary conditions or other pulmonary conditions, other measures of avoidable complications, the answer is yes.

We have measures of potentially avoidable complications for dozens of conditions, both chronic, acute and procedural. And, we've submitted those in different stages to the National Quality Forum for review.

DR. DIGIOVINE: So, I guess two comments.

One is, the data seems or the submission seems to argue that there's a performance gap because there's variability.

But, as we've pointed out, the patients are not required to be admitted for pneumonia. So, I think there's going to be natural variability because some patients are going to be outpatient and some are going to be inpatient.

Also, a lot, I think, I know James

will get to this later, but a lot of these PACs are things that I think a lot of us believe are present on admission.

So, a patient who gets -- you said you excluded patients with septic shock, but my understanding is, as long as septic shock isn't the first diagnosis, you're not excluding septic shock.

And so, I think a lot of the variability also could just be based in variability in the way people code and the way people abstract charts.

So, I'm not sure what you're positing as the performance gap.

DR. RASTOGI: So, coming back to the septic shock situation, yes, if the patient, and as the previous discusser also explained, that the claims data, the discharge diagnosis is the one that the hospitals code and report on the claims.

So, if the discharge diagnosis was changed to septic shock, then that patient will

fall out of the denominator as well. It would not be considered as a pneumonia admission, but maybe the patient was admitted and it was only the in sepsis at the time of admission.

DR. DIGIOVINE: But, you're describing administrative data as if it means that the first thing we write on a code sheet is the only diagnosis and everything else is a complication. And, that's not the way we code charts, certainly not the way physicians diagnose and treat.

So, on any given patient, I might say pneumonia with shock and then the second patient I might say septic shock from pneumonia. Those are not different.

DR. RASTOGI: So, the hospital -- the septic shock patient will most likely be hospitalized and then the hospital diagnosis will determine whether the patient will be kept in or not.

Hospitals alter the claims, the admitting diagnosis is often changed to be the -- to be modified and then the discharge diagnosis

is what was the final discussion during that admission, whatever the physicians said was the real reason that the patient got hospitalized.

If the septic shock developed while the patient was in the hospital, admitted for pneumonia and then eventually progress and went south, then the septic shock would not be the discharge diagnosis.

CO-CHAIR LANG: Are there additional comments or questions for the developers regarding performance gaps specifically?

Seeing none, we will proceed to vote on performance gap.

MS. AMIRAULT: Performance gap for Measure 0708, one being high, two moderate, three low or four insufficient.

Two high, two moderate, 13 low and six insufficient.

So, this does not pass.

MS. BAL: And, as mentioned earlier, since this is a maintenance measure, you can choose to move forward with reserve status, but

you heard the stipulations and it's really up to you, your decision to make.

CO-CHAIR LANG: Let's take a hand vote. The vote is whether to proceed with the understanding that if we do consider -- if we do continue to consider this measure, it would be a reserve status or to suspend further consideration of the measure, in which case, the measure would have failed.

Okay, so all those in favor of continuing to consider the measure with the understanding it may proceed to reserve status, please raise your hands.

Those who would wish to suspend further consideration of the measure?

Anyone who has not voted or who abstained?

I will note that among those present here, the vote was unanimous for option to, to not consider the measure further.

DR. NISHIMI: There was one vote to continue by electronic submission.

CO-CHAIR LANG: Give us just a minute here.

All right, the schedule says that we'll take a break at 10:15. We're a little -- we're going to take a break now until 10:15 and then we will reconvene.

Thank you.

(Whereupon, the above-entitled matter went off the record at 10:03 a.m. and resumed at 10:21 a.m.)

CO-CHAIR BRATZLER: All right, let's go ahead and get started.

So the one thing that we need to do that we didn't complete from the initial conversation this morning is talk about discussion of the related and completing measures. So staff has pulled the list here of all the measures that we've reviewed since yesterday on asthma. And there are quite a few of them. There are nine different measures.

I can read the numbers if you want. But I think the gist of the conversation is about

are there any other areas of these metrics that we want to relate to the developers to think about harmonizing or recognizing some competition between the measures.

So two issues that have already been addressed multiple times by this group are that some of the measures have different data sources, some of them measure different populations of patients. Now there's been a fairly consistent message about asking the developers to look at harmonizing on age ranges where appropriate because there's a number of these measures that have had different age ranges.

But any other issues that this committee -- based on the asthma measures that we've talked about -- would like to send back to the developers as consideration to work with other developers around some harmonization or alignments.

Yes, Bill.

DR. GLOMB: So you mentioned -- you mentioned the data sourcing and you mentioned the

age range harmonization. The only other one that I think that I'd consider is the diagnosis. What constitutes the diagnosis of asthma.

We have had -- we have had single, you know, a single diagnosis. We've had two-plus diagnoses. It would be nice if we could come up with a standardization for what is chronic persistent asthma, what is asthma, what is not asthma.

CO-CHAIR BRATZLER: Yeah. So harmonization of the denominators in certain ways of making sure we have consistency of the codes used to identify, particularly things like persistent asthma which the codes are sometimes difficult to -- other things which you recall from the past two days now in discussions of asthma metrics?

CO-CHAIR LANG: I don't know if you had mentioned it, but did we have something come up with risk adjustment for these folks?

DR. NISHIMI: You didn't have a standard recommendation. You discussed risk

adjustment for some. Some of them were, you know, the process measures and so it didn't really come up.

CO-CHAIR LANG: Okay.

DR. NISHIMI: We can make a general comment that those that do have the risk adjustment, they need to take a look at standardizing the approach for risk adjustment.

Is that what you're indicating?

CO-CHAIR LANG: Sure.

DR. NISHIMI: Anything else? Okay.

CO-CHAIR BRATZLER: I'm trying to walk through the list just to make sure I recognize all of the measures. We have a host of different measures by different developers, different settings, so.

Anything else that we need to talk about that we want to bring back to the developers -- and then all of that information would then come back to us as a committee on one of our conference calls in the future.

DR. NISHIMI: Okay.

CO-CHAIR BRATZLER: Ready to move forward?

DR. NISHIMI: Yes.

CO-CHAIR BRATZLER: Do you want to take it or do you want me to take it? I'll go ahead with this.

Okay. So the next measure on the list is 0035. It's Pediatric ICU Unplanned Readmission Rate by Virtual PICU Systems. On the phone, hopefully, is Matt Scanlon to do an introduction of Measure 0035. I'm sorry, 0335. I'm sorry.

MS. BAL: First, I would just like to mention that Mitch Harris is also conflicted on these -- all three of these PICU measures, just so everyone knows. And that -- Kathy, could you see if Matt from the PICU group is on the phone, or anyone else from PICU -- Virtual PICU Systems?

THE OPERATOR: Matt has not joined yet. And I don't see anyone else on the line.

CO-CHAIR BRATZLER: So what we'll do is we'll -- so as noted, Matt may be expecting the

call to start at 10:30, so that may be why he is not here yet. Why don't we go ahead, either David or Bill, and discuss evidence, and then we'll go from there.

DR. GLOMB: Okay. I'll just go ahead and start if you're okay with that, David. Then I'll let --

DR. STOCKWELL: Yeah, sure.

DR. GLOMB: I will just -- let me kind of do a little bit of a mini-discussion of our presentation since we don't have them on yet.

Again, this is about PICU unplanned readmission rates, looking at the total number of patients requiring unscheduled readmission to the ICU within 24 hours of discharge or transfer. The rationale is that this measure be used in conjunction with one of the other measures we'll discuss, 0334, which is severity-adjusted length of stay. Because theoretically you could game your length of stay measurements by prematurely discharging patients and then they get readmitted because they weren't really ready to go out the

first time. And this would balance that measure out.

The numerator -- it's a fairly simple math equation. It's the number of unplanned readmissions within 24 hours after discharge or transfer from the PICU per 100 patients. And it's in children only under 18 years of age. Excluded are all above 18 years of age.

DR. SCANLON: I'm sorry to interrupt. This is Matt Scanlon. I just wanted to let you know I have joined the call.

CO-CHAIR BRATZLER: Thanks, Matt.

DR. SCANLON: Okay.

CO-CHAIR BRATZLER: We're going to go ahead and talk evidence and then we'll let you --

DR. SCANLON: Perfect.

CO-CHAIR BRATZLER: -- give a little brief introduction to the metric.

DR. GLOMB: I'm going to throw an asterisk in and ask for discussion by the committee. This is termed as an outcome measure, readmission being -- is it really an outcome

measure or is it a process failure, that's a question that I have for this.

This measure was originally endorsed eight years ago and it's up for revisiting at this time.

So some of the evidence, outcome measure of unscheduled readmission within 24 hours, indirectly measures process, those decisions related to discharging the patient while directly measuring a PICU resource utilization due to unplanned readmissions. So in that sense I suppose it's an outcome, but not necessarily a patient-directed outcome.

The developer, again, would like to balance this measure along with 0334 to make sure that those -- that those are accurate measurements. And I think that's appropriate.

One of the questions for the committee, for our work group was, was there at least one thing that the provider could do to change based on the measured result? And there really wasn't anything identified within this, so

obviously it's thought that an unplanned readmission is a failure and, you know, maybe they shouldn't have been discharged yet or there's something missing from the process of transferring the patient to a step-down unit or the ward. And there the developer tests that the rationale and evidence of the outcome have not changed since this was last endorsed for review.

And so that's one of the things we need to decide is whether or not the evidence basis has changed and if there's any need for repeat discussion or vote on the evidence.

CO-CHAIR BRATZLER: So, David, do you have anything to add to Bill's introduction there on evidence?

DR. STOCKWELL: No. I would just suggest that if used as a measure on its own it probably doesn't make sense, but paired with the other measure as a compliance -- however that gets combined, I'm still not clear how that actually happens or what the process part is. But standing on its own it's probably less of a

useful measure and you may not be able to make changes for it. But as used as a balancing measure to tell you whether or not you are overly aggressively discharging your patients out of the ICU, it's much more helpful in that regard.

CO-CHAIR BRATZLER: Okay. So what I'm going to do at this point is ask -- before we vote on or decide whether we are going to vote on evidence or not, I'm going to let Matt give a brief introduction to the measure, which we didn't do before we started that conversation.

So, Matt, if you want to go ahead for two or three minutes.

DR. SCANLON: Sure. And I apologize for being late. I'm working in the ICU this morning, so it's been juggling things.

So as the first speaker that I caught mentioned, the unplanned readmission measure metric was first introduced eight years ago and has been reapproved once in its history. And the intent has always been to use it as a balancing measure to the issue of length of stay.

In the ICU -- in the pediatric ICU world, at least using the VPS data, there is a wide range in length of stay between organizations. But one of the factors that could influence that is the potential for gaming by prematurely discharging patients and then having to readmit them within a short time period.

When the measure was originally created, the Joint Commission actually had a metric of unplanned readmission at 48 hours. They've since dropped that. We felt that that was not necessarily sensitive enough to reflect a premature discharge, and so we developed a measure. And there was some literature that suggests 24 hours was a reasonable place to look at. And that was the foundation of that.

We have found -- and I apologize because this was done in response to the questions we received, so it's not in the original submission but it's available to provide -- that there was a statistically significant correlation with a Pearson correlation

coefficient of negative .398 between unplanned readmissions and severity-adjusted length of stay. That is, organizations that have a shorter length of stay also have higher rates of unplanned readmission, which is not necessarily surprising and is exactly why we felt these two measures should go hand in hand.

So I can speak more to specific questions. But let me stop there and let you take it from there.

CO-CHAIR BRATZLER: All right. Thank you, Matt.

Then before we go forward with our conversation about evidence I wanted to ask NQF staff that I know it's been mentioned that there is a formal pairing process of saying that one measure doesn't stand without the other. Can you tell me about that?

DR. NISHIMI: So in NQF's internal system and in the public-facing material that's exactly what is said. As with other measures, it doesn't prevent, you know, external implementers

from doing what they want to do.

In this case, however, because VPS holds the software and there's a proprietary mechanism, they in effect, you know, control much more tightly than it would for any other measure the fact that they are paired. But in terms of our public-facing information it's very clear that they're paired.

DR. DiGIOVINE: Can I just ask on that? I'm just noticing on the front of the first page of our measure it specifically when asked -- the statement is if paired, what is the reason the other measure is included? And it says not applicable. And if it says if this measure is paired or grouped give the NQF number and title, and it's blank.

DR. SCANLON: So Matt Scanlon here. And this is -- and if I misspeak, please, NQF staffers, correct me, but the implications and connotations of pairing, as I've been through this for eight years, have changed over time. And four years ago I don't remember a specific

reason but there was a reason at the NQF level that they did not want to do a formal pairing. So that's why we carried that forward. That was not because we felt they were disassociated. In fact, the language we use in the measure suggests otherwise.

But again, there was some technical reason that at the time that term pairing was felt to be inappropriate. And I can't speak further to what that rationale was.

DR. NISHIMI: Yeah, and I'm afraid I can't help either because I wasn't here. But I can tell you that in the current internal, what we refer to as OPUS, and the external Quality Positioning System search you will see that it's paired.

DR. DiGIOVINE: So you are saying it -- despite what it says on here, these actually are paired measures.

DR. NISHIMI: It indicates that this measure should be used with 0334, yes.

MS. BAL: Also a clarification. We

started that process after the form was submitted, so that's why. The way that it was submitted it was not paired at the time. But we are in the -- we've processed it to be paired now.

CO-CHAIR LANG: We have had other measures and we will have this afternoon to reach that line, was made out with the other measure listed with the number.

CO-CHAIR BRATZLER: So at this point we've had a conversation about evidence. Is there any other discussion of evidence? This is a already-endorsed measure, so it's a maintenance measure so we do not have to vote on evidence if the committee agrees.

So how many -- any other discussion?

(No response.)

CO-CHAIR BRATZLER: Then how many committee -- anybody raise your hand if you think we need to re-vote the evidence for the measure.

(No response.)

CO-CHAIR BRATZLER: Okay, seeing none,

we will go ahead with discussion of gap.

DR. GLOMB: David, do you want to take the gap discussion?

DR. STOCKWELL: Yes. Yeah, sure.

So there are recognizable gaps within VPS. The range is between zero and 1.6 percent of patients. And so that I think will satisfy the thing that was concerning that was mentioned just a bit ago was that there hadn't been any real change noted over time. But I know that that's something that we will talk about later.

And also, just jumping to disparities, there were not really any disparities noted in this measure from VPS. So we will note it in the paired measure a little bit later.

DR. GLOMB: Yeah, I just want to add, the only thing that came up disparities-wise had to do with whether a child had insurance or had no insurance whatsoever. And the uninsured kids were sicker when they got there. Their lengths of stays were shorter and their mortalities were probably higher. That was the only thing that

stuck out. And not surprising.

CO-CHAIR BRATZLER: So I just have to ask the question, is zero to 1.67 percent that substantial of a gap between PICUs? It seems like a fairly narrow range.

DR. STOCKWELL: So I guess I would say that it seems that there is -- and, yes, this is evidence based on what Matt just told us again this morning -- but if there is a reasonable spread, enough to show that there's a moderate correlation, negative correlation between the two measures then that, to me, would say that it would be at least meaningful enough to show that there is some gap.

CO-CHAIR BRATZLER: I think Tom has --

DR. LAMPONE: And I just wanted to clarify, William had made a comment on disparities. And if I heard you correctly, there was a disparity between insured and uninsured. Uninsured presented with more advanced or more complicated disease, but had a shorter length of stay and had a higher mortality.

DR. GLOMB: Not necessarily related. Those last two are not necessarily related. That's not proven by the evidence. So that's not --

DR. LAMPONE: Okay.

DR. GLOMB: -- clear to us from the evidence.

DR. LAMPONE: Okay.

DR. GLOMB: And you'll see that same statistic throughout all three of these measures we're about to talk about.

DR. LAMPONE: Okay. And did we see any statistics that looked at readmission to the ICU in those populations.

DR. GLOMB: Yes. Between insured and uninsured.

DR. LAMPONE: Okay. So there was a disparity there with the --

DR. GLOMB: There was a disparity there.

DR. LAMPONE: -- base of the measure which is looking at readmissions.

DR. GLOMB: That's correct.

DR. LAMPONE: Okay.

DR. GLOMB: That's correct. It wasn't huge but it was statistically significant.

DR. DiGIOVINE: I think throughout this I'm going to be struggling with the idea that this is -- I think we are saying this isn't an important measure by itself but it's important because it's balancing something else, which I struggle with.

And I guess I just wonder whether the developer has ever thought to do what I think we still do, which is if the patient is readmitted to the ICU within 24 hours, just to count that all as the same admission and then make their length of stay admission A plus admission B and use one measure as the quality measure around length of stay.

DR. SCANLON: So there's a couple factors there. One is that not all readmissions to the ICU can be viewed as a breakdown in care or as management-related. Some are calculated

decisions that patients are transferred and then for a change in the condition, unrelated, or maybe because of a bed crisis within a unit they move a patient out with full knowledge that the patient is coming back, which is why we have been very careful in the software and in the data dictionary to break down what the definition of scheduled versus unscheduled is.

I think we also know anecdotally, and this was true at the time these measures were originally developed, that some of the centers that had very short lengths of stay admitted that they pushed kids out to keep those numbers low. And so I think the issue is -- I don't know that we could presume to just staple all within 24 hours together. And, you know, I guess we could say unscheduled readmissions should be compiled onto the original one.

But having said that, that's not been what's been described at least in our review of the literature. And so that would be really almost creating a new metric for length of stay

which is -- for ICU length of stay which I don't know that we -- that would be breaking new ground. That doesn't mean we shouldn't do it, but I think that would take a lot more reflection and investigation of just the validity of that, never mind how the community would view that.

CO-CHAIR BRATZLER: So back in the room. Edgar?

DR. JIMENEZ: Yes. Just -- I'm completely in agreement with the statement they made.

But one more thing is that we have been seeing readmission rates that are potential problems with even the hand-off of patients, you know, where that's not appropriate, you know, transference of care. So it's something that we are paying attention very closely.

CO-CHAIR BRATZLER: James?

DR. O'BRIEN: I'm wondering -- and the developer maybe can comment about this, in the adult ICU literature there's certainly appreciation for a combined end point of ICU-free

days, which is days alive not in the ICU, which does incorporate both the possibility of people being readmitted to the ICU, how long they're in the ICU, and mortality all in one measure. I don't -- if you can comment whether or not that's reached the pediatric literature?

DR. SCANLON: Not that I'm aware of. And I think, at least from the standpoint of measure development one of the challenges is that the data from administrative data sets is often problematic because of the lack of specificity about hours spent in the ICU and rather making determinations by midnight census as to where a patient is. You know, if you could admit a kid to the ICU, transfer them out, admit them and transfer them out again, and if all that happened within a single day it wouldn't show as an ICU admission at all.

The VPS is a clinical database. And because we don't have currently the hospital days but rather just the ICU days, it was technically not feasible without trying to use the data

source that we felt was more problematic. And again, to my knowledge, although certainly there's some smart people on your panel who may feel otherwise, I have not seen the ICU-free days make it into the pediatric literature very far.

CO-CHAIR BRATZLER: So I'm going to pull us back just a little bit. So we've made several suggestions that almost sound like material changes to measures that would require us to go through the whole process that we went through yesterday about whether a whole new measure should be developed. So I'm going to pull us back to discussing the measure that we are presented with, which is what we have to decide today.

