NATIONAL QUALITY FORUM

+ + + + +

PULMONARY AND CRITICAL CARE ENDORSEMENT
MAINTENANCE STEERING COMMITTEE

+ + + + +

THURSDAY
MARCH 22, 2012

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Center, 1030 15th Street, N.W., Washington, D.C., at 8:00 a.m., Stephen R. Grossbart and Kevin Weiss, Co-Chairs, presiding

PRESENT:

STEPHEN R. GROSSBART, PhD, Co-Chair
KEVIN WEISS, MD, MPH, Co-Chair
PETER ALMENOFF, MD, FCCP, Veterans Health Administration
HAYLEY BURGESS, PharmD, BCPP, Hospital Corporation of America
MICHAEL E. CANTINE, BSAST, RRT, CPFT, Morristown Medical Center
RUBIN COHEN, MD, FCCP, Hofstra University School of Medicine
NORMAN H. EDELMAN, MD, American Lung Association
WILLIAM BRENDLE GLOMB, MD, FCCP, FAAP, Texas Health and Human Services Commission
TRUDE A. HAECKER, MD, FAAP, The Children's Hospital of Philadelphia
DIANNE V. JEWELL, PT, DPT, PhD, CCS, The Rehab Intel Network
DAVID LANG, MD, Cleveland Clinic

JANET LARSON, PhD, RN, FAAN, University of Michigan School of Nursing
MITCHELL M. LEVY, MD, FCCP, FCCM, Society of Critical Care Medicine
JOHN PELLICONE, MD, FCCP, FACP, Helen Hayes Hospital
DAVID RHEW, MD, Zynx Health Incorporated
CHRISTINE STEARNS, JD, MS, New Jersey Business and Industry Association
CHARLES STEMPLE, DO, MBA, Humana
DAVID C. STOCKWELL, MD, MBA, Children's National Medical Center
CHRISTY WHETSELL, RN, MBA, ACM, West Virginia University Hospitals
DONALD M. YEALY, MD, FACEP, University of Pittsburgh

MEASURE DEVELOPERS:

DAWN ALAYON, National Committee for Quality Assurance
MARK S. ANTMAN, DDS, MBA, American Medical Association
SUSAN ARDAY, Centers for Medicare & Medicaid Services (by teleconference)
KATHERINE AST, MSW, LCSW, American Medical Association
SUSANNAH MAY BERNHEIM, MD, Yale New Haven Health Services Corporation (by teleconference)
JOHN BOTT, MSSW, MBA, Agency for Healthcare Research and Quality (by teleconference)
DALE BRATZLER, DO, MPH, Centers for Medicare & Medicaid Services (by teleconference)
LINDY CHIN, ActiveHealth (by teleconference)
KERI CHRISTENSEN, American Medical Association
DEBORAH DEITZ, RN, BSN, Centers for Medicare & Medicaid Services (by teleconference)
JEFF DREFFORD, Agency for Healthcare Research and Quality (by teleconference)
ELIZABETH DRYE, MD, SM, Centers for Medicare & Medicaid Services

CHRISTINE GALL, MS, RN, Virtual PICU Systems, LLC (by teleconference)
BRIDGET GULOTTA, MSN, MBA, American Medical Association
LAURA MAE GROSSO, PhD, MBA, Yale University School of Medicine (by teleconference)
BENJAMIN N. HAMLIN, MPH, National Committee for Quality Assurance (by teleconference)
BRUCE KRIEGER, MD, American Medical Association (by teleconference)
DENISE KRUSENOSKI, MSN, RN, CMSRN, The Joint Commission
SHELLEY S. MAGILL, MD, PhD, Centers for Disease Control and Prevention (by teleconference)
RAJESH MAKOL, ActiveHealth (by teleconference)
DAVID NAU, PhD, RPh, CPHQ, Pharmacy Quality Alliance (by teleconference)
COLLETTE PITZEN, RN, BSN, CPHQ, Minnesota Community Measurement (by teleconference)
MARJORIE RALLINS, DPM, American Medical Association
BOB REHM, MBA, National Committee for Quality Assurance
PATRICK S. ROMANO, MD, MPH, Agency for Healthcare Research and Quality
ELVIRA RYAN, RN, The Joint Commission (by teleconference)
MATTHEW SCANLON, MD, Medical College of Wisconsin (by teleconference)
AJAY SHARMA, MD, ActiveHealth (by teleconference)
BANI VIR, MD, ActiveHealth (by teleconference)
ANN E. WATT, MBA, RHIA, The Joint Commission

NQF STAFF:

HELEN BURSTIN, MD, MPH, Senior Vice President, Performance Measures

HEIDI BOSSLEY, MSN, MBA, Vice President,
Performance Measures
ANN HAMMERSMITH, General Counsel
KATHRYN STREETER
JESSICA WEBER
REVA WINKLER, MD, MPH

ALSO PRESENT:

MAUREEN DAILEY, American Nurses Association
(by teleconference)
SHEILA HEITZIG, American Academy of Allergy,
Asthma & Immunology
MELBA HINOJOSA, Health Services Advisory
Group, Inc. (by teleconference)
DARRYL ROBERTS, American Nurses Association
(by teleconference)

TABLE OF CONTENTS

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

(8:03 a.m.)

CO-CHAIR GROSSBART: Well, good morning. I hope everybody had a pleasant evening last night. I made it down to the cherry blossoms. So it was beautiful.

Before we get started, I did want to just check in and see how many of you are unable to stay until the scheduled adjournment time at three o'clock? How many of you are going to catch an early flight? Then how early, for those of you who are leaving -- So two o'clock? Two-thirty-ish? Okay.

I think our goal will be to finish early anyway. So we can do that.