So are there any other discussions about gap? Yes, Bruno.

DR. DiGIOVINE: Just let me just then clarify. Everything we're going to ask, are you asking is there a performance gap in, for example, in readmissions as a standalone measure? Or are we asking is there a performance gap for

length of stay for which readmissions would be an important thing to look at?

DR. NISHIMI: They stand alone.

CO-CHAIR BRATZLER: Any other questions or clarifications?

(No response.)

CO-CHAIR BRATZLER: Okay. Let's move forward and vote on gap.

MS. AMIRAULT: Performance gap for Measure 0335, 1 being high, 2 moderate, 3 low, or 4 insufficient.

(Voting.)

MS. BAL: Ella, could you please resend your vote.

(Voting.)

MS. AMIRAULT: Zero high, 12 moderate, eight low, one insufficient.

We're going to do a re-vote.

DR. STOCKWELL: Does that mean even for the folks on the phone?

CO-CHAIR BRATZLER: Yes.

DR. NISHIMI: No, they're fine. No, we

have the phone votes. So, we're looking for 22.

MS. BAL: Just give us one second.

MS. AMIRAULT: Okay, performance gap for 0335, 1 being high, 2 moderate, 3 low, or 4 insufficient.

(Voting.)

MS. AMIRAULT: Zero high, 13 moderate, eight low and one insufficient.

Based on the percentages you have gray zone. And we'll move along.

CO-CHAIR BRATZLER: Okay, so we'll go ahead with the conversation. The next will be reliability.

DR. STOCKWELL: So reliability that the numerator is total number of -- all of this is stuff you would expect -- total number of unplanned readmissions within 24 hours of discharge or transfer from the ICU. The denominator is 100 PICU discharges for those less than 18 years of age. And then the exclusions are older than 18 -- 18 and older.

Data selection is essentially

electronic clinical data as well as some administrative data. And during our discussions there weren't concerns about the reliability for the specifications and for the testing. There is not new data. The developer noted that a separate proportion using previously-established methods, therefore no further reliability assessment is indicated.

CO-CHAIR BRATZLER: Any other --

DR. STOCKWELL: I don't.

CO-CHAIR BRATZLER: Yes?

DR. LAMPONE: This is for the developer. When you look at the data, what percentage of it is administrative? Could you tell us?

DR. SCANLON: Zero.

DR. LAMPONE: Okay. So it's all --

DR. SCANLON: We --

DR. STOCKWELL: I'm sorry about that.

DR. SCANLON: I'm sorry.

DR. STOCKWELL: Sorry Matt, that was just -- this is the challenge of looking at this

as an individual, versus some of the other ones that we are going to discuss in a little bit. So that's my fault, sorry about that.

CO-CHAIR BRATZLER: All right, so go ahead.

DR. SCANLON: Oh, I'm sorry, is there a question I didn't answer? I'm just --

DR. LAMPONE: So none of the data is -- it's all administrative data?

DR. SCANLON: No. None of the data is administrative.

DR. LAMPONE: Okay.

DR. SCANLON: No. It is clinical data. It's entered by clinicians. We have, to the minute, times of entry and departure from the ICU, which is why we can provide very detailed calculations.

CO-CHAIR BRATZLER: So just to clarify, because there's no new reliability data presented, because the measure is broadly in use and already endorsed, we don't have to vote on reliability either. We can make the decision to

move on with the conversation about validity.

So anyone who wants us to go ahead and vote on reliability, raise your hand.

(No response.)

CO-CHAIR BRATZLER: Seeing none, we will go ahead and discuss validity.

DR. GLOMB: I'll go ahead and talk about that.

Again, under the validity testing I guess one of the -- the question before our committee is, are the specifications consistent with the evidence. And I guess I have a counter-question. And that is, does unscheduled readmission equal poor quality of care.

The assumption is there, intuitively suggested. I don't think that I would oppose that. I just wonder whether the evidence is there and, therefore, that would mean that the validity would be in question if it's not. I know we voted on the evidence not being necessary. I just think we're making a jump looking at this measure in isolation.

CO-CHAIR BRATZLER: Chana and then we'll go to David.

MS. WEST: So I was looking through all the documentation and I didn't see a clear indication on what they're characterizing as planned versus unplanned in terms of the PICU admission, so.

CO-CHAIR BRATZLER: Matt, can you comment on that, planned versus unplanned?

DR. SCANLON: Absolutely. The operational definition that we -- in fact is part of our ongoing inter-rater reliability -- was the admission to the ICU, whether it's an admission or a readmission, known about 12 hours before it occurred?

So, for example, if we have a kid that 12 hours before they're coming to the ICU that someone calls and says we need a bed for this, that reflects that this is a conscious choice. It's not a -- either a deteriorating child or a child that was prematurely transferred out and needs to come crashing back to the ICU.

So while you could argue that the 12-hour window of knowledge -- the foreknowledge about a child is arbitrary, there was some literature -- and I apologize, I don't have the reference in front of me -- that supported that.

And then as part of the ongoing -- as I said, the inter-rater reliability -- that, as well as the internal validation checks of the software, we look for those. And anything that seems out of -- potentially misclassified, so for example, an unscheduled patient admission from the PACU triggers a review at the site level to make sure that that was properly classified.

MS. WEST: So this is reliant on manual documentation that's then entered in the software? Or am I missing something in the work flow.

DR. SCANLON: No. So at the unit level all ICUs know who is scheduled to come into their ICU. That's part of running an ICU. And if we know if they are essentially scheduled for a bed greater than 12 hours in advance, that's treated

as a scheduled readmission versus an unplanned -- I'm sorry, a scheduled admission rather than an unplanned admission.

Then separate from that is the question of were they just discharged from the ICU. So a patient who left the ICU could come back to the ICU for either planned reasons, and thus scheduled, or unplanned reasons which would be the unplanned readmission to the ICU.

Did I answer your question?

MS. WEST: Yes and no. So the time period that you're looking at -- and just, you know, to clarify, I'm a NICU nurse so I understand all the processes.

DR. SCANLON: Okay, sure.

MS. WEST: But in terms of the 12-hour time frame that you're talking about, is there some literature --

DR. SCANLON: Yes.

MS. WEST: -- that speaks to it being a 12-hour time window or --

DR. SCANLON: Again, at the time of the

measure development there was a -- there was at least one reference that we used. I don't -- and I apologize, I don't recall what that is at this point. But the rationale being that in our world it's not uncommon for a patient to be discovered to have a problem and they need an ICU bed. And so the question was what cut-off would constitute having enough warning to manage the patient.

And part of this also gets to the larger macro issue of providing adequate nursing and other resources to care for the patient. So there's wide ranges -- just speaking of -- and this may be a little off topic, but in our ICU in Milwaukee, which is a 72-bed pediatric ICU with three floors, our pure medical floor, the unplanned -- or unplanned admission rate is 90 percent, meaning 90 percent of patients we don't know they're coming 12 hours in advance, which is a giant care issue because it's hard to plan for a party if you don't know who's coming.

In the case of readmissions, we use the same definition to say was this readmission

planned to come, in other words, an elective, conscious decision, or not. And there was a reference for the 12 hours. I apologize, I don't recall what that is and would have to go back and see if I can pull that for you.

CO-CHAIR LANG: So, Matt -- go ahead, Bruno.

DR. DiGIOVINE: Sorry, David.

Matt, certainly in the adult world it is not uncommon for people to hang out in our wards for 12 hours waiting for a bed. Does that not happen in pediatric hospitals?

DR. SCANLON: Oh yeah, oh definitely. Wait, waiting for an ICU bed?

DR. DiGIOVINE: Yes.

DR. SCANLON: I think that that is -- I don't know that there's any -- there may be literature. I'm not familiar with any literature about that being an ongoing problem in the ICUs in the pediatric world. Getting into the ICU is not an issue. Getting out of the ICU back to the floors is more of a problem in our world. And I

would defer to the other intensivists on the panel to speak to that, the pediatric intensivists.

DR. STOCKWELL: David Stockwell. That's our experience as well.

CO-CHAIR LANG: Yeah. So, Matt, this is David Lang. Appreciate your elaborating on these issues. And I'm coming at this from -- I'm not a critical care physician so I don't have the kind of content expertise that others do around the table. But looking at this from more of a distance, I actually looked up and Googled the definition of an intensive care unit.

It's a designated area of a hospital facility dedicated to the care of patients who are seriously ill.

It would seem to me -- I'll just throw this out there because it's going to come up, you know, later today we're dealing with other ICU measures, and your measure is coupled with a length of stay and a mortality measure -- it would seem to me that the intensive care unit

setting in terms of what characterizes the designated area, and also the nature of the patients who are seriously ill would vary substantially from the referral center to the community hospital, yet they are both classified as quote, unquote ICUs or, in your case, PICUs.

We have dealt with a number of measures where we have adjusted for socio-demographic factors. But how are you adjusting for the case mix and other variation in, you know, the different ICUs?

DR. SCANLON: So there's a couple things. Let me -- that's a great question. Thank you.

First of all, all the unplanned readmissions are not risk adjusted, but the length of stay measure and the mortality measure, our SMR measure, are risk adjusted using peer-reviewed published algorithms for severity of illness to a handle for that aspect.

At the software level and in our reports one of the things we can do is look at,

by organization type, other differences. So you could look at SMR or unplanned readmission by both hospital type and unit type. And we can break that down by is this a community hospital, an academic hospital, a free-standing children's hospital. Granted, if people don't participate in the software, which is how I can obtain data and look at these differences, then obviously I can't comment on an organization that isn't submitting.

But what we do, if anything, is stratify by like characteristics when we are doing internal comparisons for benchmarking to say, is a hospital or is a given ICU performing on par with its peers? Because of, as you said, the wide difference.

I think -- but as a larger community of providing critical care services to children, that begs the question of should you be in the business at all? And that's more, it's not necessarily -- well, I think it's actually very germane to the NQF process because you can have

hospitals that open up -- a community hospital that opens up a PICU just to say they have a PICU, but that doesn't mean they should be caring for sick kids there.

And so by looking at differences in length of stay, unplanned readmissions and mortality, you can start to say, is this appropriate care for a child in that setting, or should that care even be offered.

CO-CHAIR BRATZLER: Okay, so we're discussing validity at this point. Any other -- yes, Bill.

DR. GLOMB: If I can ask the developer to comment on one other aspect in the validity section here that will apply to the next two as well, it has to do with questions that were raised at our work group meeting about the IRR process.

DR. SCANLON: Yes.

DR. GLOMB: You didn't go into a lot of detail.

DR. SCANLON: I can provide that today.

DR. GLOMB: If you could do that. And what are you doing in terms of inter-rater reliability ongoing testing, or was there --

DR. SCANLON: Oh, yeah, I can address both of those.

So when a given center is coming on board to use the software, first of all there's a training process for data collectors. And the data collectors have to be a nurse by background or a physician, so that we're not having administrative clerks, for example, extracting charts.

The initial IRR process is that they are provided with five patient records which are de-identified. And then they do their abstraction. And we review that compared to the gold standard answers to see how they're doing. And all sites that come on board have to do that and meet a certain threshold of concordance greater than 90 percent before they'd even be allowed to start submitting data.

That process is then done quarterly

with each data submission where they do at least one chart each quarter to make sure that they're maintaining their quality. If there's any question about the accuracy or the -- or if there's essentially discordance in that, that data is quarantined and they go through a re-training process. And until they maintain -- get their IRR up, there's no more flow of that data into the data pool.

Right now our quarterly aggregate IRR rates are greater than 96 percent. The last two quarters are 96.81 and 97.76 percent.

So I don't know if that answers your question.

DR. GLOMB: It does, thank you.

DR. SCANLON: Okay.

CO-CHAIR BRATZLER: All right. So any other conversations or questions about validity?

(No response.)

CO-CHAIR BRATZLER: Okay. At this time we will go ahead and vote.

MS. AMIRAULT: Validity for 0335, 1

being high, 2 moderate, 3 low, 4 insufficient.

(Voting.)

MS. AMIRAULT: Three high, 13 moderate, five low, and zero insufficient. And based on the percentage we can move along.

CO-CHAIR BRATZLER: Okay. So we move to discussion of feasibility. And I did want to note that, as all of you are aware, this is a proprietary measure. So we'll go to the discussion of feasibility.

DR. STOCKWELL: So it's David, I'm going to start. The feasibility aspect of this is really, as you've heard, a simple process of gathering the data. There is I think low feasibility questions in that regard.

The other item to consider in this section is that proprietary item. And we clarified during our work group meeting that proprietary measures are certainly something that the NQF allows. And I -- Matt can give you the real numbers, but I know that the vast majority of the ICUs across the country -- PICUs across

the country participate in this. And so I just -- we note that challenge to feasibility.

But if you wanted to do this, I think we have to consider this as a standalone measure. If you wanted to calculate your own unplanned readmissions rate, there are certainly no barriers to doing that in and of itself.

CO-CHAIR BRATZLER: Anything to add, Bill?

DR. GLOMB: No. Just that it is very feasible for those who are members of the software program, for which there are varying fees. And more time intensive, perhaps more room for variation in scoring for those who are not on board, who would have to do a manual review.

CO-CHAIR BRATZLER: James.

DR. O'BRIEN: So, again, as an adult intensivist, I would love to have somebody do some of this data for us. And I don't think we would ever be able to get the funding to have somebody collect all this data.

Is this -- so I guess, obviously, this

is done a lot in pediatric ICUs. What percentage of pediatric ICUs currently have this software? And how much of this is chicken and egg. How much do they have the software because NQF has these measures, and how much did they have the software before NQF had these measures.

DR. SCANLON: The software has been in existence since roughly -- well, the first generation of software started in 2004 and predates the NQF -- our involvement with the NQF measures. Or, I'm sorry, the creation of these NQF measures.

The percent -- and I've got people from VPS with me on the call, so someone may email me -- but I don't know the percent of ICUs in the nation that have it. We have over 100 ICUs -- pediatric ICUs participating. The majority -- the overwhelming majority are U.S. There's actually some pediatric ICUs in Saudi Arabia and one Canadian ICU that are on board also.

There is a separate software that's

being used by cardiac ICUs, or some cardiac ICUs. And some cardiac ICUs -- pediatric cardiac ICUs are actually using both software. But I would say -- I would venture to say the majority of pediatric ICUs is correct.

I just got an email. It's currently 135 sites are using the software. So I just don't know the denominator of how many pediatric ICUs there are in the U.S.

CO-CHAIR BRATZLER: So I'm just going to throw out that this fee-based registry type data submission is not uncommon. American College of Surgeons, American College of Cardiology -- there's a host of it's fee-based where you actually have to pay somebody to capture your own data and then pay a fee to participate, so.

James?

DR. O'BRIEN: Does the developer have any sense of what the cycle time is for completing a patient record?

DR. SCANLON: I don't know. That actually depends on how long the patient stays.

And it varies. There's a number of strategies.

We ask the patient to be registered in the software shortly after admission. But then completion, if a patient -- and as you look at the length of stay data which is truncated at 30 days, it's not unheard for pediatric patients to stay the better part of a year. Those are outliers. But so while there's ongoing updating, the patient is not submitted into the finalized data pool until their clinical experience is complete, be it through death or discharge.

So I can't give you a specific answer because of the range of length of stay.

CO-CHAIR BRATZLER: Any other --

DR. SCANLON: Having said that, for every discharge the data is then submitted within a quarter.

CO-CHAIR BRATZLER: Any other conversations about feasibility?

(No response.)

CO-CHAIR BRATZLER: We'll go ahead and vote.

MS. AMIRAULT: Feasibility for 0335, 1 being high, 2 moderate, 3 low, or 4 insufficient.

(Voting.)

MS. AMIRAULT: Three high, 13 moderate, five low and zero insufficient. Based on the percentage we'll move along.

CO-CHAIR BRATZLER: And we'll go on to discussion of usability and use.

DR. GLOMB: I'll go ahead and start off usability.

I think this is probably where some of our discussion came in our on work group. Currently the measured data are not aggregated and publicly reported. There are a couple of hospitals that do put that up on their websites. And I suspect it's because they're doing well in the scoring, so it's a useful selling point.

The public funding body in California does require that any PICUs involved -- they put this up publicly through the children's pediatric healthcare -- California Children's Services. And we've heard about the number of centers involved.

I guess one of the biggest questions -- and again maybe this is as a standalone, if we're looking at this by itself -- is that a two-year data collection from January 1, 2012 to the end of December 2014 -- I'm sorry, that's three years -- showed that there was no increasing or decreasing trend for the overall rate of unscheduled readmissions. So with that much data we've not seen any movement in the rates going on.

In terms of potential harms, the developer had said previously that there's a risk of miscapturing the time of the original discharge. Since this is a time-dependent measure, you know, missing it by a minute could mean that you get a different score one way or the other.

CO-CHAIR BRATZLER: David, any other comments about use or usability?

DR. STOCKWELL: No. I think it's well covered.

CO-CHAIR BRATZLER: One question I did

have to your point about no change in the trends over time. I know this measure is not risk adjusted. The other two are but this one is not.

So do you have any sense, Matt, about overall risk for the patient population over time? In other words, if you have a sicker population over time, no trend may actually reflect some improvement, the patients are sicker now than they used to be when they get to a pediatric ICU.

DR. SCANLON: I don't know the answer to that off the top of my head. I would have to -- I have looked at that for given ICUs as part of developing reports for them, but I honestly can't tell you off the top of my head. Again, I have people in the background who are probably trying to answer that for you and I can share it as soon as I get that.

But I don't know that there's been a dramatic upswing in severity of illness over time. I don't know, David, if you have any sense of that.

DR. GLOMB: I have two comments. One has to do with the -- and Chana brought it up and we had brought that up before -- PICUs are really not cookie cutter. As a PICU and NICU doc, I can say that they're not cookie cutter in the same way that a NICU tends to be somewhat cookie cutter. Because it revolves around what specialists are available at a specific children's hospital, what types of illness and disease they deal with primarily.

Different -- across Texas for instance, different children's hospitals have a completely different spectrum of disease that they might take care of. If you've got a patient who needs this, they go to that children's hospital, a patient who needs this, they go to a different one.

So it's bothered me that without risk-adjusting that it's a little bit of an apples and oranges comparison across the way.

The other thing is -- the other question is this doesn't point to a specific --

it doesn't get us in the direction of where do I need to improve to get my rate down. It just says my rate's too high, comparably.

DR. SCANLON: So this is Matt Scanlon. I would agree with that last point. I think at a individual level, you know, what's helpful is looking at this and then reviewing those unplanned readmissions. Once, originally in the -- I think it was the first set of endorsed measures -- there was actually a systematic review of unplanned readmissions as a mandate or as a measure through the NQF process. And that was dropped at the last cycle. And I don't recall the specifics of why.

So, yeah, I think the number in and of itself is only as useful as what you do with it. The reason we provide this in peer review or peer-type summary data to institutions as part of participation is so that they can look and see how they're doing compared to peers. And then they have to dig in and figure out why. Just as any quality improvement process.

CO-CHAIR BRATZLER: Edgar.

DR. JIMENEZ: Just a comment here that we try to take into consideration the statistics -- demographic statistics with the pressure on PICUs. And it's going to be pretty stable for the next few years, whether it's actually in the population over 65, I think we'll see as we talk over the next few measures. We're going to have a much, much higher ratio or a disproportionate ratio of demand for acuity of care in the older population. This is going to be something that we're going to be seeing over the next ten years.