Let's get started then. Do you want to say any opening welcomes or hand it over to Reva? Do we have updates that we want to share with the Committee? We sent the email out yesterday regarding the Minnesota?

DR. WINKLER: Let me just ask, does anybody have any issues or questions they

would like to ask of the staff? Otherwise, thank you all for being here again today.

CO-CHAIR GROSSBART: Then let's get started. We are going to have a little juggling around in our agenda. The first change is that Measure 0231 on the top of the last page of the agenda, top of page 4 of the agenda -- We are going to move that up to our first consideration of the candidate measures.

Dr. Patrick Romano is here, and he is going to, as a measure developer, give us the overview. What we are asking is each measure developer to provide us about a two to three-minute overview of the measure before Committee begins to review. So, Dr. Romano.

DR. ROMANO: Hello. Good morning, everyone. I am pleased to represent AHRQ this morning. I am a general internist based at UC Davis School of Medicine in Sacramento.

This measure, Pneumonia Mortality Inpatient Quality Indicator (IQI 20), is part of the inpatient quality indicators module of

the AHRQ quality indicators. It is intended for application to all payer datasets, hospital, administrative or discharge datasets such as those that are collected by 43 state health data agencies around the country.

It is intended for application to datasets that may not permit linkage of patient information across episodes of care. So it obtains the information about risk factors and outcomes from within the record of a single hospitalization.

This is a risk adjusted measure, and so it is risk adjusted using a hierarchical model that includes the patient's risk factors and hospital's in effect, and in that way it is similar to a number of the other inpatient quality indicators.

I will stop there and just take questions later.

CO-CHAIR GROSSBART: Does the Committee have any questions for the developer? All right. Then moving on -- Are

we still working on video on the screen? Okay.

MEMBER ALMENOFF: So this is an administrative risk adjusted model?

DR. ROMANO: That is correct.

MEMBER ALMENOFF: And it is an in-house death rate or a 30-day?

DR. ROMANO: It is an in-hospital mortality death rate -- in-house death rate. Correct.

MEMBER ALMENOFF: Because CMS already has a 30-day pneumonia rate. So how is that different than this?

DR. ROMANO: Right. That is correct. I think that is the next measure on the agenda. So different users have different datasets. Basically, the AHRQ quality indicators were developed in response to demand from stakeholders and users who don't have the ability to link post-discharge outcomes.

So this offers a measure of

pneumonia mortality that does not require a linkage to post-discharge outcomes.

MEMBER YEALY: If I could jump in: To have only in-hospital as opposed to 30-day or 60-day, the problem would be, as health care delivery changes, particularly the development of long term acute care facilities, you could actually have a diminishing in-hospital mortality rate with really no change in death, just because people would die in a different location.

So you really actually need both of these side by side, an in-hospital and then some other distant. Whether it was 30, 60, or 90, you could have a debate about, but if you truly wanted to measure the outcome, at a minimum both of those are needed.

MEMBER RHEW: I would completely agree with Don. I think you need both of them, but at the same time I think there is value in having just the in-hospital focus as well, so you can look specifically. Certain

measures will directly impact the hospital stay. So I think there is value there, but then the corollary is you have to complement that with the 30-day.

MEMBER ALMENOFF: That is not my point. In our system, we already do both, and I agree. You need to do both, but if one, the 30-day model, is going to be one type of model and then an in-house model is using a different administrative model, then it is not an apples to apples comparison. So I am just kind of wanting to understand, is this going to be a CMS measure for everybody?

Is this going to be one model for inpatient, one model for outpatient -- excuse me, one model for inpatient, one model for 30-day, and they are different models? I don't know how to do a comparison if they are completely different models. That is my point.

CO-CHAIR GROSSBART: I actually think we are jumping ahead to related and

competing measures.

MEMBER ALMENOFF: I know. I am sorry. A good conversation, but I do want to turn this over to Don Yealy to walk us through the measure, and we do have the documentation up on the screen now.

MEMBER YEALY: From the impact side, there was little debate about whether or not this was an important thing to be assessing. Obviously, it is a common disease with a nontrivial fatality rate that can be impacted upon by the actions of health care providers. So we had no concerns about that.

There appears to be a performance gap -- in other words, that the death rates aren't within a narrow band across sites. There also appears to have been improvement from the date that were available over an extended period of time. So it has changed, but there is still more opportunity.

The evidence behind this, there was little or no conversation about whether

there is any concerns.

Moving on to the rationale and usability and feasibility, I may as well just take them in one lump. The only question that came -- I think it, in some ways, overlaps what your concern is -- is that the administrative risk adjustment is easily done, but may not fully embrace some of the illness burden differences at onset.

Having said that, I am not sure how one would be able to do that. Obviously, one of the -- The use of one of the risk stratifying tools at time zero would be the best way to do it, but it is not easily done through an administrative dataset. So you are left with this.

What you are left with is a possibility that varying death rates or differences could be due to different illness burden rather than the actions of the providers, but it does not appear to be a systematic issue, and it doesn't appear to be

amplified in any particular band of patients. So at the end we were comfortable with this being a measure.

CO-CHAIR GROSSBART: Since we have the developer here, Dr. Romano, do you have any response to the concerns about the risk adjustment?

DR. ROMANO: It is certainly a valid concern. There is ample work in the literature regarding physiologic predictors of pneumonia mortality. We do know from some of the work from the Yale team, actually, that will follow me, that the administrative data, the comorbidity information, does surprisingly well in risk adjustment and accounts for most of the variation in apparent severity across hospitals. But having said that, there is also evidence from Michael Klein's work and others on laboratory data and physiologic parameters such as oxygen saturation that add additional value to the risk model.

Going forward, I think, into an

era of electronic health records when more states and other users are beginning to collect additional information that is available form the electronic health record, there will be opportunities to enhance the risk adjustments, and we have already begun exploratory analytic work in that area using the data, pilot data, from several states.