DR. SCANLON: I think the other thing where there's an implication which may affect these numbers is with the change in reimbursement model and a move from fee-for-service, I think there's going to be increasing pressure to move children out of ICU, which may lead to unintended consequences.

CO-CHAIR BRATZLER: That's a very good point.

Any other conversation about usability

or use? Yes, Todd.

DR. DORMAN: I guess I will just share a local. So actually for 27 years we have been tracking readmission rates to adult ICUs. We started with definitions of 72 hours. Found that to be overly-sensitive. Forty-eight hours, and then probably for about 15 years have been looking at 24 hours. And find them very useful, not as the raw rate but as the drive towards the gap analysis that Matt is referring to to understand. And it clearly identified hand-offs mentioned by Edgar as a significant player.

But there are others, including the tie to high-volume, high-pressure days and the pressure for early discharge. Not that anybody would move somebody out that they believe was unstable, but you have those unknown patients that get impacted. And that has led to a change in health system resources in order to deal with those days, which has then been reflected by a decrease in the unscheduled readmission rates.

So I think there have been some hard

examples of the data driving beneficial change.

CO-CHAIR BRATZLER: Any other conversations about use or usability?

(No response.)

CO-CHAIR BRATZLER: Okay. We'll go ahead and vote.

MS. AMIRAULT: Usability and use for 0335, 1 being high, 2 moderate, 3 low, or 4 insufficient.

(Voting.)

DR. NISHIMI: David. David, can you resubmit your vote?

DR. STOCKWELL: I didn't submit it yet but I had to send it by email. Do you see it there?

(Voting.)

MS. AMIRAULT: Zero high, 14 moderate, seven low and zero insufficient. Based on the percentage we can move forward.

CO-CHAIR BRATZLER: And then we'll have our final vote on overall suitability, recognizing remember when we discussed gap that

this measure was in the gray zone. So there will be further discussion down the road because of that point.

MS. AMIRAULT: Overall suitability for 0335, 1 being yes and 2 being no.

(Voting.)

MS. AMIRAULT: Twelve yes, nine no. And based on the percentage this is in the gray zone.

CO-CHAIR LANG: Okay, Dale and I are continuing to do a tag team here. So I'm going to take the next measure, 0334, PICU Severity-adjusted Length of Stay.

Matt, or would you, would developers wish to make a brief statement regarding this measure before we proceed?

DR. SCANLON: Yes. The concept of severity-adjusted length of stay was initially put forward by Dr. Murray Pollack who has been a pioneer in real metrics around the ICU, both in terms of severity-adjusted algorithms and looking at utilization. And we have been using this for a number of years to allow for risk adjustment,

but also look at utilization as a -- and the impact on, potentially, patients and families.

Again, in our minds to look at this without looking at unplanned readmissions is short-sighted and leads to the potential for gaming, which I think for any national measure is always a potential concern.

So I think that sums up the overview and then we can take it based on individual questions.

CO-CHAIR LANG: Thank you, Matt.

David, Dale, you're up.

DR. STOCKWELL: Sure. It's David, I'll do it.

So it sounds this is an maintenance evaluation of an outcome measure. There is not really new evidence presented at this time, although there is, you will hear about an updated adjustment, severity wellness adjustment algorithm, but the data is internally adjusted and there's evidence to that point. But in terms of evidence of the use of this, the focus on the

ICU and the length of stay has been well-documented as something that is worthy of reflection for any medical director of an ICU.

CO-CHAIR LANG: Thank you, Matt. Evidence.

CO-CHAIR BRATZLER: And the only thing I would add is this is reported as an, essentially, an observed to an expected ratio based on severity of illness with the child.

CO-CHAIR LANG: Thank you. I meant to say David. Thank you, David, and thank you, Dale.

The underlying rationale for this outcome measure hasn't changed since the last NQF endorsement review. So this also is appropriate to proceed without a vote, unless anyone wishes to do otherwise. Please raise your hand, if so, otherwise we will proceed.

(No response.)

CO-CHAIR LANG: Thank you.

Performance gap. Dale, David.

DR. STOCKWELL: So in the performance

gap the data that the developer provides shows a severity-adjusted length of stay ratio between .66 and 1.8, with the median being 1.01. And they note that there is no decrease or increase in trends over time. So that part I think is fairly well satisfied.

The question about disparity really is highlighted here, even further compared to the last measure, where again you see that uninsured children have a disparity with shorter length of stay in the PICU. They note that there is also a greater physiologic derangement on admission, and that may lead to a higher mortality.

And the developer also notes that it may reflect pre-hospital practice independent of the care provided by the ICU. They do then move into stratifying by rates next to the age groups, gender and insurance again. And there were several differences amongst all those. We can go through that if the committee would like.

Ultimately the committee felt that there was a performance gap but using this as

something for disparities may not be, may not be as useful as it is for just the overall ICU performance.

CO-CHAIR BRATZLER: And the only thing I'll add is that I think when we talked we were actually -- you know, I was impressed that there is much more substantial variation in risk severity-adjusted length of stay between units than we saw with the previous measure. I mean the ratios were substantially different across the various PICUs.

CO-CHAIR LANG: Further discussion? Questions for the developer?

(No response.)

CO-CHAIR LANG: Seeing none, we will proceed to vote. Janine.

MS. AMIRAULT: Performance gap for 0334; 1 being high, 2 moderate, 3 low, and 4 insufficient.

(Voting.)

MS. AMIRAULT: Six high, 13 moderate, 2 low, and 0 insufficient. Based on the

percentage we move on.

CO-CHAIR BRATZLER: So we'll go on to discuss reliability. Reliability, essentially the same conversation that we had about the last measure. The measure site is captured, all of the measures, but the three measures that this group produces come out of the same data set. So based on inter-rater reliability testing we heard the conversation early, the reliability testing is done at the data element level. And they provided the evidence of the consistency of the data collection across programs. And they have ongoing continuous reliability testing as a part of submitting and participating in the database.

CO-CHAIR LANG: Additional discussion?

MR. STOCKWELL: I have nothing further.

CO-CHAIR LANG: David, were you going to say something?

MR. STOCKWELL: No, sorry. Just that I didn't have anything further.

CO-CHAIR LANG: Okay, thank you.

Okay, we will proceed to vote

reliability.

MS. AMIRAULT: Reliability for Measure 0334; 1 being high, 2 moderate, 3 low, 4 insufficient.

(Voting.)

MS. AMIRAULT: Nine high, 12 moderate, 0 low, 0 insufficient. Based on the percentage we'll move forward.

CO-CHAIR LANG: Validity.

DR. STOCKWELL: So with validity the developer, as you said, had taken an established, published method called the PRISM III, and has now updated that with their own internal data and provided validity testing of that. They essentially just compared a training data set to an independent validation set and found that they had reasonable validity after looking at the various different components of that.

And I believe that the committee was comfortable with the assessment that was provided.

CO-CHAIR BRATZLER: And I don't have

anything to add.

CO-CHAIR LANG: Further discussion? Yes, Bill?

DR. GLOMB: I have a question for the developer.

There was a decision to truncate the length of stays at 30 days. And I know why that is, because all of our units have that one child who's been there for 9 months, 10 months, a year-and-a-half. But is there any validity in the literature for doing that? I have not seen it and I just wondered how you made that decision?

DR. SCANLON: That decision was not our decision. That was the decision of Dr. Pollack based on his review of his -- the collaborative database he was part of or headed up. And that was the, I think that was the PICU use project that led to that originally.

And so why they chose 30, I don't recall, recall the exact rationale, but it was based on analysis of their data that they drew the line at 30 days to deal with those outliers

that you mentioned. And again, that's a peer-reviewed manuscript that we essentially used the methodology of.

DR. GLOMB: Thank you.

CO-CHAIR LANG: Further discussion?

(No response.)

CO-CHAIR LANG: We'll proceed to a vote. Janine.

MS. AMIRAULT: Validity for 0334; 1 being high, 2 moderate, 3 low, and 4 insufficient.

(Voting.)

MS. AMIRAULT: Six high, 13 moderate, 1 low, and 1 insufficient. Based on the percentage we'll move on.

CO-CHAIR LANG: Feasibility.

CO-CHAIR BRATZLER: So for feasibility, exact same conversation as the previous. This is a proprietary measure. We heard that about 135 different PICUs are actually participating. The data abstraction happens at the level the individual unit captures the patient level, and

is submitted via a registered software that's used.

So I don't have anything different to add from the previous discussion.

DR. STOCKWELL: Yes, it's David. I agree.

CO-CHAIR LANG: Bruno?

DR. DiGIOVINE: I just think in the prior discussion we did make the point that you could probably get your readmission rates without having the software. I think there is no way to get a severity-adjusted length of stay without having the software.

DR. SCANLON: I'm not sure -- this is Matt Scanlon -- I'm not sure that I would agree with that. I think that, again, you would have to have, you would have to track your length of stay. I am nervous about the use of administrative data for those purposes, but that's just my perspective.

And separate from that you can calculate severity of illness on all of the

children and then do the math accordingly. It would be an odious task, which is why we built the software and automated it, but there is no barrier to actually doing it long-hand, if you will.

CO-CHAIR BRATZLER: And I think, you know, you're right that the risk model actually is in the public domain; it has been published several times.

CO-CHAIR LANG: Further discussion?

(No response.)

CO-CHAIR LANG: If not, we'll proceed to vote.

MS. AMIRAULT: Feasibility for 0334; 1 being high, 2 moderate, 3 low, or 4 insufficient.

(Voting.)

MS. AMIRAULT: Three high, 13 moderate, 5 low, and 0 insufficient. Based on the percentage we'll move forward.

CO-CHAIR LANG: Usability.

DR. STOCKWELL: So the usability is also very similar as the last time. There are no

publicly-supported measures utilizing this. Although there are a large number of ICUs that do report it.

Sorry for the background noise, by the way.

So very similar to the last conversation we had on the last measure.

CO-CHAIR LANG: Further discussion? Going once. Going twice.

(No response.)

CO-CHAIR LANG: Janine.

MS. AMIRAULT: Usability and use for 0334; 1 being high, 2 moderate, 3 low, or 4 insufficient.

(Voting.)

MS. AMIRAULT: Zero high, 14 moderate, 6 low, and 1 insufficient. Based on the percentage we'll move on.

CO-CHAIR LANG: We are now considering the overall suitability of the measure for endorsement.

Comments from members of the

committee, questions for the developer?

(No response.)

CO-CHAIR LANG: Seeing neither, we will proceed to vote.

MS. AMIRAULT: Overall suitability for 0334; 1 being yes and 2 being no.

(Voting.)

MS. BAL: David, did you vote?

DR. STOCKWELL: I did. I need to do it by email though.

MS. BAL: Okay.

MS. AMIRAULT: Eleven yes, 10 no. This is in the gray zone.

CO-CHAIR BRATZLER: Okay. So we'll move on to the next measure, which is the third of the measures from Virtual PICU Systems.

CO-CHAIR LANG: If I could just interrupt you for a moment. There is an issue regarding scheduling that I just want to raise. And it is that after, if you could look ahead beyond this measure which is the third of three that relate to PICU, we have two measures, 703

and 702, that are related: one concerning mortality, the other concerning length of stay.

It would seem that it would be -- I mean to me it would make more sense to couple those measures and either do them both before or both after lunch rather than splitting it up.

We also have a member and public comment session. I just wanted to get the views of the group as to whether there was a preference as to how to proceed and deviate from the schedule as it, as it exists.

MR. BENSON: I would suggest we do the two coupled after lunch. We are a half hour off from the schedule anyway.

CO-CHAIR LANG: Excellent. Is that a motion, Ken?

Does somebody want to second?

(Motion seconded.)

CO-CHAIR LANG: Very good.

All in favor.

(A show of hands.)

CO-CHAIR LANG: Excellent.

Opposed? Extensions?

(No response.)

CO-CHAIR LANG: Very good; motion carries. Thank you, Ken.

CO-CHAIR BRATZLER: All right. So we'll proceed with Measure 0343, PICU Standardized Mortality Ratio or discuss -- well, Matt, do you have any initial introduction? Very similar to the previous ones.

DR. SCANLON: Yes, it's a variation on what we've discussed. I mean the concept of standardized mortality ratio is well-published. It's been adopted by the IHI as part of their Move Your Dot campaign. And essentially it's a way of looking at observed over expected mortality based on appropriate risk-adjustment modeling.

And so that's what we've been doing, providing that both at a unit level. It's available at a patient level, although it's really inappropriate to use for an individual patient as it's more of a population metric.

And then for comparisons across clusters of units.

CO-CHAIR BRATZLER: Okay. So our discussants are Bill and Ella.

DR. GLOMB: Ella, do you want me to start or do you want to start? She's not on?

CO-CHAIR BRATZLER: It's the Bill show.

DR. GLOMB: It's me. Okay, let's roll.

So essentially this is using the PRISM III software which has, again, been used, identified, not only measures but in literature, to calculate a predicted mortality based on physiological risk: vital signs, chemistries, et cetera. And that is the numerator.

And then the denominator is all deaths that occur within the unit. So your predicted rate should ideally be the ones who are dying. And those who are not predicted, shouldn't be, to put it succinctly.

It's all children under 18 years of age who have been in the unit for greater than two hours, with at least two consecutive sets of

life-compatible vital signs. So you're not dying when you roll through the doors of the ICU. If you are, you don't get counted in this, in this study.

And moving on to the evidence, again this relates only to that single outcome, mortality. I think we can all agree this is an outcome measure here. And there's discussions here about summary of the evidence and the use of the tool, the PRISM III tool, looking at the morality ratios.

I will just read through that real quickly. There's three caveats from the developer with regard to the literature on the value of using this SMR calculation. Use of a calibrated tool for severity adjustment has been identified as important.

Recent publication in "Critical Care Medicine" 2012, identified that the use of a physiology-based tool in calculating standard mortality ratio is superior than using any administrative data, and that the premature

transfer of patients from the ICU can lower the SMR -- guess people are dying elsewhere -- creating a potential gaming. But use of this, of Measure 0343 in combinations with this one also then addresses that potential.

So the question for the committee, is there at least one thing the provider can do to achieve a chance for the measure results. This doesn't specifically point to something you can do differently other than not let them die if they weren't expected to.

You know, that's, that's the evidence basis.

CO-CHAIR BRATZLER: So again, since this is a maintenance measure, really no change in the evidences is there that -- since the previous endorsement. And we have to decide whether or not we vote on the evidence or not.

DR. GLOMB: That's correct. The only evidence that changed is the addition of another tool. I believe that came along chronologically since the last endorsement.

CO-CHAIR BRATZLER: Right. The updated algorithm.

DR. GLOMB: Right.

CO-CHAIR BRATZLER: Right.

DR. YEALY: Yes, and that's my question. Do we know that this stratification tool retains its precision and accuracy over time or if it drifts away and, therefore, any differences are really not good care but just the tool not performing well? I don't know if the developer has information about that.

DR. SCANLON: We have actually recalibrated the tool, as I understand it, for exactly the reason of drift over time. And while it wasn't published in a peer review, there has been a white paper that I believe was presented at a meeting on that issue by a number of the physicians involved with the software product.

So we are keenly aware of that and the need for recalibration. The original, one of the initial severity of illness tools, PRISM II, which was the obvious predecessor to PRISM III,

suffered that exact issue. So it got to a point where Dr. Pollack and others were telling members of the community not to use it because it hadn't been recalibrated.

So we've got a very large number of patients in the data set now. And because of that we are able to reassess and recalibrate as needed.

CO-CHAIR BRATZLER: Any other questions or comments about the evidence? Bruno?

DR. DiGIOVINE: It's not about evidence. I don't know if you want me to hold. I can hold it. It's about the definition of mortality.

My only question is you talk about gaming with pediatric -- counting a death only if it happens in the ICU. Why not use hospital mortality? Or is there background that you could share with us as to why just ICU and not hospital?

DR. SCANLON: There's a couple different reasons. And again, I would also ask

that the other panel members weigh in if they disagree with me.

The majority of deaths in pediatrics that are non-hospice-related deaths occur in pediatric ICUs. Deaths outside the ICU are relatively rare phenomena. And actually while there were some centers that have published that they had a problem with that, the institution of rapid response teams have made those incredibly rare.

So other than hospice patients, I'll tell you the patients who are other status, who are palliative and moved out of ICUs in a subset of organizations that have those resources, most pediatric deaths occur in ICUs, right or wrong.

CO-CHAIR BRATZLER: Any other questions of conversations about evidence?

(No response.)

CO-CHAIR BRATZLER: So I'm going to ask for a hand vote. Do raise your hand if you want to re-vote the evidence base here.

(No response.)

CO-CHAIR BRATZLER: Okay, seeing none, we will go on and discuss the performance gap.

DR. GLOMB: I'm going to go ahead and talk about gap here.

This current performance data came from 79 PICUs, as we learned earlier, up to 135 using software.

And the unit level standardized mortality rate was fairly broad, between almost none to 2, which is twice what would be expected. So during 2014, the median unit level mortality rate was .92, and the mean level was .97; so pretty much right where you would expect them to be at the 1.0, which is what is predicted based on the tool.

The patient level mean SMR for 2014 was not statistically different from 1. So again, using the tool, looking at the individual patient, predictions were pretty accurate.

For performance over time, similar to the previous two measures there has been no trend noticed, no increasing or decreasing trend over

the three years of data collection for this particular measure.

The disparities are similar to what we've seen all along, that the uninsured children have significantly shorter lengths of stay, greater physiological derangement on admission, and that the hospital mortality was higher.

In terms of race, ethnicity, age groups, gender, insurance, payer, the younger the child, the higher SMR -- though not very much higher. And if you were a teen in a kids' ICU, your rate of mortality was slightly lower, .89. But there were no statistically significant differences in race, in the race or ethnic groups. There were no statistically significant differences in managed care versus commercial insurance, versus Medicare, managed care, self-pay, et cetera. And there were no statistically significant differences found in gender or sex.

So that, at least looking at disparities, there doesn't appear to be a gap. There is a gap if you run this across units.

CO-CHAIR BRATZLER: Yes, I was shocked by the breadth of the gap, or at least the SMR. I think it does need to be in the public domain. But anyway.

Don, I think your name's been up and you didn't have anything. Yes, nothing.

So any other conversation or questions or discussions about gap?

(No response.)

CO-CHAIR BRATZLER: Okay, seeing none, we will go ahead and vote.

MS. AMIRAULT: Performance gap for 0343; 1 being high, 2 moderate, 3 low, and 4 insufficient.

(Voting.)

MS. AMIRAULT: Four high, 11 moderate, 5 low, and 0 insufficient. Based on the percentage we'll move on.

CO-CHAIR BRATZLER: Okay. Discuss reliability.

DR. GLOMB: So under reliability I think we've discussed the logistics of the

measure. There is no calculated algorithm stated. That's part of the proprietary software package here.

With regards to reliability testing, this takes us back again to that inter-rater reliability review, which is an ongoing process.

I don't have anything else. Again, we had those, we had those concerns expressed -- we've discussed them already -- that different units have different characters, different types of disease processes which might alter the predicted unit, unit-based predicted mortality. But I think that's been explained pretty well by the developer because that's done at the patient level.