CO-CHAIR GROSSBART: I think, at this point with this measure we can step through the voting, and then after that we will move on to the CMS measures. So, Don, in terms of --

MEMBER YEALY: In none of these were there any concerns. They were all high or strongly positive. Our risk adjustment concern, while voiced, did not temper or alter the overall. So we could go through each one, one by one, but there was fairly strong support across every evaluative part of the process.

CO-CHAIR GROSSBART: So again,

going through our process yesterday, let's vote on the impact -- Well, first of all, other comments from the Work Group? I am sorry. Then let's open it up for the Committee. Any questions for the Work Group or Don on the impact assessment? Okay, Jessica, let's go. Let's vote, a one through four scale again. High is one, moderate is two, low is three, and four is insufficient.

It appears that the batteries are well rested. So we have 17 votes for High and one vote for Moderate.

Let's move on to the next area, which is the performance gap. Don, any comments?

MEMBER YEALY; No, unless someone has a specific question. I seem to be pithy today.

CO-CHAIR GROSSBART: Yes. Work Group, Committee, any questions on the performance gap, again a one to four scale, one being the highest. Seventeen High and two

Moderate.

Then moving on to the evidence base, this is a simple yes or no or Insufficient. Any questions or comments that you want to add, Don, or the Work Group. Any questions for the Work Group from the Committee? Then let's vote. One yes, two No. It is unanimous, 19 Yes.

Now let's move on to reliability and validity. We touched on some of these points already. Don, do you just want to give us an update?

MEMBER YEALY: No. Again, I think, while there are some concerns, any of the stratification opportunities don't appear to be systematic or isolated in a particular band and don't really threaten the measure as it is stated.

CO-CHAIR GROSSBART: So let's vote on the reliability, again a one to four scale. Fifteen votes for High and four votes for Moderate. No other votes.

Now validity? All right. Excuse me?

MEMBER ALMENOFF: I was asking what the C statistic was.

CO-CHAIR GROSSBART: So that was a question.

DR. ROMANO: It is reported in the measure submission form. Someone else may find it before I do.

MR. BOTT: Yes, this is John Bott with AHRQ; .849.

CO-CHAIR GROSSBART: Pretty good for government work. Any other questions, comments from either the Work Group or the full Committee? Then in terms of the validity vote, a one through four scale.

Seventeen votes for High and two votes for Moderate.

Now we move on to usability and feasibility. So usability?

MEMBER YEALY: Again, these seem fairly straightforward and easily described

and communicated.

CO-CHAIR GROSSBART: Reva just mentioned, as well as currently publicly reported. Any questions for -- Any comments from the Work Group, and any questions from the full Committee?

MEMBER YEALY: The only questions about this would deal with the risk stratification, really, which we have already essentially assessed on a different metric.

CO-CHAIR GROSSBART: All right. yes?

DR. ROMANO: I do want to stress, and Dr. Drye just asked this also in response to one of the earlier comments, that this is not being used by CMS for independent reporting and Hospital Compare. So it is not in direct competition with the Yale measure in that respect.

CO-CHAIR GROSSBART: Thank you for that clarification. So we are voting on usability. First of all, are there any

questions or comments from the Work Group or the full Committee? If not, this is again a one through four vote, and let's vote.

The vote is 16 with a vote of High and four with a vote of Moderate. No other votes.

Then feasibility. Any comments, Don?

MEMBER YEALY: Dead or alive is not usually a challenge to identify.

CO-CHAIR GROSSBART: You would be surprised. Any questions or comments from the Work Group or the full Committee? So let's move to our voting, a one through four scale.

How many votes do we have recorded so far? We have 18 with a rating of High and two with a rating of Moderate.

Now our final question, the overall rating and endorsement of the measure. In favor of endorsement, vote one; opposed, vote two. One more vote. Everyone voted? There we go.

It is unanimous, 20 in favor.

Next on the agenda, we are going to ask Elizabeth Drye, Dr. Drye from Yale new Haven Health System, I believe, representing CMS, to discuss the four mortality and readmission measures for pneumonia and COPD. Please take a few minutes to give an overview of the measure, and feel free to address any concerns that you have heard along the way from the Committee.

DR. DRYE: Thanks so much. I am Elizabeth Drye. I am from Yale, and I think -- I just want to confirm I have on the phone the rest of our team up in Connecticut. Are you guys there? Wonderful.

So I am going to briefly go over the four measures, the mortality and readmission measures for pneumonia and COPD, and I think, after talking to Reva, the most useful thing would be to actually talk about the two -- all four of them together, but I will start with the two mortality measures and

then the two readmission measures, because the mortality measures and the readmission measures, are structured very similarly, and they are just covering different patient group.

As you know, the pneumonia measures have been around for several years, and they are publicly reported on Hospital Compare, and the COPD measures are newly developed.

For mortality, I just wanted to briefly describe our approach to the measures. They are risk-standardized, all-cause mortality measures that look at mortality within 30 days of admission. We do include transfer patients. We basically evaluate an episode of care, which starts at admission to the hospital.

So if the patient is transferred after that in another acute care setting, we attribute to the outcome, death or not death, to the first admitting hospital. We exclude

patients who we leave against medical advice, actually, from all four of the measures. The measures are risk adjusted, as you know, using claims data.

For the pneumonia measures, we were able to validate that risk adjustment very extensively against a national dataset of chart abstracted data and, as Patrick mentioned, the performance of the model is really good.

The rates that are produced by chart based and clinic based models were highly correlated at the hospital level.

For readmission, our modeling approach is the same, but our exclusions and our time frame are a bit different. We start the 30-day clock at discharge, and it is the acute care hospital that is discharging the patient to the non-acute setting.