So I think we were pretty comfortable with reliability.

CO-CHAIR BRATZLER: So any questions or other comments about reliability?

(No response.)

CO-CHAIR BRATZLER: Seeing none, we'll vote.

MS. AMIRAULT: Reliability for 0343; 1 --

CO-CHAIR BRATZLER: So I'm going to ask real quick, do we need to vote? Raise your hand if you think we need to vote on reliability because it hasn't changed.

(No response.)

CO-CHAIR BRATZLER: So we'll go on and discuss validity

DR. GLOMB: Again, these are all standardized definitions. It's an established method, established in the literature, and has been part of multiple measures in the past. It relates to some of the measures we have already discussed today. The IRR seems to speak to that.

In terms of threats to the validity, let's talk a little bit about the exclusions. For the purposes of this measure, and I think within the software utilized to calculate the SMR itself, it's relegated to children under 18 years of age only. Those who have been in the unit over two hours and/or more than two consecutive

sets of vital signs consistent with life.

And they exclude the palliative care cases, those who were likely to die. Those patients are left out of this mix. It also does not include the pre-term infants, post-gestation 36 weeks and below because the tool was not validated on that population.

Again, it kind of takes us a little bit back though to the character of the unit. There are some pediatric ICUs that are an extension of the neonatal intensive care unit. A lot of their volume relates to former preemies moving up either post-operatively or at a cut-off time designated in the hospital. That might play into the statistics a little bit.

I think that our concerns, there is some concern that missing data could also skew results in a small volume PICU and make some big changes. And that, again, doesn't really point us to a specific actionable change that one might see in care.

That's all.

CO-CHAIR BRATZLER: And then I actually, so I have one other issue that I think I'd raised on the work group. And I don't imagine that you guys have looked at it, but I was struck by the fairly wide distribution of the SMR and wondered if we needed, if there needed to be some consideration of hierarchical modeling at the unit level, such as is done by a number of other mortality and other measures for different programs, which might take into account some of that difference in case mix, but also some of that difference in maybe just issues related to pure quality of care.

DR. SCANLON: Yes, we've started to look at that. And I think that's worth exploring.

In part, one of the issues that the algorithms that the calculations are based on were not developed with any hierarchical modeling. So part of it is do we -- understanding what are the implications of applying that. Where we handle it at an

individual unit level is presenting like data to like in terms of unit characteristics.

So when we report data, units pick, for the purpose of ongoing quality, units pick characteristics where they want to match to other ICUs, such as number of beds, do you offer this service versus that service. And then we benchmark against those to provide accurate reporting but not at the national level.

Again, the issue varies. This is not necessarily nationally reported because there is no -- and while we can publish that on the VPS website, there is no market for that data, sadly.

When we looked at variation within the SMR, we certainly found that the standard deviation within the lowest performing set, those with SMRs 1.2 to 3, have actually more variation but and they don't, they drift within that area. But in the lower SMRs, .8 to 1.2 diversity, 0 to .8, there is very little movement of those ICUs, suggesting that those ICUs are fairly stable in their outcomes. But it's the low performers that

tend to have a lot of variation from period to period.

CO-CHAIR BRATZLER: Edgar.

MR. JIMENEZ: Just a quick thing. In the adult population and something that we'll be seeing in the public reporting will be an issue with initiatives playing leapfrog and stuff like that. They're looking at using the standardized reporting algorithms as seeing ICU performance because there needs to be, we probably get the problem has been we haven't had a good tool so far to do it, but there needs to be a stratification, as you were mentioning, of systems, you know, that would allow basics, programs there where they can handle basic things and more sophisticated move over to higher acuity centers.

CO-CHAIR BRATZLER: Don.

DR. YEALY: I have a question for the developers about the palliation exclusion. Are there any guardrails around that with respect to either timing or location? I don't, we don't

have a pediatric unit in our place.

Are there similar measures in an adult way if you -- one way to change your mortality measure is to get everybody on palliation pretty quickly. They still die, they just die differently. That may be a wonderful thing but it's probably not the intent of the measure isn't to improve, you know, the deployment of palliation.

So my question is how do you handle that? Are there some guardrails around when someone's excluded for palliation? Had to be X amount of time before the time of death or location of it?

DR. SCANLON: So I'm going to have to have my behind-the-scenes people send me the operational definition. But my understanding is these are patients admitted to the ICU purely for palliative reasons, not a patient who has altered code status or limitations on support in the ICU.

So if a patient came in because of some devastating event, was cared for, and then

was proceeding to or heading towards death and was placed in palliative status, that was not -- those are not excluded. Those are incorporated in the model so that we don't have that problem of gaming there.

It's patients that -- there's a few centers, and this is where this came in, who would admit patients purely for palliative purposes. And that's where the exclusion applies.

CO-CHAIR BRATZLER: Any other questions about validation, validity testing?

(No response.)

CO-CHAIR BRATZLER: Okay, we can go ahead and vote.

MS. AMIRAULT: Validity for 0343; 1 being high, 2 moderate, 3 low, or 4 insufficient.

(Voting.)

MS. AMIRAULT: Two high, 15 moderate, 3 low, and 0 insufficient. Based on the percentage we'll move on.

CO-CHAIR BRATZLER: And I think the

last question is about -- oh, I'm sorry, yes, feasibility.

DR. GLOMB: So feasibility is similar to the previous two measures. You have access to the software because you're a member of the group or you have an electronic medical record, then you'd be able to get this data.

If you have no electronic medical record you could do -- well, like what Matt said, it was an odious manual review and data entry. And obviously they're paying for the man hours for that. And you could have access to scoring as well. But it is certainly doable in both environments; one easily and one less easily.

CO-CHAIR BRATZLER: Any other discussion of feasibility?

(No response.)

CO-CHAIR BRATZLER: Okay, we'll vote it.

MS. AMIRAULT: Feasibility for 0343; 1 being high, 2 moderate, 3 low, or 4 insufficient.

(Voting.)

MS. AMIRAULT: Three high, 14 moderate, 3 low and 0 insufficient. Based on that percentage we're moving on.

CO-CHAIR BRATZLER: And then usability/use.

DR. GLOMB: So nothing new on the usability. The statistics are the same. This really isn't being publicly reported. Some hospitals might use this voluntarily, and others in California Children's Services if this is posted information.

No changes in the trends over time.

I'm curious if the developer has any thoughts about why we've not seen any movement there based on their data collection.

DR. SCANLON: I'm sorry, movement in the SMR or movement around public reporting?

DR. GLOMB: No, movement in the SMR itself. I'm sorry.

DR. SCANLON: Well, again I think at a unit level the SMR, the median should be around 1 if the calculation is accurate. I think that the

issue is that you wouldn't -- first of all I would say compared to adult mortality in ICUs, pediatric mortality is about 2.7, I think, percent, so it's pretty low to start with. So moving that dot is, in a statistical fashion is pretty challenging.

I think a second issue is the question of whether all deaths are, one, preventable and, two, whether a death is appropriate or inappropriate. There are, at the risk of being controversial, one could argue there are such a thing as good deaths.

But I think the other issue is that because of the lack of a mandate for public reporting of this, while NQF has to date provided a mechanism to publicly report these, or a validated metric, there is not a market force. And while one of the previous speakers spoke to leapfrog demanding that from the adults, the relative cost of pediatric ICUs, while very expensive, pales in comparison to adult care. And so we're not on their radar, rightly or

wrongly.

I think, looking at the range of the SMR that we -- one of the previous speakers commented on, it's disturbing to me that there's not attention to this and that you could actually see that SMR would at least be 1 across the board, if not below 1. And there has certainly been dramatic improvement in CLABSI rates in pediatric ICUs. We're pretty good at keeping those kids alive. Even if they did a CLABSI for this, I don't know that that was a big source of mortality.

So that's a multi-pronged answer. But I think the lack -- where there is room for improvement, the lack of national attention to this, be it by joint commission or any public reporting body with the exception of the California system, certainly has not put any pressure for centers to improve.

CO-CHAIR BRATZLER: James.

DR. O'BRIEN: I think some of this also just goes back to recalibrating or respecifying

the model, is that you're going to wind up then again regression to median, you're going to get back to a 1 as every time you do it, which makes it challenging with these risk-adjusting models as we change the model year over year of knowing whether or not there's actual improvement.

CO-CHAIR BRATZLER: Kenneth.

MR. BENSON: As I understand the purpose of these measures is to help the process improve, improving the quality of healthcare. And what I find troubling in this and the two previous ones has to do with gathering information and presuming the data is correct, and it comes to conclusions that could lead other people to improve their healthcare. that that information is not being gotten out there.

Now, the developer said there's a lack of interest. I don't understand this. I'm just at a loss on how, if we're going to take the time to do this, there is not a mechanism to get it out to the people who could use this to improve their quality. Keeping it locked in a box

doesn't help anybody.

CO-CHAIR BRATZLER: Yes. So I mean I will simply comment that for -- I think Matt made the point very well -- for programs like Medicare, the adult population, which are very, very expensive and they use a lot more ICU care and other extensive care, lots and lots of payers now are mandating, whether it's Medicare or other private payers, mandating that this data be put into the public domain.

That simply hasn't happened, it doesn't sound like, for pediatric ICU care, and there just aren't payers out there mandating that this information -- the payment models haven't pushed it to go to the public domain. So.

DR. SCANLON: This is Matt again. I'm sorry. I couldn't agree more with your frustration. We would love that, not just because we have this software package, but as someone who is passionate about improving the care of children, it's very frustrating to me that there's no national hunger for this.

So, you know, you can use the NQF sway or whatever within the bailiwick of NQF, but use your sway to influence it. That would be great. I would love to have an audience let's say with a national metric that needs to be on the newspaper and whatnot because, you know, the range from 0 to 2 for an SMR is, is disturbing.

DR. O'BRIEN: So I guess for the developer then you mentioned a requirement that for participating centers they have to pay a fee. They have to have a nurse or a physician who is specifically trained and show that they are able to abstract this data correctly. Is it possible for you to include in your participation agreement that those organizations that are going to participate have to publicly report their performance?

DR. SCANLON: I don't know the answer to that. And I say that as I am one of the clinical developers. I am not one of the officers of the program. And so I need to be very careful not to speak outside my range of

expertise and authority.

Is it possible? Yes.

I don't know what that would mean for participation. And I think one implication is that programs that have a poor SMR may just choose to drop out so they don't have to report. I think, you know, saying that's the terms of using the software, again, people who look good will use it, people who have problems, the quickest way is just quit paying to use the software and then you don't have a problem.

And that's a cynical view but I think that's where having nationally-mandated public reporting, be it by payers or other bodies, is what's missing here.

CO-CHAIR BRATZLER: Steve.

DR. GROSSBART: I think one of the quandaries here is that this NQF endorsement is for a proprietary product. And, you know, the cynic in me says this is a marketing strategy. The data is not really being used for what the NQF expects the data to be used for, which is for

public reporting, accountability. And it's just it's we're giving an endorsement and it's going nowhere except within the confines of the vendor's, of the developer's business as a vendor.

And so when we talk about usability, but it may place significant limits on usability.

DR. SCANLON: Well, as a self-avowed cynic I can tell you that the organizations are blinded. But I think that there would be -- to each other in the software because of the legal implications of sharing and whatnot at the present.

I think that it is entirely feasible for us to publish the range of SMRs on the website at a site level but it's not within our current ability without rewriting a bunch of contracts. Which, again, I'm not saying we're not opposed -- or we're opposed to, to identify those organizations.

But that doesn't address the issue of what happens if someone is a poor performer, and

why would they be incentivized to pay to report poorly. And so while you, you know, you can say that we are not being good stewards by keeping this data in a lockbox, that's not the spirit of what this is for.

And again, I'm not trying to sound Pollyannaish here but our goal is to get the data out there. But in the absence of an audience for it, or some external pressure, we're not -- we are currently not able to publish un-blinded results in a public fashion. Maybe that's a fatal flaw in the mechanism, but that result was necessary to get programs up and running in this in the first place.

And this has been an iterative process. As I said, the software has been existence since before 2004. Certainly we can push for that.

I think the pediatric community at large has been in the forefront of transparency. And so we can push that envelope. But, again, you may just drive poor performers, if it's at

the level of the VPS we may just drive poor performers to drop out and hide their problems.

CO-CHAIR BRATZLER: So I'm going to let Todd have the last word here on this before we vote on usability.

DR. DORMAN: Well, I was just going to add a comment that's maybe an academic comment and sounds a little bit strange when you first hear it. But there is -- I'm not aware of data that says that you can use such a measure to compare ICU to ICU. They're internal quality measures. So publicly reporting would create the impression that a unit that is at 1.5 is somehow worse than a unit that is at .7. And I don't believe that there is data that has actually supported that fact. They are different and there are many reasons why they may be different.

And so I think there is an unintended consequence that has led people to be concerned about transparency with these numbers because then the utilization of the number becomes potentially misused based upon our understanding.

And I know that sounds crazy because it's compared to predicted and so, but.

CO-CHAIR BRATZLER: Steve, one last.

DR. GROSSBART: But isn't the ability to create consensus measures that allow for comparisons across different users, different facilities, different populations, I mean isn't that an expectation of NQF consensus measurement development?

DR. NISHIMI: Yes.

CO-CHAIR BRATZLER: So that all goes into your decision about how you vote on usability and use. So any other conversations there?

(No response.)

CO-CHAIR BRATZLER: We'll go ahead and vote on usability.

MS. AMIRAULT: Usability and use for 0343; 1 being high, 2 moderate, 3 low, or 4 insufficient.

(Voting.)

MS. AMIRAULT: Zero high, 8 moderate,

12 low, and 0 insufficient. So based on the percentage this is in the gray zone.

CO-CHAIR BRATZLER: All right. The last one is overall suitability for endorsement. Any other comments or questions?

(No response.)

CO-CHAIR BRATZLER: If you would call the vote.

MS. AMIRAULT: Overall suitability for 0343; 1 being yes and 2 being no.

(Voting.)

MS. AMIRAULT: Nine yes and 11 no. This is in the gray zone.

CO-CHAIR BRATZLER: All right, thank you. Thanks, Matt, for your conversations today.

DR. SCANLON: Yes, thank you.

CO-CHAIR BRATZLER: So we're a little behind schedule. We have one more measure that we were supposed to do before lunch but it's almost time for public comment.

So I think the suggestion up here is to get the pulse of the committee to let's go

ahead and do public comment. And then, as we did yesterday, get our lunch and then do a working lunch to keep moving so that we can do the best we can to get all the measures done today before we leave.

Does that seem reasonable?

MS. BAL: Operators, any public comment?

THE OPERATOR: At this time if you would like to make a comment, please press star then the number one.

(No response.)

THE OPERATOR: There are no public comments at this time.

CO-CHAIR LANG: Also, we're going to draw straws to determine whether the term that each of us will serve on the Standing Committee will be two years or three years; is that correct?

MS. BAL: That's correct.

CO-CHAIR LANG: So it's either two or three. I don't think there are any other

possibilities, are there?

CO-CHAIR BRATZLER: Two.

MS. BAL: Oh, sorry. Your name and the years.

CO-CHAIR BRATZLER: Dale Bratzler, two.

CO-CHAIR LANG: David Lang, three.

DR. NISHIMI: Into the mike you need to announce it.

MR. BENSON: Ken Benson, three.

DR. DORMAN: Todd Dorman, two.

DR. GLOMB: William Glomb, three.

MS. WEST: Chana West, two.

DR. OHTAKE: Patricia Ohtake, three.

DR. ELLIOTT: Kim Elliott, two.

DR. RILEY: Crystal Riley, three.

DR. YEALY: Don Yealy, two.

DR. LAMPONE: Thomas Lampone, two.

DR. COLLINS: Curtis Collins, three.

DR. DiGIOVINE: Bruno DiGiovine, two.

DR. BAULDOFF: Gerene Bauldoff, three.

DR. JIMENEZ: Edgar Jimenez, two.

DR. GROSSBART: Stephen Grossbart,

three.

DR. O'BRIEN: Jim O'Brien, three.

DR. MURRAY: Richard Murray, three.

DR. SCHINDLER: Christine Schindler, three.

MS. BAL: So for the people on the phone, for Mitch Harris is two; Susan Pollart is two; David Stockwell is three; and Ella is two.

DR. NISHIMI: Thank you. Lunch is ready.

CO-CHAIR BRATZLER: So let's take about 10 minutes, 10-15 minutes max and then by 12:30 we'll get back and get started again.

MS. BAL: Adam, we'll be back at 12:30 and we'll go over your measures. Thank you.

(Whereupon, the above-entitled matter went off the record at 12:15 p.m. and resumed at 12:30 p.m.)

CO-CHAIR BRATZLER: All right. Is everybody ready to get started again? And, Adams, are you on the line?

(Pause.)

DR. NISHIMI: Operator, is Adams Dudley on the line?

THE OPERATOR: No, they haven't joined yet.

(Pause.)

CO-CHAIR BRATZLER: So, we're waiting on our developer to get on the line. If it's okay, I guess we can go ahead and start.

Okay. So, this is a maintenance measure. So, our two discussants are Edgar and Patricia and we'll go ahead and do the conversation about evidence. And then once Adams rejoins the call, we'll ask him to give an overview of the measure.

DR. OHTAKE: Thank you. So, this is Measure 0703, Intensive Care In-Hospital Mortality Rate. And it's for all adult patients admitted to the ICU and the percentage of patients whose hospital outcome is death, both observed and risk-adjusted mortality rates are reported with predicted rates based on the Intensive Care Outcomes Model-Mortality.

And so, the rationale is that death is the reason -- preventing death is the reason why people are admitted to the ICU. This is an $81 billion enterprise to care for our patients in the ICU and we certainly want to be sure we're doing the best -- providing the best care possible.

The numerator is all of -- the total number of eligible patients whose hospital outcome is death. And the denominator is the total number of eligible patients who are discharged, and this includes both deaths and transfers out to other hospitals. And this is an adult measure. So, individuals less than 18 years of age at the time of ICU admission are excluded.

ICU readmissions are excluded. Short stays in the ICU are excluded, less than four hours, or primary admission for trauma burns or immediately post-coronary artery bypass grafting, or admitted with a diagnosis of rule-out MI.

It's an outcome measure and currently

it's being -- the data is being abstracted from paper medical records, although the developer indicated that an eMeasure is in development and they hope to have that available in 2016. And the level of analysis is at the level of facility. It was first endorsed in 2001.

As far as the evidence goes, the developers have indicated that there was some new evidence. But when the NQF staff reviewed that, there didn't seem to be a lot of updated literature available.

And when we discuss this in our workgroup, our workgroup agreed with that that the evidence seems to be consistent with what was available at the previous endorsement.

DR. JIMENEZ: Nothing to add, except that it is paired with 0702, right? Just to keep in mind as we go through that this is a paired measure.

DR. OHTAKE: Thank you.

CO-CHAIR BRATZLER: Todd.

DR. DORMAN: So, really a point of

clarification. So, it's not really an evidence question, but this point of it moving to an eMeasure, does that impact our discussion?

And if we -- will they then dovetail that under this and be allowed to continue with if we approve this as an eMeasure, or does that have to come back?

DR. NISHIMI: You aren't approving it as an eMeasure. You're approving what's before you.

They would have to bring an eMeasure, because there are some very specific criteria.

DR. DORMAN: Okay. Thank you.

CO-CHAIR BRATZLER: So, Operator, can you tell us -- oh, I'm sorry. James, go ahead.

DR. O'BRIEN: So, the one evidence that I might suggest that has some relevance is a paper from Jack Iwashyna in Michigan, which looks at confounded by indication risk adjustment.

The foundation of that is based around observational studies, but I think has applicability when it comes to benchmarking, but

suggests that when there is -- is truly confounded by indication, which may be the case in ICUs that sicker patients may be transferred or present to different hospital.

Risk adjustment really doesn't help you to overcome that and you still wind up with misclassification as far as the actual underlying effect.

CO-CHAIR LANG: I just wanted to clarify since Edgar mentioned this, that the next measure that we're going to be considering concerning length of stay in the ICU, it was stated on our phone call by Adams Dudley that that measure, the length of stay measure, is only to be used if you're also doing mortality reporting. So, that clearly is not a standalone measure.

DR. DIGIOVINE: And I just want to make a comment to James' point. I just want to make sure I understand it.

I think Jack Iwashyna's research is around transfers from another hospital, which I

understand is an exclusionary criteria for this measure. I just wanted to be sure that was correct.

DR. JIMENEZ: That's correct.

DR. O'BRIEN: So, I think the other applicability, though, I can certainly see with a tertiary care hospital in urban environment, there may be a different indication for patients who are admitted to ICUs there relative to a hundred-bed hospital in a rural area.

DR. NISHIMI: Adams, have you rejoined?

THE OPERATOR: Adams has not rejoined yet.

DR. NISHIMI: I think we just have to go ahead with the --

CO-CHAIR BRATZLER: So, at this point we have to decide whether we're going to vote on evidence. The subcommittee actually recommended that we didn't need to vote on evidence, but I'll leave it up to the group. So, raise your hand if you think we should go ahead and vote on evidence

or move forward.

Do you think we should vote on evidence? So, we have two votes. So, are we doing this as a majority or -- majority, okay.

So, I guess we will not vote on evidence at this point and go on and discuss gap.

DR. OHTAKE: Okay. So, the developers provided information of performance scores from data collected between 2010 and 2011. And it looked at almost 70,000 patients and there were just over 8,000 deaths for an overall ingested mean mortality rate of 11.67 percent.

This is down 2.18 percent from the previous data period. So, there has been some movement in performance. However, there still is felt in our discussion that there's definitely a gap that -- a performance gap that can be addressed by continuing to use this measure.

As far as disparities, there is -- were not disparities of actual patient measurements reported. However, the literature information identified -- that was presented

identified that there is disease-specific racial variation among African-Americans.

There's also disparities for the elderly. Particularly, the older women seem to fare worse than men. And also based on insurance status as well.

Edgar, is there anything else you want to add?

DR. JIMENEZ: Nothing to add.

DR. DUDLEY: Hi. I just wanted to let you know I'm here. This is Adams Dudley.

CO-CHAIR BRATZLER: Yes, Adams, this is Dale Bratzler. So, we're going to in just a moment have you give a brief overview of the measure.

DR. DUDLEY: Okay.

CO-CHAIR BRATZLER: Don, can you --

DR. YEALY: Just one question for the developer. The most recent data we have are four plus years old. There's nothing more recent that we can sink our teeth into regarding the gap?

DR. DUDLEY: No, we do not have more

recent data.

DR. YEALY: Any particular reason why that is?

DR. DUDLEY: We have not -- right now we are not collecting this data. We're waiting for CMS evaluation for the possibility of national adoption. So, there isn't a current dataset.

CO-CHAIR BRATZLER: Bruno.

DR. DIGIOVINE: Yes, just sort of relative to what we talked about last time, your -- what is being shown here is crude mortalities and you talked a little -- we talked a little bit about a prediction score.

So, is the measure a raw mortality score, or a standardized mortality score?

DR. DUDLEY: It's a risk-adjusted mortality score.

DR. DIGIOVINE: Can you elaborate on what you mean by that?

DR. DUDLEY: We use the measures that were originally reported in what's called the MPM

Model, Mortality Prediction Model. It's one of three major models used in ICU risk adjustment.

There are hundreds of published papers with the three models. And what we did in developing the model that was used for public reporting is to compare the work required to collect the data for each of the three models and also to compare the performance assessments one would generate using each of the three models.

We found very high correlation between the performance assessment for the three models above 0.9. And found that the model we're currently using required much less work to get the data collected than other models, less than 15 minutes a patient.

So, in the balance of the burden of data collection and the benefits in terms of prediction, it seemed like it fell towards the model that we're currently using.

DR. DIGIOVINE: I'm sorry, Adams. Maybe I didn't ask my question clearly. So, if you're MPM and you have a predicted mortality for

every patient, why don't you report for our review standardized mortality ratios rather than just crude mortality rates?

DR. DUDLEY: Report to you for your review, or report to the -- we report both to the -- to the hospitals that are part of the program, and we'll report to you anything you would like.

DR. DIGIOVINE: So, the numerator is the number of deaths. The denominator is the number of patients.

DR. DUDLEY: Right.

DR. DIGIOVINE: So, the score would be a percentage. How does that percentage then become risk-adjusted?

DR. DUDLEY: So, we calculate -- so, that's the raw data. And then for each patient we have all the variables that go into the risk-adjustment model we calculated and expected mortality rate. And we also report and observe to expected mortality rate.

CO-CHAIR BRATZLER: So, other questions about performance gap?

(No response.)

CO-CHAIR BRATZLER: So, before we vote on performance gap, Adams, I was wondering if you wanted to just give a brief -- we kind of jumped into the measure without you, but if you wanted to give any brief introduction to the measure?

DR. DUDLEY: And I apologize for not being here at the beginning. I didn't catch the schedule change.

So, the -- obviously mortality in the ICU is extremely important. And we decided in -- a group of volunteer hospitals in California started measuring this in 2006 and publicly reporting it in 2007.

And along with the public reporting, developed some performance improvement collaboratives. And that is -- it's from those groups that this data comes.

It eventually expanded to every hospital with at least 200 beds in California.

CO-CHAIR BRATZLER: Is it being used outside of California, Adams?

DR. DUDLEY: It is not being used outside of California.

CO-CHAIR BRATZLER: Okay. All right. So, we've had a conversation about performance gap.

Any other conversations about performance gap? And if not, we'll go ahead and vote on performance gap.

MS. AMIRAULT: Performance gap for 0703. One being high; two, moderate; three, low and four, insufficient.

(Voting.)

MS. AMIRAULT: If everyone could just do it one more time? Sorry about that.

(Voting.)

MS. AMIRAULT: Five high, 13 moderate, two low and two insufficient. Based on the percentage, we'll move along.

CO-CHAIR BRATZLER: Okay. And so, we'll move to a discussion of reliability.

DR. JIMENEZ: With reliability, there was about 94 percent with a range of 85 to 97

percent.

The difference in performance can be identified and requires quality data collection, though. I mean, it is an intensive, manual, at this point in time, operation.

And then at the level it performs, it can be varying in quality outcome. So, based on our algorithm, it can be rated as high in quality.

DR. OHTAKE: I just add that they looked at inter-rater reliability, as Edgar said, with trained auditors compared to the hospital's data abstracters.

CO-CHAIR BRATZLER: And did -- as we've discussed with previous measures, is there any ongoing periodic reliability testing to participate in the database?

DR. OH TAKE: I think we'd have to ask the developer that. But from the measure, they indicated that this was a critical point, as you bring up. So, perhaps the developer could tell us about any ongoing quality checks on the

particular hospital abstracters or data collectors.

DR. DUDLEY: It is important. And, actually, even when someone new comes on to learn to collect the data, I would -- we strongly recommend training beforehand.

The training isn't terribly onerous, but it is important. And then we would recommend approximately every year or two, we found with every two-year auditing that the data stay pretty well on track.

CO-CHAIR BRATZLER: James.

DR. O'BRIEN: Can the developer just clarify -- I saw that transfers into a hospital are excluded from the denominator. Transfers out of the hospital, though, look like remain and are considered a patient who survived the hospitalization; is that accurate?

DR. DUDLEY: That is correct.

CO-CHAIR BRATZLER: Any other questions/comments about reliability?

(No response.)

CO-CHAIR BRATZLER: Okay. All right. So, we'll go ahead and vote on reliability.

MS. AMIRAULT: Reliability for 0703. One being high; two being moderate; three, low or four, insufficient.

(Voting.)

MS. AMIRAULT: Four high, 15 moderate, three low and zero insufficient. Based on the percentage, we'll move along.

CO-CHAIR BRATZLER: All right. Validity testing.

DR. JIMENEZ: To validity. Great. The risk adjustment follows a Bayesian statistical model. The areas under the curve are 0.81. And if you correlate with others, the system measure as 0.92. And they allow differentiation across the measured entities.

The exclusion criteria are appropriate through SDS parameters. And the usual groups that have been excluded also like burns, trauma, cardiothoracic are included in the database and usually are seen by more specialized hospitals,

too. So, it qualifies for moderate validity.

I don't know, Patricia, if you have anything else.

DR. OHTAKE: I have nothing further to add other than the risk adjustment variables. There are 15 of them that are pretty commonly used in risk adjustment with this particular patient group.

CO-CHAIR BRATZLER: So, any committee conversation or --

DR. O'BRIEN: Does the developer have any information on the number of patients who are excluded due to transfer and then also the number of patients who are transferred out and considered alive that are included in the dataset?

DR. DUDLEY: Both are quite small. We actually -- in the -- the issue of transfers came up when this was first endorsed. And the particular concern was raised by the American Thoracic Society that academic medical centers would be penalized by accepting in transfer

patients who were particularly ill.

The -- we analyzed at that time, we have not reanalyzed since, the impact of excluding all transfers, or just transfers in each direction, and the performance score correlations were about 0.95 with -- comparing with transfers versus without transfers.

And so, the decision was made since the area of particular sensitivity was transfers into centers that were accepting high-risk patients, that was the part that we decided to exclude.

But as of the analyses when we went through the initial endorsement, there is essentially no impact on performance ratings.

Would this be an opportunity for me to add another piece of information back on performance gap? Is that okay?

CO-CHAIR BRATZLER: You can. It's already passed on that criterion.

DR. DUDLEY: Sure. I just wanted to get out there for future consideration, because I

know there will be more levels, I think perhaps -- I was surprised by the low number of people who thought the performance gap was high.

And then I thought about the context and in reality it might look that way because we've been measuring this and reporting it for six years, but when we started, there were large groups of hospitals that were three-fold difference in risk-adjusted mortality.

So, I think that in thinking about the performance gap, it might be worth considering we're assessing that in a group of hospitals that have been working on this for a long time, but where talking about application of it almost all the hospitals in the country haven't yet had this applied. And so, the performance gap will be much larger in that as it were a naive population.

CO-CHAIR BRATZLER: And that will be relevant to our conversation about use.

Todd, I know you had --

DR. DORMAN: Yes, I'm not sure if you

-- if you stated this earlier or if I missed it in the paperwork. Which version of MPM are you using for the -- Three? 3-0?

DR. DUDLEY: Actually, what we recommend is using the variables and then recalibrating the coefficients to the new population frequently.

So, we actually recalibrate quarterly, because treatment changes and sort of the essence of what puts someone at risk doesn't change, but the relative contributions of particular risks do change over time.

And so, we actually recommend that you recalibrate to the new population and the new data frequently.

CO-CHAIR LANG: Yes, for your calculation algorithm you stipulate on S-18 that the death rate for each hospital is adjusted according to average case mix.

Could you elaborate on that for us so it can help me understand better how that's done?

DR. DUDLEY: Sure.

CO-CHAIR LANG: That would seem to be a critical element of this measure.

DR. DUDLEY: Sure. So, I'll explain the -- I'll start back with the last comment about the recalibration.

So, what we do is we take all the data from all the patients across all the hospitals and calculate the impact of each risk factor across all of the hospitals.

And then each of those variables contributes to for an individual patient, calculated risk of mortality.

And then looking at a particular hospital you add up the predicted risk for each of its patients. And that gets you an estimate of what the predicted mortality should be.

And then you look at the number of observed deaths and compare the observed to expected.

So, for hospitals that have more risk factors, you're applying to each of the individual risk factors the average weight of

that across the state as recalibrated each quarter. And then, therefore, they're getting a higher predicted death rate per hundred patients that they have and capturing their severity of illness in that way.

Is that clear?

CO-CHAIR LANG: Yes.

CO-CHAIR BRATZLER: Any other questions or comments about validity?

(No response.)

CO-CHAIR BRATZLER: Okay. Seeing none, Janine, go ahead and vote.

MS. AMIRAULT: Just a note that the highest eligibility for validity for 0703 is moderate. So, options are two, moderate; three, low and four, insufficient.

(Voting.)

MS. AMIRAULT: If you could just do it one more time? Thank you.

(Voting.)

MS. AMIRAULT: Zero high, 13 moderate, nine low and zero insufficient. And based on the

percentage, it's grey zone.

CO-CHAIR BRATZLER: Okay. We'll go ahead and discuss feasibility.

DR. JIMENEZ: Okay. With the feasibility, this is pretty much basically chart abstraction and manual. So, that's a burden in the collection, but the usefulness has been referred to as outweighs the burden of the collection.

And that's pretty much what I have. I mean, it's the, you know, it's the, I mean, it has been used in California extensively with no problems at all reporting, except for the manual collection.

DR. OHTAKE: I'd just like to add to bring attention to the fact that there's no costs or licensing requirements with this particular measure. So, it's freely available.

CO-CHAIR BRATZLER: Is there any sampling in cases, or is this a hundred percent of the population at the ICU?

DR. DUDLEY: We request the first 400

consecutive -- sorry -- 100 consecutive patients per quarter for a rolling annual sample of 400 patients.

CO-CHAIR BRATZLER: Thanks. Any other conversation/questions about feasibility?

(No response.)

CO-CHAIR BRATZLER: Okay. Janine.

MS. AMIRAULT: Feasibility for 0703. One being high; two, moderate; three, low and four insufficient.

(Voting.)

MS. AMIRAULT: Two high, 14 moderate, six low and zero insufficient. And based on the percentage, we will move forward.

CO-CHAIR BRATZLER: All right. Then usability and use.

DR. JIMENEZ: The measure has, like we said, has only been used in California. It is expected to have some acceptability from CMS as it moves to an electronic format, but it's not there yet.

So, besides the California

participating hospitals, there hasn't been any other usability.

Now, I know from discussions that I've had with the Leapfrog Group, that they are looking for some measures that would eventually -- would -- in the public format of reporting would supply information on ICUs for the intensivist physician standard, but -- and this has been looked at by the -- will probably have to be in electronic format before that.

DR. OHTAKE: I just add that the developer stated that it was -- the measure was discontinued use in 2013 because there are other measures -- or in favor of measures required by CMS.

So, I guess my question for the developer is, is this measure currently available for use as paper-abstracted literature should --

DR. DUDLEY: Yes. All of the requirements to collect the measure are available for free.

DR. OHTAKE: Thank you.

CO-CHAIR BRATZLER: So, I just want to make sure I understand. So, in 2013 because of competing priorities, it's just not being -- I think, Adams, you said it was being voluntarily collected by a number of California hospitals and actually publicly reported by some.

DR. DUDLEY: Right.

CO-CHAIR BRATZLER: That's not happening anymore?

DR. DUDLEY: The concern was that the CMS data collection burden was increasing and would push hospitals on their resources and they were going to focus on those things, but we want to move this over to a CMS-preferred format.

CO-CHAIR BRATZLER: Any other questions/comments about use or usability?

(No response.)

CO-CHAIR BRATZLER: All right, Janine.

MS. AMIRAULT: Usability and use for 0703. One being high; two, moderate; three, low and four insufficient.

(Voting.)

MS. AMIRAULT: One high, 11 moderate, 10 low and zero insufficient. And based on the percentage, this is in the grey zone.

CO-CHAIR BRATZLER: All right. Our last question is overall suitability for endorsement.

Any other -- yes, Bruno.

DR. DiGIOVINE: So, I'll make my -- Todd made the comment earlier about using mortality ratios as quality. And I just happened to come across a quote from Dr. Hofer at the University of Michigan who said that publication of hospital mortality rates misinforms the public about hospital quality and described them as seriously inaccurate.

So, I just thought it would be -- I think that's weighing on -- certainly weighing on how I vote and thought the developer at least should have an opportunity to respond to that.

DR. DUDLEY: Sure. I think -- so, I publish with Tim on that topic and I do think it's hospital to do hospital mortality reporting

wrong.

That's why you see us; A, having tested the various available risk adjustments very carefully to -- risk adjustments allow us very carefully to be sure that we -- the one we used performs adequately, and; B, using a relatively largish sample size.

So, when you end up with 400 patients per hospital, that's a much larger sample size than what you're getting with most of the currently publicly reported measures. So, most hospitals that have MI reports and so forth from CMS don't have 400 patients in the measure.

So that we reduce the risk of misclassification both by having better risk adjustment by far than the currently used CMS models and by having large enough sample sizes that the probability of risk adjustment is -- I'm sorry -- of misclassification is much lower than would be expected.

CO-CHAIR BRATZLER: James.

DR. O'BRIEN: I think probably in a

similar vein working in a health system that has 12 hospitals that are connected by an eICU in which we're collecting APACHE IV measures, we see that in our lowest acuity hospitals despite adjustment, those have the lowest also observed to expected mortality rates.

Part of that is due to the fact of them being extreme outliers and the indication bias for where they wind up being. And then the performance of the risk adjusting focusing on the middle part. And so, these extreme outliers of hospitals even within, again, a 12-hospital system, it just doesn't perform well. And so, it doesn't even have face validity within our system.

DR. DUDLEY: We -- again, we're publicly reporting with this measure and we didn't have an instance of the hospitals complaining to the media that we were being unfair to them or that we had misrepresented their patient population.

I think it's possible that within a

system you get even more dramatic shifts in where the patients go because it's within system issues and that could be perhaps dealt with by transfer rules.

But overall despite public scrutiny, we did not actually have a problem with hospitals complaining about the data, the accuracy of the data, the accuracy of other people's data or how they were rated.

Again, I think a lot of that comes down to making sure people are carefully trained to collect the data correctly and that you're auditing them to make sure they're doing that as well.

CO-CHAIR BRATZLER: And you have a big sample size per hospital.

DR. DUDLEY: That does help.

CO-CHAIR BRATZLER: Yes. Any other conversations before we vote on suitability?

(No response.)

CO-CHAIR BRATZLER: All right, Janine.

MS. AMIRAULT: Overall suitability of

0703. One being yes, and two being no.

(Voting.)

MS. AMIRAULT: Would you mind just one more time? Thank you.

(Voting.)

MS. AMIRAULT: 13 yes, nine no. And based on the percentage, this is also a grey zone.

CO-CHAIR BRATZLER: All right. Thank you guys very much. So, we'll move on. We're about back on schedule, and go to Measure 0702, Intensive Care Unit Length-of-Stay.

Adams, we'll let you introduce the measure and then our two discussants will be Todd and David.

DR. DUDLEY: Sure. So, first thing, this measure comes from the exact same variables and the exact same data collection as the prior measure so that one ends up with both an outcome -- a clinical outcome and an efficiency measure from the same data collection.