So if a patient is transferred between two acute care hospitals, it is the second hospital, the discharging hospital,

that would be assigned the outcome of readmission or not.

Our rational for that is that we are really looking at quality, but also at transitions of care and the management of the movement of the patient out of the acute care setting. We also exclude patients who leave against medical advice, as I mentioned before.

The readmission measure for pneumonia is publicly reported on Hospital Compare, and COPD, as I mentioned, is new.

I wanted to mention a couple of -- The main changed term to the pneumonia measure -- I just want to mention a couple of things, the changes to the pneumonia measure since it was endorsed several years ago, and then respond to a couple of issues raised in the Working Group that reviewed the measures for this Committee in February.

The main change to the pneumonia measure is that we respecified the measure for patients 18 and over. As you probably know,

we developed the measure in the Medicare fee for service population, which is a population in which we have wonderful national data, including inpatient and outpatient history on all patients in that age group.

We were able to obtain data from the state of California, as you know, a very large state, and look at how the measure worked in the population 18 and above, and we were really pleased with what we saw.

We had to basically focus our testing on two issues. One was that we didn't have data for non-admissions for either non-admitted -- or data from patients who were seen at the hospital but not admitted or data for patients in the outpatient setting, in the physician office setting.

So we had less data available for risk adjustment. We only had admissions data for risk adjustment, and then of course, we were looking at a different age group, and we really had to ask whether the risk adjustment

variables we were using were the same -- had the same relationship to the outcome of mortality or readmission in those age groups as in the older age group.

So we tested both aspects of those differences, and for all four of the models they performed really well in the 18 and over age group. In fact, the patient level discrimination was a little bit better in those age groups, and we also tested the interaction between age and the risk adjustment variables and adding interaction terms with the thought that perhaps these variables behave differently in younger patients. It didn't really change the performance of the model at all or the rates estimated by the model.

So it was convenient for us, but also, I think, hopefully, helpful for the provider and payer and user community for these measures that we could respecify these measures as 18 and over measures, and they can

be used then by states to assess COPD or pneumonia mortality or readmission.

Let me just speak to the couple of issues that came up in prior meetings. One was there was a question about whether use of a readmission measure would incentivize hospitals potentially to increase their use of observation stays in lieu of admitting patients who come back to the hospital within the 30-day time frame. That is a great question, and it is one that, actually, CMS is already aware of.

Part of our work is to follow what is happening with observation stays. I wanted to respond directly to it. We have a report that we did looking at the rate of observation stays across hospitals, and we looked from 2007 through the end of 2009.

The AMI, heart failure, and pneumonia readmission measures were just posted publicly beginning in 2009. So if we are looking for an effect of that public

reporting, it would be very hard. It would be, really, too early to see.

There is only six months of this data we had that post-dated that public reporting, but if you look up on the slide, the top line is AMI. The middle line is heart failure, the red line, and the bottom line is pneumonia, and the x axis is the year. We looked at six-month intervals across the three-years of data, 2007 through 2009, and the y axis is the mean hospital level observation rate within 30 days of discharge for these conditions.

The y axis -- I don't know if you can see the numbers, but they are very small. The highest bar is 2.5 percent, and pneumonia tops out just over 1 percent. So the typical hospital is really -- Actually, more than half the hospitals really had no use of observation stays, but the median hospital was just over one percent at the very end of our time period, so a very small use of observation

services.

When you look at patients using those services without being readmitted -- in other words, without otherwise being captured in the measure -- the numbers are even smaller. That is the next slide.

I apologize. We changed the axis on you, but the top bar is now 1.8 percent. So this is something we need to track, particularly if readmission measures are used for payment, and we hope that enough will continue to track it. We are very interested in tracking it.

I think that is their plan, but right now there are very low levels of uses of observation stays for these patients.

Does anyone have any questions about that before I make one last comment? Okay.

Another issue that came up was the potential use of an environmental factor, particulate levels -- this is in the COPD

discussion -- to risk adjust for risk of mortality or readmission.

As I mentioned, I am sympathetic to that, because the biological mechanism potentially is very clear. There are county level data that EPA collects, as mentioned, on particulate levels, but the step of incorporating that or other environmental factors into our measures is a big step.

So since the meeting in February, we took a very cursory look at the literature. There is not much yet on relationship. We did see some studies on relationship to admission, but not a lot on relationship to readmission or --- you know, these are very specific outcomes, 30-day readmission or mortality following hospitalization for COPD.

There was a study recently done by the United Kingdom that looked for the relationship between air pollution, ambient air pollution levels, and they also did some modeling, and the outcome of COPD admission.

They are focused on that because of the very, very high cost to the UK of caring for those patients, but they really didn't find anything.

The strongest association was with nitric oxide and not particulates. Not to say it isn't there, but the step of linking the actual levels to the clinical exposure and then to our outcome is quite a big step, and we, I think, will continue.

In our group, we are starting to look at environmental factors and how they may be affecting the outcomes of interest, but we are the beginning of that work, and really not able to incorporate it in this short time frame.

So I will stop there and see if you have any questions.

CO-CHAIR GROSSBART: Norm?

MEMBER STEMPLE: For the 30 days, if someone goes to a LTAC Smith rehab, does the 30 days start at discharge from that

alternative level or from the acute inpatient?

DR. DRYE: For our measure we just looked at the 30 days post-discharge from the acute care hospital setting, and we are indifferent of where you go, but I would just note that CMS is working on measures that look at readmission in post-acute care facilities.

CO-CHAIR GROSSBART: Norm, you had your hand up?

MEMBER EDELMAN: Yes. Thank you very much for addressing the pollution issue, and I was the one who raised it. I raised it primarily with regard to all-cause mortality. So you are not measuring mortality due to COPD, which I think you referred to in your discussion. You are measuring all-cause mortality.

There is very, very strong data going back 20 years to the famous Six Cities study that air pollution, largely PM2.5 small particles, explains a significant degree of unexpected mortality for cardiopulmonary

disease, not necessarily for pulmonary disease alone or cardiac disease alone, but for cardiopulmonary disease.

The level of air pollution is not under control of the hospital. So this is a potential source, in my opinion, of unintended bias which, I think, could be quite significant. I understand the difficulty of including such a metric in the standard that is going forward, but I do think it is incumbent upon the developer to do a pilot study, and that wouldn't be hard, simply to test this hypothesis; because put in air pollution levels or you can take a sample, if it is too much work to do it for the entire cohort, and see -- Those are parametric measures. Your model is designed to deal with parametric measures, and see if you significantly reduce the variants.

That can't be a difficult job, and I really think it is incumbent upon you to prove me wrong.

CO-CHAIR GROSSBART: Any other questions for the developer? Mitchell?

MEMBER LEVY: I think these are important metrics to track. You have three or four years of data. Is there any evidence that the mortality rate or readmission rate is changing?

DR. DRYE: Thanks for asking. I meant to mention. So far, there is not really a trend, except in AMI mortality, which has been dropping steadily, actually.

CO-CHAIR GROSSBART: Any other questions from the Committee? I do have one. Oh, go ahead, Trude.

MEMBER HAECKER: What about hospice care and the patients that have a predisposition? Is an exclusion in here?

DR. DRYE: Yes. That is another good question. For the mortality measures, we really -- What we would love to be able to know is: Is the patient coming into the hospital for palliative care only? In other

words, their goal is not survival, because we are trying to use mortality as a quality signal.

We have looked really extensively at the best approach to doing that, given the data that we have. We modified the measures up after they were -- I think it is for all of them -- since they were endorsed to exclude patients who had a history of enrollment in Medicare hospice up to and including the first day of admission. We still would apply -- That exclusion stands for use in the Medicare population. We don't really have a comparable indicator for the 18 and over.

We have looked at other indicators extensively, the V66.7 code, which is a concept for palliative care, and discharge to hospice. Believe it or not, these patient groups that have those different codes really do not overlap, and we are really trying to capture -- V66.7 is increasingly used, which I think is a good thing, just for pain

palliation, not only for end of life toward management.

So we continue to think about that question, and we welcome any suggestions on how to do a better job, but right now we think the most accurate way to handle it and make sure we are not adjusting -- What we don't want to do is adjust for a patient -- or include patients who transition to a hospice status due to poor quality of care.

So we really looking for indicators we can get at or close to admission, and we welcome suggestions, but I think we reaffirmed our standing approach for now.

MEMBER RHEW: I'm sorry. Could you clarify. Did you say that it is only at the time of admission? So if a patient during the hospitalization was deemed a candidate for hospice and then they were sent to hospice, they would be included in the measure for mortality?

DR. DRYE: They are included in the mortality measure, if it is not at on or before the day of admission.

MEMBER RHEW: That is only if they die within 30 days, though.

DR. DRYE: Right. I mean, they would be included in the measure. Their outcome is what it is.

CO-CHAIR WEISS: A question I asked yesterday broadly about this dropping of the age group down to age 18 in the COPD environment. At least in my mind, an 18-year-old with COPD feels like a different thing in terms of clinical scenario than an individual who is older.

What have you learned so far as you have dropped the age to 18 in the analysis you have done in terms of how much that younger group is contributing to this, and is that contributing enough that it is a real -- it is important to put those lower age groups in, or are we this just as a political

gesture, so we can say it is 18 and older?

I don't quite understand why we are going down so early in age.

DR. DRYE: That is a good question about COPD, in particular, I think. I am a pediatrician by training. So I confess to not being an expert in any way in COPD.

I can't remember. I am going to ask my colleague, Laura Grosso if she remembers, but I know that at least one COPD measure we looked at -- or it was maybe in the literature -- looked at 40 and over for COPD patients.

I think it is good. I just don't think you see many patients in that age group with that diagnosis. I guess the question is the ones that you are seeing, if you just went to 18 and over, would including them sort of create a bias against hospitals that took care of certain kinds of younger patients who had obstructive disease?

It is a good question. We didn't

look at that age group specifically, but the measure -- I think that you probably could apply it. You could draw that cutoff wherever you felt was clinically reasonable. It was cleaner for us to specify it at 18 and over.

CO-CHAIR WEISS: I am just thinking, if there is centers of excellences in pulmonary medicine who really were tackling these difficult early diagnosed patients. I don't know if there are that many of them.

DR. DRYE: Yes. I can look and get back to you on that on what we are seeing in the California data, if you like. I don't know off the to of my head, and we can look at the death rates there, too, in that 30-day outcomes in that age group.

CO-CHAIR GROSSBART: Mitchell?

MEMBER LEVY: I want to go back to my question about tracking rates over time, because we all assume this is a quality measure, but do you look at the hospitals that are outliers and see if the reporting has

changed their rates of readmission or mortality; because we all assume it is a quality indicator, but if over four years in the hospitals that are outliers for both readmission rate and mortality there is no change, I just wonder what the effect of the reporting is.

DR. DRYE: Right. I don't think we know yet. Because they are outcome measures and it is important to get as many cases as we can to get the reliability of the measure results, CMS uses three years of data when they publicly report the pneumonia mortality measure.

So this year, when they put the results out for 2012, it will be basically on 2009, '20 and '11 data. So there is a lag in the effective quality improvement efforts. We are starting to look at those shifts. In readmission, it has just been really recent.

I can tell you, we know that there is a lot -- Let me just shift to readmission.