We do not recommend that this measure

be used without the mortality measure, because we wouldn't want people to focus only on length of stay without having the clinical outcome also on the dashboard because of potential unintended consequences.

The approach to risk adjustment is similar in that the exact same variables are used. The modeling is slightly different because it's a continuous variable that's skewed instead of a binary outcome variable.

Because it's the same data, it's the same training and it's the same auditing that we recommend. And I think that's all I have to say unless there are other questions.

CO-CHAIR BRATZLER: All right. Todd.

DR. DORMAN: I'm going to start. So, Intensive Care Unit Length-of-Stay is the measure title. It's an outcome measure. Paper medical records. Level of analysis is at the facility level.

It is a maintenance measure that was previously endorsed in 2011. It is, I guess,

technically a paired -- or at least it's recommended to be paired with the previous measure that we just finished discussing.

Under evidence, which is where we're starting, the developer tested that there is new evidence since the last NQF review in 2011, but only really provided explanatory information and we'd like to hear a little bit more about that.

It should be pointed out that I guess the phrase was used, "efficiency." This is really connected to a contributor to cost, not patient outcome at least as presented here.

And I think I'll stop there. I don't think we saw any -- other than the comment from the developer that there's new evidence that they didn't provide a lot of information, I don't think we were aware of any new evidence.

CO-CHAIR BRATZLER: Adams.

DR. DUDLEY: New evidence about -- I'm sorry.

DR. DORMAN: The evidence.

DR. NISHIMI: New evidence for the

underlying rationale.

DR. DUDLEY: So, we -- the new evidence -- perhaps I misunderstood the form. The new evidence is just their updated validity and reliability reports that we discussed with the last variable -- the last measure, I mean.

CO-CHAIR BRATZLER: All right. Since this is a measure that's for maintenance, does the committee want to vote on evidence, or not?

If you do, raise your hand. Okay. We'll go ahead then and discuss gap.

CO-CHAIR LANG: Data were collected between 2010-2011. 224 hospitals contributing, about 70,000 patients and more than a quarter million ICU days.

The overall unadjusted mean length of stay was 3.4 days with the standard deviation of 0.8 days.

The developer does indicate that disparities exist among different population groups, diagnosis, level of care and reported racial disparities such that African-Americans

have an adjusted ICU length of stay that was significantly shorter than that of Caucasians.

CO-CHAIR BRATZLER: Anything else, Todd?

DR. DORMAN: I'd like to hear the developer comment a little bit on the size of the gap here and the difference in interquartile ranges is 2.8 to 3.9, which is quite small.

DR. DUDLEY: Again, so this gets at the gaps that you see in data that's been reported to hospitals for six years.

In the beginning, the gaps were quite a bit larger, but I think overall we found that the length of stay gaps were less than the mortality gaps.

So, in the original mortality pilot study that we did back in 2006 -- 2005 and '06, there were threefold differences between stable groups of hospitals, you know, ten hospitals -- compared the top ten hospitals with the bottom ten hospitals. There was threefold difference in observed-to-expected mortality ratios. There

were not gaps that large in the length of stay.

Nonetheless from a payer perspective the gap that is present even now is very important. It's a very big difference to pay for an extra -- even an extra half day on average of ICU stay versus ward stay is very large.

DR. O'BRIEN: If the developer maybe can just comment on that in the world of DRGs how the payer winds up being on the hook, and then also maybe considering Jeremy Kahn's work looking at the effect of reducing ICU length of stay and whether that actually impacts total cost.

DR. DUDLEY: The -- taking the second one last, of course it's all part of a, you know, it's ICU days are one input into a total cost of care, but I think they're a very expensive input.

And so, to the extent that one can optimize the use of that input, one is better off.

Sometimes payment is based on a DRG. That is true. In that case in the short term, the savings don't accrue to the payer. But in

the long term, they recalibrate DRG rates based on cost reports.

In addition, there are many payers, for instance, many Medicaid programs that are paying still on a per diem basis where the per diem is higher if it's an ICU day than if it's not.

So, we have a very eclectic approach to payment, but in the long term it all comes out in the end. We want to optimally use our resources and if there's a way to use less and get the same clinical outcomes, which is why we always would use this with the mortality measure, then that ought to be sought.

So, we manage to reduce statewide risk-adjusted mortality substantially over the reporting period while reducing the variation of length of stay.

DR. DORMAN: Thank you. So, the question I'm going to ask really I was going to deal with under one of the other sections. But since you brought up that the gap is growing, I

feel like I have to ask it now.

I'm trying to get at a better understanding of what's being tracked and how chart reviewers are figuring out whether a patient in the ICU is actually an ICU patient or a patient in the ICU.

And I don't think that that's something that's commonly documented in a note. I don't know how they would abstract that. And so, it would not surprise me that ICU length of stay is growing, but with patients who are actually floor status patients who are staying in the ICU.

So, since you brought it up under gap even though it's probably closer to validity, I'm trying to understand how you know these patients are actually ICU patients and not step-down patients, telemetry patients or floor patients who happen to be physically located in an ICU.

DR. DUDLEY: So, it's part of the training. It's something that we strongly recommend is a matter of ongoing discussion among

the hospitals participating in any system.

We -- actually, let me go back for a second. I don't think I said that the gap is increasing. In fact, the gap isn't increasing. It's actually decreasing because we've been reporting it for a long time. What I meant was that the mortality gap was greater than the length of stay gap.

But at any rate, who meets criteria for being an ICU patient is a very important issue. And we discuss with the hospitals what constitutes telemetry, how they've renamed their units. We try not to just use walls and instead talk about the intent of the care.

And the -- part of the rationale for doing it a hundred consecutive patients at the start of each quarter is to get the data capture in real-time so that you don't have to go back and figure out from the chart backwards was the patient in the ICU.

eMeasures are also bad in this respect in that it can be difficult to tell when the

patient officially left ICU-type care.

So, this will be an ongoing issue that applies to any kind of ICU measure, but, again, was something that the hospitals reporting this were not as concerned about. Certainly didn't complain to the media about any of this or anything.

CO-CHAIR BRATZLER: Bill.

DR. GLOMB: I was going to ask whether you'd consider using one of the commercially available authorization and claims programs like InterQual or Milliman to help you make those determinations.

Those are used by everybody from payers to hospital claims departments now. They're universally reproducible in their results. It's not geared toward the payer base or the provider base and it seems like that might solve part of that problem.

I'm frequently having to tell folks who are unhappy with their claims resolution that geography, where the patient is located in a

hospital, is not what constitutes level of care. It's level of care that constitutes level of care.

And so, that would avoid this question about whether someone truly meets the ICU level of care or not.

DR. DUDLEY: Right. So, the issue there is that NQF measures can't have any proprietary component to them or aren't supposed to have any proprietary component to them. And CMS isn't supposed to adopt things that have a proprietary component to them.

So, we have tried to mimic those without using official intellectual property of someone else.

DR. NISHIMI: I just wanted to clarify for you, Adams, actually NQF has for quite a while now accepted measures that have proprietary components.

DR. DUDLEY: Okay.

DR. GLOMB: And CMS does endorse, I think, both -- I know InterQual, but I think they

also endorse Milliman.

CO-CHAIR BRATZLER: All right. So, I'm going --

DR. DUDLEY: Well, that could be here.

CO-CHAIR BRATZLER: I'm going to bring our conversation and make sure we're focused on gap before we go on to some of these issues that I think are around validity.

Any other issues around gap?

(No response.)

CO-CHAIR BRATZLER: Then let's go ahead and vote. Janine.

MS. AMIRAULT: Performance gap for 0702. One being high; two, moderate; three, low and four insufficient.

(Voting.)

MS. AMIRAULT: Two high, ten moderate, ten low and zero insufficient. And based on the percentage, this is grey zone.

CO-CHAIR BRATZLER: Reliability.

DR. DORMAN: So, the developer attest that there's been no change in the

specifications. The measure is risk-adjusted. The developer used data element reliability testing that's been published with about 11,000 patients out of 35 California hospitals.

Inter-rater reliability was assessed and was 91 and a half to 98.8 percent. Reliability testing was both at the measure squared and the data element, as I mentioned.

I'm not sure that that's -- let's see. I think I'm going to leave it at that. And the algorithm comes out to being eligible rating as moderate.

CO-CHAIR BRATZLER: Bill, did you have a comment?

(Off microphone comment.)

CO-CHAIR BRATZLER: So, I assume reliability here is largely the same as the measure we discussed previously, same data elements.

Any other questions/comments about reliability?

(No response.)

CO-CHAIR BRATZLER: Okay, Janine.

MS. AMIRAULT: So, reliability for 0702. Two, moderate; three, low or four, insufficient.

CO-CHAIR BRATZLER: Remember you can only vote two, three or four here.

MS. BAL: I'm sorry. Before we continue, I just want to make an announcement that Mitch Harris is actually conflicted with this measure. So, he won't be voting. Thank you.

(Voting.)

MS. AMIRAULT: One high, 14 moderate, seven low and zero insufficient. And based on the percentage, we can move along.

CO-CHAIR BRATZLER: So, we'll discuss validity.

CO-CHAIR LANG: Validity. So, again, this is not an eMeasure, but the data are obtained from paper records.

The developer stipulates that agreement was assessed between trained auditors,

the authoritative source and hospital data collectors for all individual risk model elements and the percent agreement was 94 percent.

There are a number of potential threats to validity, including a handling of transfers, which was mentioned, and appropriate case mix and risk adjustment.

And there are a number of individuals who I should preface my comment by stating have more content expertise than me, but -- or than I do, but my understanding is that APACHE has some limitations for risk adjustments as reflected on Page 27 where there's -- or 28, rather, where there's an r-square of 0.42.

The argument here that there's a strong correlation coefficient r of 0.89 between a simplified method and APACHE and I guess if the developer could help me understand the -- whether this fully addresses the lower independent r-square of 0.28 for the simpler model.

DR. DUDLEY: So, our focus is on the public reporting of this. And so, if your -- and

where it's the r-squared -- and that -- sorry.

Our focus is on the public reporting and that focus then is what's the score for a hospital?

R-squared is a measure of the explanation of variation in predicted scores for a patient.

So, if there is noise at the patient level that cancels out when you get up to the hospital level, it's possible to have a lower r-squared at the patient level and still have good correlation between two different methods of assessing risk and performance.

And so, the r of 0.89 is between the simpler model and the APACHE model and is an indication of very high correlation for the hospital-level score between the two models.

And since our focus is on what in the end do we say about the hospital's performance, we find that helpful.

CO-CHAIR BRATZLER: Other comments about validity?

DR. LAMPONE: I just had one comment, and this was brought up earlier about the actual length of stay.

I think we're talking about the validity and reliability of the risk adjustment. So, whether you have a patient that is admitted to the ICU, continues to need to be at -- is appropriate for that care setting.

Moving forward, how do you know to your point that you have patients in the ICU that weren't continued ICU stay, which seems to be the driver of the -- of the basic question that the measure is supposed to be answering.

And without having some structured criteria, I think you get a lot of noise in the measure, because there may be patients waiting for a bed to open in the step-down unit, patients where they're waiting on care decisions or things aren't being delivered -- care isn't being delivered efficiently that may skew this.

So, I wonder not only the committee's ideas about that, but also the developer and

whether that is felt to be relevant enough to play into this measure.

DR. DUDLEY: So, if the patient -- so, there are two types of I'm in the ICU, but in the heads of the ICU doc, for instance, I'm not an ICU patient. There are two types of situations you just described.

If the patient has been -- it has been agreed that the patient should be discharged and go to another floor and it's just that the bed isn't there, then those orders are written and our data collectors would not count the patient as in the ICU.

If some care decisions have not been made or other things have occurred that are -- potentially reflect inefficiency and it is not clear that the patient is to leave, then we count them and I think the perspective -- so, these measures were developed and vetted by a group, a multi-stakeholder board that oversees the whole thing and it includes consumer and payer groups as well.

And I think their perspective would be if you haven't made the care decisions that another hospital would have, then we want to hold you responsible for that, because we're paying for the lower efficiency and getting people to talk to each other or whatever it is that it takes to get to the care decision.

DR. LAMPONE: Yeah, the only thing I would add is that I think you see many times those patients who have had complex issues and they're in the ICU still getting some treatment, but basically being monitored. And there's a subjective decision made that that level of monitoring could not occur in another portion of the hospital in another setting. So, I think that's where it gets a little grey.

DR. DUDLEY: Right. And I've heard preference for some of the proprietary methods of assessing a patient's level of care, but I think there's no way in the end that there isn't some subjectiveness left in these decisions.

But I think that from a management

standpoint we have to pick a decision and -- or a point at which we implement the idea that the length of stay is ended.

And the orders are a clearer thing that someone can abstract and be confident about. And much of the other stuff that's uncertain and varies and is subjective isn't stuff that we would necessarily want to take account for and we might even want to hold the hospital responsible for that.

CO-CHAIR BRATZLER: James.

DR. O'BRIEN: So, just a -- it may be -- I'm not sure if it's an error related to what we have as far as the data dictionary, but what's included on the NQF website says that the date of discharge from the ICU is that latest documented data of the patient being physically in the unit, not when there's an order for discharge or transfer.

DR. DUDLEY: I will look into the data dictionary, but it -- they have to physically be in the unit and -- I'm trying to think of how

that plays out differently. I'm not -- I'm not sure about that.

CO-CHAIR BRATZLER: Bruno.

DR. DIGIOVINE: We just finished reviewing a similar measure for pediatric ICUs where they felt it was important to have readmission data as a balancing measure when looking at length of stay.

Do you think that is a threat to validity in your -- in looking at it as an adult ICU without knowing whether there may be premature discharges that are leading to readmissions?

DR. DUDLEY: Yeah, I thought that was interesting. I think there's a big difference in the importance of that in the pediatric arena where mortality often is lower and -- but even so, I wondered about that. I mean, what we care about in the end is the clinical outcome, most important clinical outcome, and then efficiency in getting to that clinical outcome.

So, I think that at least from the

perspective of participants in our group, it was not thought that going to the effort of collecting readmission data was worth it and that it was most important to have the clinical outcome and then some efficiency measure.

CO-CHAIR BRATZLER: Other comments?

(No comments.)

CO-CHAIR BRATZLER: So, hearing none, we'll go ahead and vote on validity.

MS. AMIRAULT: Validity for 0702. One, being high; two, moderate; three, low and four, insufficient.

(Voting.)

MS. BAL: Ella, can you please resend your vote?

(Voting.)

MS. AMIRAULT: Zero high, 12 moderate, 10 low and zero insufficient. And based on the percentage, it's in the grey zone.

CO-CHAIR BRATZLER: All right. Feasibility.

DR. DORMAN: So, feasibility, I think

is -- gets us back into the same discussion to some extent we've had. It's chart abstraction and the primary concern really exists around that concept. It was discussed -- or concerns were raised around the statement that chart abstractors took between 11 and 15 minutes. And we've just spent a lot of time talking about the complexity of figuring out whether the patient was critically ill and required a critical care service.

And I think the developer said earlier that the people are trained to look for the intent. Seems like it would take longer than 10 to 15 minutes to do that. So, there was significant concern about that aspect.

CO-CHAIR BRATZLER: Any other --

DR. DUDLEY: I may have been unclear. So, obviously looking for an intent is impossible to do from a chart. I -- so, the -- I'm not certain right now because I haven't gone through an auditing process in a while, about the use of orders versus use of the physically in the ICU,

but both of those are relatively easy to implement.

CO-CHAIR BRATZLER: Yes, Chana.

MS. WEST: I was trying to be quiet here, but in the -- in my past life I actually was responsible for the people that were collecting the data on this measure and the previous measure. And it did take a significant amount of time for them to pull the data from the records. And it was an electronic measure -- I'm sorry -- an electronic record. So, it wasn't paper where they're having to flip every single page in the chart. It was electronic, which is easier to navigate and it did take a significant amount of time to extract.

CO-CHAIR BRATZLER: Other comments? Feasibility.

(No comments.)

CO-CHAIR BRATZLER: Okay. Hearing none, Janine.

MS. AMIRAULT: Feasibility for 0702. One, high; two, moderate; three, low and four,

insufficient.

(Voting.)

MS. AMIRAULT: Zero high, 11 moderate, 11 low and zero insufficient. Based on the percentage, this is in the grey zone.

CO-CHAIR BRATZLER: And use and usability.

CO-CHAIR LANG: Usability. Until 2013, the measure was used for internal QI in California. In 2013, the developer transformed -- began transforming the measure into an eMeasure for consideration by CMS. Currently, a model using data from two hospital EMRs is in progress.

Among the potential unintended consequences from this measure, one was mentioned previously, a premature discharge from ICUs. Another potentially could be that hospitals may seek to avoid high-risk patients who due to their severity of illness may require longer ICU stays.

CO-CHAIR BRATZLER: Do you have anything else, Todd, or anything?

(No response.)

CO-CHAIR BRATZLER: Any other questions/comments about use or usability?

(No comments.)

CO-CHAIR BRATZLER: Okay, Janine.

MS. AMIRAULT: Usability and use for 0702. One, high; two, moderate; three, low and four, insufficient.

(Voting.)

MS. AMIRAULT: Zero high, 10 moderate, 12 low and zero insufficient. So, this is also a grey zone.

CO-CHAIR BRATZLER: And then our last question about overall suitability for endorsement.

Any other comments? I think we've chatted with the developer quite a bit about suggestions, concerns particularly about level of care versus location of care, particularly about whether there needs to be a balancing measure on readmission or not. So, I think we've covered all those points.

Any other comments, and then we'll go

ahead and vote on overall suitability?

(No comments.)

MS. AMIRAULT: Overall suitability for measure 0702. One yes, two no.

(Voting.)

MS. AMIRAULT: Six yes, and 16 no. So, this fails.

DR. NISHIMI: So, measure 0702 is not recommended for endorsement.

CO-CHAIR LANG: We will be proceeding with Measure 0468, Hospital 30-day, all-cause, risk-standardized mortality rate following pneumonia hospitalization.

Please.

DR. OHTAKE: I just have a quick question. Since 0702 is meant to be paired with 0703 that failed, I'm just curious how that's handled from a procedural, like --

DR. DUDLEY: No, 070 -- the mortality measure is fine on its own. The length of stay measure is the one we would not recommend being used by itself.

CO-CHAIR LANG: So, we will readdress the mortality measure on our phone call since it's in the grey zone.

So, we're proceeding with Measure 0468. Are there developer representatives on the phone who wish to comment or describe?

Wow. All day we've had people on the phone and you're here. Thank you for being here. Wow. You're right here.

Please, would you like to describe or discuss the measure for two to three minutes to introduce it to us?

DR. BERNHEIM: I'm going to let Karen do that. This is Susannah Bernheim. I'm the project director.

DR. DORSEY: I'm Karen Dorsey and so I'll start by saying that this measure we are sending back with some changes since the last endorsement.

Specifically, we've expanded this measure's cohort to include patients with sepsis who have a diagnosis of pneumonia that's present

on admission and to include patients with aspiration pneumonia. Those two sets of patients were not included last time we came before the committee.

This measure is an outcomes measure measuring 30-day mortality. It has been in the IQR program for several years. We present new information about measure reliability, new testing for the risk adjustment model because we've expanded the cohort. So, there's quite a bit of new testing in this endorsement maintenance application.