There is a ton of focus on quality improvement in readmission right now that wasn't there several years ago at our own hospital and nationally.

I think everyone is aware of that, but whether those high outliers, which I think is a really good question, are coming down, we really haven't sorted that out yet. We need to keep following it.

CO-CHAIR GROSSBART: I was going to add to that point. Overseeing quality in a 24-hospital system, we can't really use the CMS data for process improvement, because it is so old. All we can do -- plus we don't have the post-discharge data, although we are working to get it -- excuse me, the Social Security death files. We are working to start looking at our own rates, but it is really tough, because the data is so untimely.

So we are focusing on in-house mortality, which is something we feel we could control and measure, but we do realize that we

are discharging patients to hospice too late in their life, and we are working on earlier recognition and moving a patient to hospice. But it is hard to drive change with this measure. Payment might help a little, or payment penalties.

I do have one technical question. As I read the measure specifications for readmissions, an index admission is defined as not being preceded by an admission in the previous 30 days, and a readmission is defined as one or more admissions within 30 days post-discharge.

So is every patient at risk of no more than -- being used no more than once in the numerator? So in other words, if I -- Does one patient discharged within 30 days and then readmitted once have the same impact on a hospital's readmission rate as a different patient that is readmitted three times in a 30-day window?

DR. DRYE: Right. So that is a

good question and one we grappled -- You know, we grappled a lot with how to structure this measure, because repeated readmissions or admissions for the same patient are statistically correlated.

So if you put them in, you get your results to some degree, but we need the measure to be actionable. That is, if you really get -- You know, a hospital is really effective at bringing down the patients who are readmitted frequently. We want that to show in the measure score.

I don't mean to confuse you with this answer, but I am just going to contrast a bit here. For this measure, you are exactly right. I appreciate your careful reading of the spec.

If a patient is admitted, as you mentioned, January 1st, and they are readmitted twice in January, the outcome is just binary. Were they readmitted once or more, or not? We don't take those next two

admissions, the readmissions, and use them as index, really, basically, for statistical reasons. We are being careful statistically. But that patient could be in the dataset more than once in a year, because they could get admitted again in February, March, April, May, and every time we move out of that 30-day window, we will take the next admission.

I will just say that, for another measure that is actually before NQF right now, too, we have a hospitalwide readmission measure. We made a different decision to allow every admission to count as an index, even it was a readmission. We did a bunch more analyses to see if that is really problematic, and the trail seemed like the right one to make.

So, yes, you can be in more than once, but not in the same essentially 30 days.

CO-CHAIR GROSSBART: Thank you. That was very helpful. We have given the developer a tremendous amount of time, and I

do think we need to move on. So are there any final critical questions that need to be asked? Charles?

MEMBER STEMPLE: As a health plan, we focused on the admissions and, not to toot our horn, but last year at CHF, our readmit rate decreased by 18 percent. In commercial population, we took it down 11 percent for readmission rate, because it was our number one clinical focus.

So I think as hospitals become accountable and ACOs and medical homes and all these things take grist for the dollars, whether there has been an improvement to date that we can see, I would certainly anticipate renewed energy from the hospitals that now are at risk for these readmit dollars going forward.

So though today we might not see huge impacts from this, I assume over the next two years that we would see hospitals really focusing on this area, and the data is

important just from future going, because I would assume the hospitals who are going to lose those readmit dollars are going to be very focused on these.

CO-CHAIR GROSSBART: And I would agree with you. My job is on the line, if we don't move our numbers next year. All right, with that, let's turn to our measure assessment.

Also, as we move on to the pneumonia outcome measures, I know there is going to be, in terms of impact and evidence, a lot of redundancy in our data for pneumonia readmissions, pneumonia mortality, and same thing with COPD.

So if we could -- I was going to say, if you guys could tag team a little on some of the early discussion, and we will merge them together and then we have to vote on them separately, but just at a high level. So to start off, I think, John, you are up first for 30-day, all-cause risk-standardized

mortality for pneumonia.

MEMBER PELLICONE: Yes. I think we have heard mostly about the importance. The only other issue here is the importance of taking outcome related to hospital care. That is the obvious message.

MEMBER STEMPLE: And for the readmit rate, I think we have heard everything. I think it is very important data. The one thing that was missing, at least the data I showed in the California model, they did look at disparity, different groups, and basically, the data seemed to be the same across all different groups, socioeconomic, race, etcetera. So I think they did a good job of validating that the outcome is agnostic to various differences that you might put in. So I think the data element sets were very good and valid.

CO-CHAIR GROSSBART: Well, with that, let's go into our more detailed review and voting for the pneumonia mortality

measures. Let's start with the impact question. So, John or the Work Group, any additional comments you want to make about impact? Any questions from the Committee?

Well, then let's vote. Again, a one to four scale with one being the highest.

How are we doing on the count? Everyone -- Down one? Okay. Get a couple more votes. Try voting again. So we have 18 with a rating of High and one with a rating of Moderate. No other votes.

Then moving to the -- You would think I would have this memorized by now. Moving to our next category, the performance gap and opportunity.

MEMBER PELLICONE: With regard to the mortality rate in the 2007 to 2009 report, there was a significant gap and, importantly, it was not linked to the proportion of minority patients being treated.

CO-CHAIR GROSSBART: Any further comments from the work Group or the Committee?

Hayley?

MEMBER BURGESS: I may have trouble articulating this, but I guess my question is: Over time, now that we have had a chance to play with this data, if you will, have you gone back -- this is to the developers -- Have you gone back and looked? Is there a way to correlate the all-cause mortality back to the pneumonia?

I am from a hospital system. So I worry about this. Right? That, if they get hit by a truck or if bad things happen in other circumstances, we are getting blamed for this. So I am just curious. If the ball hasn't moved that much over four years, have you gone back to look at -- I mean, is there a way to look at all -- You know, whatever it is of the cause of death, can you map that back?

DR. DRYE: So we chose to go with all-cause mortality rather than -- I think the alternative would be pneumonia related mortality -- because when you look at, as you

are suggesting, the causes of death, it is not just one thing for these patients. So for both mortality and readmission, you don't want to try to sort out what was the aspect of quality of care potentially that would have marginally affected this patient's risk of death.

We are not -- These are not measures where the goal is zero. We know that there are going to be patients who die from mortality, particularly in the Medicare population, and that rate is not -- it is not going to go to nothing. What we are trying to encourage hospitals to do is to lower the risk of mortality across the board with respect to any of the patient's conditions or any of the risk factors.

Random events, we don't think, are going to influence the rates too much year over year. So it is completely bad luck and, you know, your patient gets hit by a car. That is not going to -- It is not something

that should be sort of -- You know, this report is only affecting one hospital over another, but for both mortality and readmission, the goal is to try to lower risk and look at the patient as a whole.

When we focused in on related causes, first, it is hard to know what is related, if it was a medication, too much medication, the patient fell and broke her hip. Is that related or unrelated? It may not be related to pneumonia, but it is related to the care.

So we stay with all-cause, because it is most consistent with our goal of sort of whole patient care and lowering risk across the board, but you have to accept that the rates are not going to go to zero.

CO-CHAIR GROSSBART: Kevin?

CO-CHAIR WEISS: This is a great question, because as you think about competing risks across the age spectrum, as we all know, they vary dramatically. If one were to do

just a simple frequency distribution of causes of death for these 18-year-olds versus these 85-year-olds, it is a different list.

For a health system, interventions are going to look dramatically different to try and actually impact. So there is a hazard that was created when you went from a very tight age range, age banding, to a very broad age banding in terms of what it means in terms of how you can actually intervene on this process when you deal with an all-cause mortality.

I think we will address that. It technically comes into ours as to usability, I guess it would be, in some sort of sense.

CO-CHAIR GROSSBART: Any other questions or comments about performance gap then? Go ahead.

CO-CHAIR WEISS: So what would be thought -- I mean, since you are here, it would be great, because your group does a lot of thinking, and for the folks on the phone:

What would you all think is a good rate? We know zero is not the rate, but is there a theoretical good rate that we should be going toward? If one looked at preventable mortality within this bandwidth of 30 days, any idea what we are aiming for, or just a best practice?

DR. DRYE: This is a measure of relative performance. So we are trying to assess hospital performance relative to hospitals with similar patients, patients with similar risk factors. You can look at the distribution, and it centers always on the average rate in the nation, which for pneumonia is -- hold on; I am trying to find the distribution for you.

So what you can do is look at the lower end. You know, look at the 25th percentile, at the 10th percentile, and see where are those hospitals. What is their rate when they are doing really well? It may just be one or two percentage points down.

I would say that is true for mortality where, I think it is fair to assume that hospitals have tried for a long time to try to do as well as they can on mortality.

In readmission, our sense is there hasn't been a focus, as we all know, on reducing readmission risk in hospital care until recently. So there, we think -- We don't know what the target is, because we really want to bring that whole curve down. We think it is high, and with some focus we should be able to get the whole distribution down.

DR. BURSTIN: Just one brief response. The other thing we have found is really, for almost any adverse event, unless it is classified as something so serious and incredibly unusual, this is very typical. The C section rates, episiotomy in the perinatal world -- very similar. It is hard to know what the target is, and I think the response is really most appropriate, really just

looking across hospitals and starting to see trends.

Whether we are actually moving the curve down is really, I think, the key to those, but it is that question we hear every time one of these rate based measures come up that don't have a clear target.

MEMBER LEVY: Now I am really confounded by so many different factors. That is the thing that makes us all so nervous, beyond risk adjustment. Once we start publicly reporting it and it is pay for performance, it is what everybody complains about our field, that we are leading ourselves down a path.

CO-CHAIR GROSSBART: Well, that is a much broader philosophical question that I don't think we want to -- and most of us want to catch our planes at least by tomorrow. So let's move on with the work at hand, and performance gap. I believe we are ready for voting. So again, a one to four scale. This

is the pneumonia mortality measure.

The vote is 13 with a rating of High and six with a rating of moderate.

Then moving on to the evidence for the measure. Any questions or comments?

MEMBER PELLICONE: No evidence per se other than the rationale regarding the need to think comprehensively for the patient's overall care.

CO-CHAIR GROSSBART: And it is an outcome measure. Any comments from the Committee or the Work Group? Hearing none, let's move on to voting, and this is a one to two scale, Yes/No, three for insufficient.

Fifteen, Yes; and four, Insufficient.

Now we move to our reliability and validity questions. So reliability first. John, any comments?

MEMBER PELLICONE: if I understand it correctly, I believe there is a built-in reliability test here in that they do a random

subset and then retest. So that is where the reliability was, and apparently it was rated as moderate.

CO-CHAIR GROSSBART: Any comments by the Work Group? So the Work Group selected moderate. Any questions by the Committee?

MEMBER LEVY: Has the logistic regression risk adjustment ever been published

DR. DRYE: For the pneumonia measures, there are two papers in the literature. I can give you those. I think they are in the -- Hopefully, we put them in the application or I can give you the citations. For COPD, we are still working on those.

CO-CHAIR GROSSBART: Any other questions or comments? Well, then let's move on to the reliability question, again a one to four scale.

Five, High; 13, moderate; one, low. No insufficient.

And validity of the measure, again

a one to four scale. Before we vote, John, any comments?

MEMBER PELLICONE: No.

CO-CHAIR GROSSBART: Work Group? Committee, questions? All right, let's move on with our voting, again a one to four scale.