The -- I think that we're sort of prepared with questions that came up from the working group to talk a little bit about rationale for expansion of the measure cohort. And so, we're eager to get into that discussion with you all.

CO-CHAIR LANG: Great. Thank you very much. I want to mention that we have two individuals on our committee who are conflicted and will not be participating. That's Mitch

Harris and Dale Bratzler.

And the measure will be reviewed for us by Chana and Rich. Take it away. Benson, please speak closer to the mic.

DR. MURRAY: Not the work, but that should be a winning strategy. Okay. So, this is hospital 30-day, all-cause, risk-standardized mortality rate following pneumonia hospitalization.

The steward is CMS. And so, mortality is defined as death for any cause within 30 days after the date of admission for the index admission discharged from the hospital with a principal discharge diagnosis of pneumonia.

And as was said, some of the subcategories here we can get to in more detail probably during validity. And CMS annually reports this measure for patients who are 65 and older although this, as I understand it, this can be used to present data for over 18 or older than 65.

The level of analysis is the hospital

or acute care setting. And the data source for the measure is administrative claims.

There is quite a bit of information that has been reviewed for previous documentation of the evidence and I don't think a huge amount has been added since. Although, there have been a couple of papers that are more recent that have been cited and they are said to support some of the changes in the numerator.

So, basically the rationale for this is that the healthcare -- this healthcare outcome developer states hospitals are able to influence mortality rates through a broad range of clinical activities, including preventing complications, provision evidence-based care, discharge planning, management of care, transitions, medication reconciliation and patient education.

So, this is eligible for a pass rating. So, I guess we're asked at this point whether you want to actually review the incremental evidence.

I think the question is, is how

important is the incremental evidence to the decision to change the numerator?

I know it helps justify it, but the change in the numerator was actually driven by, I think, expert input.

DR. DORSEY: And so, there was some evidence that we allude to in the materials of changes in coding practices or increased use of sepsis coding for patients who also have pneumonia when they present to the hospital.

And that was -- well, I describe that as sort of the starting point, what brought our interest to looking into this in the fee-for-service population specifically related to this measure.

So, then we conducted a great deal of analysis to look at whether or not this expansion was good and whether the validity of the measure would be enhanced. And so, it was that kind of coupling of the evidence and our own investigation.

CO-CHAIR LANG: So, unless --

DR. LAMPONE: I just had a clarification on that. When you included sepsis you said "not severe." Did that include hard clinical evidence, or was that based upon presentation evidence of hypotension, tachycardia, etcetera, etcetera?

Do they have to have a bacterial or culture positive data?

DR. DORSEY: So, this is all claims-based. So, when we refer to severe sepsis, we're referring to having received a principal discharge diagnosis of severe sepsis according to the current kind of coding guidelines.

DR. MURRAY: So, you don't know what's behind that in any individual case, right?

DR. DORSEY: Only what the coding guidelines instruct.

DR. MURRAY: And since we're talking about sepsis, I understand there's been a recent change sepsis-3 to the definition.

Is that something that we should be discussing at some point? Are there any threats

to validity or now or --

DR. DORSEY: At some point I think for certain that I think is not two weeks old yet --

DR. MURRAY: Right.

DR. DORSEY: -- that change to guidance. So, it wouldn't change our sort of conclusions about the appropriateness of this cohort expansion right now, but it's certainly something that we would have to revisit and come back to the committee as hospitals begin to uptake the new guidelines and if that requires changes to what we're proposing today.

DR. YEALY: Maybe that's something I could comment on. I just wrote an editorial on this.

I'm not certain how broad the uptake will be on the new sepsis-3 definitions. And essentially all that happened in them was getting rid of the term "severe sepsis."

DR. MURRAY: Right.

DR. YEALY: You're either septic or septic shock. And I think there will be some

coding changes, but I don't think fundamentally it will change it.

May I ask one question? How do you restratify here? It says that they're risk adjusted.

DR. DORSEY: We risk adjust for a series of patient co-morbidities -- I'm sorry.

DR. YEALY: Okay. So, how exactly do you do it?

DR. DORSEY: Well, we select variables using the hierarchical condition categories which group similar ICD-9 codes according to their, sort of, likeness and condition.

So, we use the -- for this version of the measure, the ICD-9 compatible map, which is Version 12. And that's what makes up the candidate risk variables.

And then we select which of those variables are sort of the best predictors in our models.

We use a hierarchical model that adjusts for clustering of similar patients in

individual facilities.

DR. YEALY: So, I guess my concern is two things would -- all right.

MR. SPEAKER: Get back to threats to validity, I guess.

CO-CHAIR LANG: Focusing on evidence. So, unless there is sentiment among the group that we should vote, we'll pass and move to performance gap.

MS. WEST: So, the developer ran four years of data in order to calculate and see if there were any discernible differences in performance. And when they ran the data, it seemed that there were some discernible gaps.

It seemed that African-Americans were disproportionate in terms of -- in terms of performance and -- but I think one of the things that we talked about and we only discussed this very briefly, this entire measure at the end of the workgroup call due to timing constraints, but the question was whether or not the measure was actually having any impact, because the mortality

rates were actually increasing.

It went from with the initial -- or, I'm sorry -- with the last evaluation in 2012 it went from 11.7. And then with the most recent data run it went to 16.4 percent. So, that was one of the discussion points.

CO-CHAIR LANG: Please go ahead.

DR. DORSEY: So, the increase that you're speaking of is directly and wholly related to the expansion of the cohort.

So, the patients who have a principal discharge diagnosis of sepsis or aspiration pneumonia carry a higher mortality risk.

And so, when we brought them into the cohort, it increased the average mortality rate for the entire cohort, but we also present the trend in mortality rates over three years.

And with the expanded cohort, you do see a decrease in the mortality rates in the national sample over the three-year period.

So, it decreased, but it -- the whole thing increased because we included sicker

patients in the cohort.

CO-CHAIR LANG: Additional discussion or questions for the developers? Yes, Bruno.

DR. DIGIOVINE: Yeah, so I'm trying to understand the disparity issue. And this will come up for the next one as well.

So, I'm reading this as every group you point out has the same median rate of mortality. So, that would strike me as a lack of disparity.

And so, on what data are we using to say there's a gap in performance? I don't know what we would say would be the appropriate mortality for patients admitted with pneumonia for over 30 days.

DR. BERNHEIM: So, this has been something that's actually always confused me about this NQF form as well. So, I'll just admit that.

So, there's two things about performance gap. One is just as this measure is used, do we think there's evidence that there's

still room to improve, that it's useful to report this because we think there's evidence that hospitals can improve and we're seeing that in two ways.

One, there's still a substantial distribution among all hospitals in mortality rates. That is a hint that there's some room to improve. And we're seeing some decreases over the last three years and we think there's probably still room to move.

The disparities piece is kind of stuck in that same section of the application, but you're right in this case. We're not seeing huge disparities among hospitals based on their mix of patients.

There's still a performance gap nationally on this measure, but we don't see a huge disparities gap.

Does that help?

DR. DIGIOVINE: So, just that last -- you said there's a performance gap on this measure nationally based on what?

DR. BERNHEIM: So, based on the two things I noted before that what -- how I interpret the question of performance gap is, do we think that there's evidence that there is room for our nation's hospitals to do better?

And so, the fact that they have been improving in recent years and that there's a wide distribution across all hospitals suggests that there is a gap in performance nationally, but it's sort of distinct from the disparities question even though those are lumped in the form, which I think often confuses committees and, quite honestly, developers.

DR. NISHIMI: Just as a point of clarification why they're together. We have seen measures for which if you were just to look at the performance overall, one might conclude that there's no longer a gap and no room for improvement. But when you drill down and look at disparities by race and ethnicity, there's a clear breakdown.

So, in other words, there is room for

improvement. That's actually why it's very important to look at those together.

DR. DORSEY: Thank you.

DR. YEALY: So, I'm -- again, it's hard to know about a gap or disparities based on care if we can't be certain that the illness burden at presentation is the same, not illness burden that eventually develops. That's a whole separate -- that's actually what you're trying to measure is the care quality.

So, absent using a pneumonia-specific tool, which is virtually impossible to do from an administrative dataset, how do we know that these all -- that different collections of patient have the same illness burden to start off with? I don't -- that's why I was asking other questions about how do you risk adjust.

What I'm hearing I don't think would fully embrace all these and disparities could just be different groups of people with different initial illness burdens.

DR. BERNHEIM: All right. So, I think

you're asking -- and I'll just build off of what Karen said, you know, how do we handle the fact that either individual patients, subgroups of patients or hospitals are going to have different case mix of patients when they come in.

And so, you know, we have a model that takes for each patient that links back to the prior 12 months and collects all diagnoses from their both inpatient and outpatient settings.

We have a fair amount of information about patients, it's claims-based information, but we have a fairly good sense for patients about how sick they are and that is -- that risk of adjustment is built in.

And in the original measure, we were able to then validate it against the model using clinical data for severity and show that the hospital profiling was the same if you use the claims data for the risk adjustment or the clinical data.

Does that answer your question? I'm not sure that I understood the question. So, I

just want to make sure I answered it.

DR. YEALY: It still doesn't sound pneumonia-specific to me. I mean, so some of the things you'd never be able to know.

For example, the respiratory rate is a big driver of initial illness burden. You'd never pick that up in a million years.

Minor alteration in sensorium you would never -- is the single biggest driver. Beyond that, you'd never pick that up.

DR. BERNHEIM: Right. So, very important point. So, what we don't have is exactly what you said, is we do not in this measure have, you know, minute clinical data on -- we don't have respiratory rate, we don't have oxygen saturation. And that's why the original chart validation was so critical.

So, in that case, what we did was we took the exact same set of patients from the same hospitals and we ran a model that used the claims data for risk adjustment, and we ran a model that did have that clinical data, and we asked the

question whether or not it profiled hospitals similarly.

So, even if it doesn't get each exact patient identical when you're at the aggregated level of trying to understand the risk burden of the patients even though it's a little bit counterintuitive, we've done this with many of our measures, we find that at the aggregate level of the hospital the claims data does equate the job at the risk adjustment. And so, we get very, very similar outputs in the model as if we had clinical data.

Not that we aren't eager to move towards clinical data, but that's what we can do with these measures.

DR. LAMPONE: I'm sorry, can I just -- I was just going to ask and I saw it in some of the data provided in one document I couldn't open, what are your age bands for risk adjustment?

And the reason why I ask, I think, you know, and this is a CMS measure, as you get

higher in age with co-morbidities, chronic renal failure, even Stage 2 or 3 is not the same as chronic renal failure, you know, in a younger patient, because that's often coupled with other co-morbidities such as congestive heart failure, diabetes and other things that have been longstanding.

So, how does your model adjust for that? Because my experience with that age band of over 65, they tend to bounce back to the hospital much more frequently mainly in many instances not because of the initial diagnosis, but because the initial diagnosis exacerbated an existing co-morbidity that wasn't really evident at the time of discharge.

DR. DORSEY: Yeah, we do include age as a continuous variable in our risk model.

DR. LAMPONE: And so, what is that? What are those age bands? That's what I was interested in.

DR. DORSEY: I don't -- I don't think I understand your question, what are the age

bands?

DR. LAMPONE: What were the age cutoffs? What are the different individual age --

DR. DORSEY: So, we use it as a continuous variable for --

DR. LAMPONE: Okay.

DR. DORSEY: -- all fee-for-service patients over 65.

DR. LAMPONE: Okay.

DR. DORSEY: That's who's in the measure.

DR. LAMPONE: Does your risk model take into account age into those -- into the risk model?

DR. DORSEY: Yes.

DR. LAMPONE: Okay.

DR. DIGIOVINE: So, just because I just want to make sure I understand this, your data is again showing us no difference in mortality rates over the period of time that you're reporting. I thought I heard you say

that's because the definition changed.

I thought the definition change was from here forward -- if the definition changed in the past, when did it change? And can you show -- do you have data using the same definition over a period of time that is in our packet somewhere?

DR. DORSEY: So, the way that we report the measures is that we use claims from previous years. Right? So, we use claims for three years previous to whatever year it's reported in.

So, when we did the retesting of this measure, we're actually using data from 2012 through 2015. So, that's why we're talking about what's happened over the past few years.

So, we can change the definition of the cohort, look at those years in the past, and we can see the trends in the mortality rate in 2012, 2013, 2014.

And so, that's what we mean by a decline over those past years even in the -- with the expanded cohort definition.

Does that answer your question? So, overall the mortality went up over that same period of time from 2012 to 2015, because we expanded the cohort and brought sicker patients or patients with a higher risk of mortality into the cohort over the -- compared to our old definition that did not include sepsis or aspiration pneumonia.

DR. DIGIOVINE: But it is true that between 2011 and 2015 there's been no improvement in --

DR. DORSEY: No, there has been, or improvement in the outcome --

DR. DIGIOVINE: But the data you have in front of us have a mean of 17, 17, 16 and 16.

DR. DORSEY: So, that's the -- that's actually the between group of comparison. When you look at the table right above that in the summary document here, you can see the rates over each year of measurement.

DR. DIGIOVINE: Right. And they go 16.8, 16.7, 15.5, 16.4.

DR. DORSEY: Right.

DR. DIGIOVINE: You're interpreting that as an improvement?

DR. DORSEY: Yes.

DR. DIGIOVINE: Okay.

DR. BERNHEIM: The 16.8, 16.7, 15.5, the 16.4 is mislabeled. That's the three-year combined.

DR. DIGIOVINE: I see.

DR. BERNHEIM: So, just look at those first three columns. That's the trend, that's -- those first three columns.

DR. DIGIOVINE: I see.

DR. BERNHEIM: The fourth column is the three-year --

DR. DIGIOVINE: Okay. Thank you.

DR. BERNHEIM: -- combined.

CO-CHAIR LANG: Is there additional -- oh, go ahead. I'm sorry.

DR. GLOMB: Just for clarification, that's a statistically significant improvement?

DR. DORSEY: We actually don't test

the statistical significance of it.

DR. GLOMB: Okay.

CO-CHAIR LANG: Are there additional comments or questions for the developers pertaining to performance gap?

(No response.)

CO-CHAIR LANG: As I see no further comments/questions indicated, we'll proceed to vote on performance gap.

Janine.

MS. AMIRAULT: Performance gap for 0468. One, being high; two, moderate; three, low and four, insufficient.

(Voting.)

MS. AMIRAULT: One high, 11 moderate, eight low and one insufficient. Based on the percentage, it's grey zone.

CO-CHAIR LANG: Reliability.

DR. MURRAY: Do you want to continue, or do you want me to go? I can go. So, reliability has been tested. And the testing shows that the measure data elements are

repeatable producing the same results a high proportion of the time when assessed in the same population and same time period.

The developer has conducted reliability testing at the measure score level. And as mentioned, there have been updates to testing.

And the data being used include a more recent cohort from 2011 to 2014 Medicare fee-for-service in 4600 hospitals or so.

They used a split sample methodology and reported the statistic of the intraclass correlation coefficient which had a numerical value of 0.79 which we're told is a strong number, but perhaps you can speak to that. And given the algorithm, it is potentially eligible for a high rating.

So, could we ask the developers to say a little bit more about the statistical method and reliability testing?

DR. DORSEY: Sure. We actually have one of our analysts on the phone and maybe she

can weigh in if the operator could open up her line, but I'll say that we take the full three-year measurement period and we randomly split it into two equal -- two equal parts, two equal samples, and then we look at the characterization of hospitals in each sample and compare them.

And we calculate the intraclass correlation coefficient in the hospital performance.

DR. MURRAY: And this is on the raw data, or is this on the risk --

DR. DORSEY: No, this is adjusted.

DR. MURRAY: This is adjusted. Okay. Thank you.

Any other questions on this?

DR. NISHIMI: Operator, can you open Jackie's line?

THE OPERATOR: I did not see that Jackie is on the line.

CO-CHAIR LANG: Okay.

DR. NISHIMI: Only if the committee has more questions about it.

DR. MURRAY: So, the question is, do the results demonstrate sufficient reliability so the differences in performance can be identified?

CO-CHAIR LANG: Are there additional comments/questions for our developers?

(No response.)

CO-CHAIR LANG: Going once. Going twice. We will proceed to you, Janine.

MS. AMIRAULT: Reliability for 0468. One, being high; two, moderate; three, low and four insufficient.

(Voting.)

MS. AMIRAULT: Five high, 13 moderate, three low and zero insufficient. Based on the percentage, we'll move along.

CO-CHAIR LANG: Proceeding to validity.

DR. MURRAY: Do you want to take that one? No? Go ahead.

MS. WEST: Okay. I'll take this, but I will say that I'm not a methodologist or a statistician. So --

MR. SPEAKER: That's okay. That makes two of us.

MS. WEST: -- I'll definitely need some other people to jump in here. So, with the specifications as I noted earlier, several changes were made.

They expanded the definition in terms of the cohort, added "aspiration pneumonia," as well as "severe sepsis."

They document inside of -- inside of the measure information form where the literature supports the increase in -- or the broadening of the definition in adding severe sepsis in there.

And I know that we discussed on the workgroup that there was some earlier reasons for why they added aspiration pneumonia in there as well.

In terms of the actual testing, the -- they conducted additional testing of the risk adjustment model and the broadened definition -- I'm sorry -- they -- excuse me -- they conducted additional tests of the risk adjustment model,

but did not conduct additional testing on the respecified measure itself.

In terms of threats to validity, they examined frequencies and proportions of the total cohort use for each exclusion criterion and provided percentiles for each of those used there.

I'm not sure if anybody else wants to jump in and be able to explain some of the numbers that we're actually seeing in there, but that was on page 7 of the actual document if we wanted to post it for everyone to take a look at it.

DR. MURRAY: So, I guess a couple questions for the developers. The validity testing results, are these -- they're not a testing of the validity, they're a testing of the risk stratification scheme, is that -- are they a testing of just the risk stratification, really?

DR. DORSEY: So, we describe what we do for measure validity. We describe the risk model validation that was done as part of

original measure development and that we've replicated from many of our claims-based measures around other conditions.

And we talk about sort of our vigorous adherence to the methodology around the development of the measure itself. So, yeah, it's mostly focused around the validity of the risk model.

DR. MURRAY: So, I guess the question stands as to whether the validity construct is different now because the numerator has changed so much and you've seen such a change actually in the actual numerical value that comes from that computation.

DR. BERNHEIM: And so, I think the key thing there is that we felt like the expansion of the cohort was largely a response to a threat to validity.

So, the literature was indicating that as hospitals were increasingly coding pneumonia patients as septic and they were doing this at very different rates at different hospitals, we

were losing many pneumonia patients in the measure and we were losing different proportions of them across hospitals.

And so, there was a threat to the validity of the measure without these patients. And so --

DR. MURRAY: Right.

DR. BERNHEIM: -- the other thing we think is that it makes the measure a lot more valid to bring these sepsis patients in who based on the coding guidance before two weeks ago, really anyone who was sick enough to be hospitalized with pneumonia would meet criteria for sepsis.

So, it was sort of just a choice for hospitals whether to call these exact same patients sepsis or not. So, we really needed to bring them into the measure to keep the measure valid.

The one other test we did, which isn't sort of typical measure of validity, but I think helped us feel very sure we were moving in the

right direction, is that we did an examination of the relationship between a hospital's proportion of pneumonia patients that were coded as sepsis and how they did on this measure.

And what we saw in the old measure was that hospitals that had a very high proportion of their seeming pneumonia patients coming in with a diagnosis of sepsis and therefore excluded from our prior measure, tended to have pretty low mortality rates because we were losing their sickest patients.

And there was a relationship between your tendency to code and how you performed on the measure, which isn't clinically sensible. And when we changed our approach, we no longer saw this relationship.

Now, there's not a strong relationship between what proportion of your patients are septic or not and your performance.

Now, there's still a lot of variety in the proportion, but we think we're getting a more coherent and consistent population across

hospitals.

So, that's not a classic validity test, but for us it was a sign of validity of the change we had made to the measure.

DR. MURRAY: Thank you.

CO-CHAIR LANG: Bruno, go ahead.

DR. DIGIOVINE: Yeah, certainly a key part of this measure is having pneumonia identified as present on admission. I don't know -- I know very little about coding, but how reliable or how valid is coding to actually identify those things that are present on admission, and those things that truly do develop later in the hospitalization?

DR. DORSEY: We have done some analysis of present-on-admission coding and said that there's, you know, been a good uptake since it was mandated to be used for hospital-inquired conditions. And those hospitals used these and they applied them according to the guidelines and they are applied correctly.

DR. LAMPONE: I just have one closing

comment. On the risk adjustment, I saw in the documentation where you do have a number of diagnoses that are used in the age range over 65, which I think is good.

The question I have about this as it relates to validity is when you have patients that score high risk that come in with pneumonia and they die, is there any adjustment in the reporting where you would have patients that would have a high expectation of nonsurvival of that event within 30 days?

Because I think you're going to have hospitals depending upon their geographic area, they may have -- or just the demographic area of that hospital where they may have higher numbers of these folks coming in.

DR. DORSEY: So, the purpose of the risk adjustment that we apply in the measure is to account for differences in-house, patients are at some hospitals versus others. Their burden of disease when they walk in the door. When they first get/seek treatment at the hospital. So,

that's the purpose of the risk model.

And what the measure does is it basically takes into account how sick the patients are at each individual hospital and it assesses what we would predict the mortality rate would be given how sick the patients at that hospital are. And then it compares it to a calculated national average of how -- an average hospital would do with that patient -- with that hospital's case mix.

DR. LAMPONE: So, you can predict the mortality --

DR. DORSEY: Right. We predict it based on -- and then compare it to the nation.

MS. BAL: Anyone listening on the phone, please mute your computer. We're getting a little bit of feedback of an echo. Thank you.

CO-CHAIR LANG: Additional comments/questions concerning validity.

Bruno.

DR. DIGIOVINE: One more.

CO-CHAIR LANG: Go for it.

DR. DIGIOVINE: Mortality. So, in the Medicare data patient population that I think is easier for you to gather, how confident are you that you have a valid way of assessing mortality in patients under the age of 65?

DR. DORSEY: So, to address that I'll basically explain how we developed our all-payer model for the measure. And we actually did retest it in the expanded cohort, but we use state data from California which is one of the states that has an all-payer database that combines mortality information with inpatient claims information.

So -- and outpatient claim -- just inpatient claims information. So, we're able to look at -- we're able to look in that setting in California state, 18 and over, we're able to look at -- build an all-payer measure and test it.

Not all 50 states obviously collect mortality data or link it with claims in a way that we would be able to use it and we have not explored the implementation of something that

could work across 50 states. And recognize the concerns about the current environment with respect to HIPAA and moving away from unique patient identifiers and restricting the kinds of entities that can collect identifiers and linkable data.

So, it's a, you know, it's a valid concern about implementation, but currently the measure is only reported in Medicare fee-for-service over 65 for exactly that reason.

CO-CHAIR LANG: Additional discussion? If not, Janine.

MS. AMIRAULT: Validity for 0468. One, high; two, moderate; three, low and four, insufficient.

(Voting.)

MS. BAL: Ella, please send in your vote.

(Voting.)

MS. AMIRAULT: Two high, 14 moderate, four low and one insufficient. Based on the percentage, we can move along.

CO-CHAIR LANG: Feasibility.

DR. MURRAY: So, all data elements are in fine fields in electronic claims and generated or collected by and used by healthcare personnel during the provision of care.

The data are coded by someone other than the person obtaining the original information. So, the feasibility should be pretty good.

Any questions or discussion on that?

MS. WEST: I think that the only thing that we had discussion about, and it's similar to what -- when I just mentioned the concern about capturing death --

DR. MURRAY: Right.

CO-CHAIR LANG: Any further discussion? Do you want to say something?

DR. BERNHEIM: Just to respond to that, I think, you know, the issue would be if an entity wanted to use this measure in a broader population. In its current use, again, it's used just for the over 65 in Medicare. So, the death

capture is not an issue.

We purposely specified it to have potentially use in other settings. So, any given state that does have the ability or a health plan that's got the right linked data obviously could use it in a broader population we've shown that the model works, but it shouldn't be used if there's not a reliable source of death information. And so, I think that's just a limitation to the use of the all-payer measure.

CO-CHAIR LANG: Janine.

MS. AMIRAULT: Feasibility for 0468. One, high; two, moderate; three, low or four, insufficient.

(Voting.)

DR. NISHIMI: Crystal, can you submit your vote?

(Voting.)

MS. AMIRAULT: Eight high, 13 moderate, zero low and zero insufficient. Based on the percentage, we will move on.

CO-CHAIR LANG: Usability and use.

MS. WEST: Okay. As was already mentioned, it's being used, the measure right now, in the IQR program, the CMS IQR program, as well as value-based purchasing.

CO-CHAIR LANG: Discussion.

(No response.)

CO-CHAIR LANG: All right. Seeing no hands, we'll proceed to you, Janine.

MS. AMIRAULT: Usability and use for 0468. One, high; two, moderate; three, low and four, insufficient.

(Voting.)

MS. AMIRAULT: Nine high, nine moderate, three low and zero insufficient. Based on the percentage, we will move on.

CO-CHAIR LANG: Now, we will consider the overall suitability of the measure. Discussion, additional questions for the developers who are both here and I think they're on the phone -- doesn't matter. They're here.

Any additional questions or comments?

(No response.)

CO-CHAIR LANG: Seeing none, we will proceed to vote on the overall suitability for endorsement of the measure.

MS. AMIRAULT: Overall suitability for 0468. One, yes; two, no.

(Voting.)

MS. BAL: Ella, please send in your vote -- oh, never mind. We received it.

(Voting.)

MS. AMIRAULT: 17 yes, four no. So, based on the percentage, this will be recommended.

DR. MURRAY: David, where's the related pneumonia mortality rate that comes up? When do you want to talk about that? Should we talk about it now?

DR. NISHIMI: No, we're going to discuss --

DR. MURRAY: Do that later?

DR. NISHIMI: -- the last measure. I'd like everyone to move crisply, and then we'll be able to finish on time and we can have related

offline --

CO-CHAIR LANG: So, we're moving forward to coup de grace for our two-day experience, which is 1893, hospital 30-day, all-cause, risk-standardized mortality rate following chronic obstructive pulmonary disease hospitalization.

MS. GORHAM: Before we move on, can I ask how many of you have to leave before 2:30?

(Off microphone comment.)

MS. GORHAM: Okay. So, we'll still have quorum. So, we can do this last one. Thank you, gentlemen.

CO-CHAIR LANG: Thank you. You will be conspicuous in your absence. Thank you.

Would the developers wish to make some brief comments regarding the measure?

DR. DORSEY: I'll be very brief. This measure assesses mortality within 30 days of admission to the hospital for chronic obstructive pulmonary disease for Medicare fee-for-service for those 65 and older.

It has not changed since it was previously endorsed. Although, we do present some updated data.

CO-CHAIR LANG: Bruno, evidence.

DR. DIGIOVINE: We didn't think there was any need to vote on the evidence.

CO-CHAIR LANG: Any further discussion, comments? If not, we will pass and then proceed to performance gap.

Bruno.

DR. DIGIOVINE: Yes. So, I think performance gap is going to be the same. So, I'll just highlight what I'm seeing.

So, at this time I don' think this data is summarized. So, if it does look like it's going 7.7, 8.1, 7.4, 7.8 over four consecutive years and there doesn't seem to be a difference across the different performance groups, so I think there is still a concern that we had about showing that there's a gap in care.

DR. DORSEY: That last column is a three-year average. Sorry.

MS. BAL: Yeah, sorry about that. That was a typo on staff's part.

DR. DIGIOVINE: But the -- okay. So, again, I guess 7.6 -- 7.7 to 6.7 you're interpreting as an improvement in care.

DR. DORSEY: It's small.

(Laughter.)

DR. DORSEY: But I'll also just, you know, direct your attention to the range that we still see, you know. It's a little narrow, even narrower than the last measure we talked about, but there still is a range of performance.

DR. BERNHEIM: Mark, our clinical expert, is on the phone. Mark, do you want to add any points to that sort of evidence of a performance gap for mortality for patients with COPD?

DR. METERSKY: Yeah. There have been studies, although some of them are a little dated, showing that some of the established processes of care, that there are gaps in the performance of those.

We certainly see problems with transitions of care. Now, I'm speaking anecdotally, but we see that all the time and a lot of these patients will die after their discharge, but granted it has been tough to show improvement.

Probably many of you are aware of the study seeking to improve post-discharge outcomes looking at intensive contact with patients and it actually increased mortality. So, it is a difficult problem, but clinicians who work with COPD do see gaps and certainly the evidence in some process gaps is clear.

CO-CHAIR LANG: Is there additional discussion?

(No response.)

CO-CHAIR LANG: Janine.

MS. AMIRAULT: Performance gap for 1893. One being high; two, moderate; three, low or four, insufficient.

(Voting.)

MS. AMIRAULT: Two high, 14 moderate,

four low and zero insufficient. And based on the percentage, we can move along.

CO-CHAIR LANG: Validity -- reliability, yes.

DR. DIGIOVINE: That's okay. So, in terms of reliability the test was sort of an interclass correlation coefficient with the main testing we saw.

The split sample analysis was 0.51 which is lower than the sort of minimal acceptable 0.7. There was a high, however, that I didn't understand. So, I guess the developer can help with the "however" piece of that, but that was, I think, our concern in terms of reliability.

DR. BERNHEIM: The "however" is mostly that we use a very conservative approach. So, in our measures that have lower volumes, we tend to see slightly lower numbers, you know.

We're not allowing for any overlap between the patients that are tested in the test and retest. So, we're sort of doing the most

conservative testing.

And in both COPD and AMI where we have lower volumes, we don't get quite to the same level as we do in the higher volume measures.

We do have a lot of things built into our measure to prevent us from misclassifying small volume hospitals. I mean, that's one other thing I'll say is that we exclude hospitals with fewer than 25 patients from reporting.

And we, you know, use an interval estimate for classifying hospitals that also ensures that we're really confident about how we classify them in the IQR program. So, that helps with the reliability.

CO-CHAIR LANG: Are there any questions/further comments regarding reliability? I'm told that we can proceed to validity without voting on reliability.

Is there an objection to proceeding to validity?

(No response.)

CO-CHAIR LANG: With all due respect,

Janine, we're going to proceed to validity. Take it away, Bruno.

DR. DIGIOVINE: Okay. So, validity was done based on face validity. And, again, I think I was somewhat -- the -- what is the summary there says that 90 percent of the expert panel agreed that it had face validity, but conversely 60 percent were able to agree at a level of -- that was either moderate or strong, which meant that 40 percent did not agree that it was either moderately or strongly had face validity in terms of that this measure was an accurate reflection of quality.

And so, that was certainly my concern and the concern of our group.

DR. BERNHEIM: Yes. So, when we developed this measure, we had developed a number of claims-based measures where we had the advantage of having the chart data to show the validity of the risk model with the chart data.

For COPD we didn't have the same advantage. We didn't have a national dataset of

chart-abstracted data. So, we have had to rely a little on the fact that we've done a number of measures where we've successfully done a chart validation.

So, we've done that for stroke, AMI, pneumonia, heart failure, many measures where we've had chart data and showing the claims-based models work.

I think when we don't have that data, it's a little harder to get that strong validity. So, we've depended on clinical experts.

This measure has now been out and reporting for a number of years pretty successfully. So, it's gaining face validity since the time that it was developed, but the numbers are what you see.

DR. O'BRIEN: There was -- I can't remember if it was this measure of the prior one. There was discussion -- previous discussion about the hospice exclusion of the first day versus subsequent days and looking at those.

For pneumonia, obviously as an acute

illness construct I get the notion that you can enroll people and hospice changes their goals of care as a result of patients getting worse because of poor care.

With a chronic illness like COPD, that actually may be a patient-centered outcome as enrolling them in hospice yet they may die within the next 30 days.

Have you had the chance to look at, again, the enrollment in hospice not just on day one, but subsequently during the hospitalization to see how much of an impact that change in exclusion might be?

DR. DORSEY: We've not looked at it freshly. We did address this when we first brought the measure before NQF and we found that there's very, very little enrollment actually during the index stay. And most enrollment happens on discharge.

So, it's a very small, small group of people who we identify by looking at days during the index, or even before the index. Even before

the index it's less than two percent. So, we haven't renewed that analysis.

DR. O'BRIEN: It just, for me, raises the question of if it was actually included as a longer exclusion, might that actually prompt greater enrollment in hospice?

DR. DORSEY: Right. And the balance -- the counterbalance to that is that we don't want to mask signals of quality. So, we purposely don't enroll after day one, because decisions to move towards hospice could be related to problems in care or harmed patients during hospitalization. And so, that's the -- that's the balance.

DR. BERNHEIM: And, Mark, I'm going to ask if you want to weigh in at all on this issue of sort of whether it is different in a COPD chronic condition population than in the pneumonia population thinking about the exclusion in the later days of stay.

DR. METERSKY: I think it's the same concept, you know. If you don't do a good job

early and they end up on the ventilator, then it's -- you're more likely to have that conversation about hospice in a patient who if care had been good, wouldn't have ended up on the ventilator.

And I think it's a bigger issue for COPD in that pneumonia is more often, as you said, an acute issue and the expectation is that most patients will get off the ventilator as opposed to COPD where many patients may not once they end up on the ventilator. So, it is a balancing act. I agree with that.

CO-CHAIR LANG: Additional discussion?

(No response.)

CO-CHAIR LANG: So, I'm told that similar to reliability with respect to validity, although according to the algorithm this is eligible for moderate or low, not high, we can proceed to skip over pointing our blue gadgets at Janine again and move forward to feasibility.

DR. DIGIOVINE: I'm going to take this personally, David, you realize, but that's okay.

CO-CHAIR LANG: I understand. I understand.

DR. DIGIOVINE: I'll keep moving.

CO-CHAIR LANG: I understand. Unless there's an objection, we'll proceed to feasibility without voting.

Bruno.

DR. DIGIOVINE: Feasibility, electronic data, again, other than mortality which we've already, I think, gone through, don't need to go through it again, there's really nothing else about it that makes it difficult to collect.

CO-CHAIR LANG: Additional discussion/questions for the developers?

(No response.)

CO-CHAIR LANG: Janine.

MS. AMIRAULT: Feasibility for 1893. One, high; two, moderate; three, low or four, insufficient.

(Voting.)

MS. AMIRAULT: Would you mind just

giving it one more shot? Thanks.

DR. NISHIMI: We have the votes on the phone.

MS. AMIRAULT: Okay. 10 high, nine moderate, zero low and zero insufficient. Based on the percentage, it's grey zone.

(Off microphone comment.)

MS. AMIRAULT: Okay. Sorry about that. It passed.

CO-CHAIR LANG: Couldn't wait for everybody.

(Laughter.)

CO-CHAIR LANG: Usability and use.

DR. DIGIOVINE: Usability and use. So, it's publicly reported in Hospital Compare. I guess on use the question becomes one of improvement whether it's actually showing that there's been improvement. Other than that, I don't think there's any issues.

CO-CHAIR LANG: Discussion?

(No response.)

CO-CHAIR LANG: Move forward to a

vote, usability and use.

MS. AMIRAULT: Usability and use, 1893. One for high; two, moderate; three, low or four, insufficient.

(Voting.)

MS. AMIRAULT: Five high, 12 moderate, two low and zero insufficient. Based on the percentage, we can move on.

CO-CHAIR LANG: Overall suitability for endorsement of the measure. Discussion?

(No response.)

CO-CHAIR LANG: Seeing no desire for further discussion, we will proceed to Janine for a vote.

MS. AMIRAULT: Overall suitability for 1893. One for yes, and two for no.

(Voting.)

MS. AMIRAULT: 18 yes, one no. And based on the 95 percent, it can be recommended.

CO-CHAIR LANG: Thank you, Bruno. That was excellent.

All right. So, now we have

harmonization or, I mean, the first two measures, one of them failed. Go ahead. I'm sorry.

MS. BAL: Yeah. So, just out of respect to everyone's time, we're going to not discuss ruling competing. We'll bring that up at the post-draft call.

We will not have a follow-up call, but Janine will go over that in a second. But so, we're just going to bring it up in the post-draft call and give everyone a little bit extra time today.

Would that be okay, or did you guys want to discuss those?

CO-CHAIR LANG: I see no objection.

MS. AMIRAULT: So, as Poonam just mentioned due to the amount of grey zone measures and other things to continue discussions, we're going to hold two post-meeting calls. So, you can expect a doodle poll -- excuse me -- two post-draft comment calls.

And we will be sending a doodle call out to everybody just to capture the availability

and organize.

MS. BAL: Okay. Just a little clarity. There's two different types of calls. So, we have a post-meeting call that was originally scheduled for this coming Tuesday.

We will be canceling that call. We were able to get through all the measures, because you all are very amazing. So, thank you for that. So, we do not need the post-meeting call.

However, the post-draft call or the post-comment call which happened after the open commenting period, we will be scheduling an additional call for that due to the number of grey zone measures. And we want to make sure that you have enough time to really discuss those measures instead of the scheduled two-hour call.

Were there any questions about next steps? Yes.

DR. O'BRIEN: Can you just comment maybe about what the format of those calls are going to be like for the post-draft comment

calls?

MS. BAL: It will be very similar in concept to this. However, since most of them fell in the grey -- or not fell -- they all -- many of them have grey zone, we would just start at the end.

So, even if they had grey zone throughout the review, we would start at the end to see what your vote would be and you basically review overall suitability for that vote.

Obviously you can bring up -- we would ask the developer to provide more clarification based off your notes and such.

DR. NISHIMI: If we have a quorum, we will vote on the call.

MS. BAL: Any additional questions?

(No response.)

MS. BAL: All right. Thank you so much, everyone. Especially the people on the phone. We really appreciate your active participation.

MS. MUNTHALI: And, I'm sorry. I just

wanted to -- I'm hoarse today, but wanted to thank you all again adding to Poonam's gratitude and especially to your co-chairs.

This was a great meeting. I know it's very long and a lot of the issues are very difficult, but we think it was very successful. So, we'll be seeing you online. Thanks again.

DR. KAZEROONI: Thank you.

DR. POLLART: Thank you.

(Whereupon, the above-entitled matter went off the record at 2:40 p.m.)

D

U

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Pulmonary and Critical Care Standing Committee Meeting

Before: NQF

Date: 03-16-16

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