Seven, High; nine, Moderate; two, Low; one, Insufficient.

Now we move on to the usability and feasibility sections. So in terms of usability.

MEMBER PELLICONE: There was a dry run in 2007 before it went completely public to the hospitals. It appeared successful.

CO-CHAIR GROSSBART: Any questions or comments from the Work Group or questions from the Committee? With that said, let's move on to our voting, again a one to four scale.

The vote was 13, High; three, Moderate; two, Low; one, Insufficient.

And now feasibility.

MEMBER PELLICONE: I think the point here is that there is access to more data in the CMS group than there is in the general all payer, over 18 group.

CO-CHAIR GROSSBART: Any questions from the Work Group, comments from the Work Group, or questions from the Committee? All right, let's move on to our voting, again a one to four scale.

Fifteen votes for High; one, Moderate; two, Low; one, Insufficient Information.

Now the overall vote: Yes or No question. One is Yes; two is No.

We have 17 in favor of endorsement, and two opposed.

All right, let's move on to the pneumonia readmission measure. Charles, you are up, and we have already had the introduction. So I think we can go into our voting sections. So beginning with the importance questions.

So do you have any specific comments about the importance that you want to add?

MEMBER STEMPLE: Nothing more. Just the Work Group clearly felt this was an important measure as we move forward looking at readmissions.

CO-CHAIR GROSSBART: Any questions for the Work Group from the Committee or any comments from the Work Group? With that, let's vote, a one to four scale on the importance of the measure or the impact of the measure.

We have 19 votes High; No other votes.

Then the performance gap?

MEMBER STEMPLE: The readmission rate as we have talked about now, at least Medicare reports out 18.2 percent in the Medicare world. Since this is a new measure for the commercial population age 18 and above, we really don't have background right

now but, clearly, there is a performance gap. As we talked about the opportunity to improve, I think, is in the future in that we haven't seen improvement over the past two years. There hasn't been dollars at risk in the hospital system. So I think that has been a key driver of lack of improvement.

CO-CHAIR GROSSBART: Any questions or comments from the Work Group or questions for Charles? All right, let's move on to voting. One to four scale again.

The results are 13 with a score of High; five with Moderate; one with Insufficient Evidence.

Now we are moving on to the evidence, and again this is an outcomes measure. Charles?

MEMBER STEMPLE: I have, really, nothing more to say. I think the evidence is there, and what I think wasn't brought out, that there is a 12-back look-back, and each member's claims to risk adjust that particular

hospital and that particular hospital system. So there is an extensive risk adjustment that has been validated. So the evidence is pretty good.

CO-CHAIR GROSSBART: This is a Yes/No question: One, Yes; two, No.

The results are 19 voting Yes; no negatives.

Now we move on to reliability and validity section of our voting. So in the area of reliability, Charles, any comments?

MEMBER STEMPLE: Nothing, really, to add. I think the data has been well validated, and I think the Work Group felt it was very validated and reliable.

CO-CHAIR GROSSBART: Any questions or comments from the Work Group or the Committee? With that, one to four scale on reliability.

We have a vote of 14 High on reliability and five Moderate. No other votes.

And now validity? Charles, any additional comments?

MEMBER STEMPLE: No, not really.

CO-CHAIR GROSSBART: Any questions or comments from the Work Group or the full committee? Let's move on to voting then.

In terms of validity, we have 11 votes High, seven votes Moderate, one Insufficient.

Now we move on to our -- I'm sorry, that was usability -- or now we move on to usability. Okay. Moving quick there. So, usability and feasibility are coming up. Usability, any additional comments?

MEMBER STEMPLE: Again, the Committee felt that it was very high and rated this very high. Really, as we have talked about, the data, I think, will be more critical as we move forward, and particularly expanding it to all populations over 18 and not just the Medicare population.

CO-CHAIR GROSSBART: Any questions

or comments? Kevin?

CO-CHAIR WEISS: So this is where my brain is giving me a strange itch, because it is just at 18 -- Extending of the population makes -- It just doesn't -- For mortality, rates are low. Deaths, in particularly deaths around 18-year-olds, are sentinel events anyway. You really should track them down, mobility for a person who has been in with pneumonia probably represents something that may be associated, but the likelihood of the next hospitalization for an 18-year-old having anything to do with that pneumonia is just, from a probability of frequency distributions of hospitalizations from 18-year-olds, is pretty darn low. It is going to be trauma. it is going to be trauma/alcohol related. I mean, it is not going to be pneumonia related.

The same thing for the 22-year-olds, 30-year-olds. You are not going to start until you get to mid-forties and early

fifties before that a readmission for pneumonia has a real likelihood of having anything to be associated with the care of the pneumonia that took place 30 days earlier.

So I think that, when it was developed as a remission -- in my mind, and this is where the itch is, is that for a Medicare population totally makes sense. If you are admitted for pneumonia, you probably have got some sort of a pulmonary thing going on, maybe hip fracture, all those things that we know of in older population's morbidity risk is drastically different for a readmission risk in a younger population. I don't see -- I think it is compacted for mortality. So I wasn't as jittery in my mind.

So I am just a little discomforted here on the usability piece for the readmission for pneumonia.

CO-CHAIR GROSSBART: Any other questions, comments?

MEMBER BURGESS: Can the developer

speak to that?

DR. DRYE: Yes. Again, we haven't -- We could come back to it. We haven't looked at sort of different age called out specifically. I would say that the risk adjustment variables that predict mortality in readmissions do better in the younger age group, I think, because when there is a comorbidity, it means more. Right? There are fewer patients in that 18 to 65 group that have comorbidities.

So in that group, the model is discriminating well against who is at risk for mortality and who is at risk for readmission, and even better than it is in the older age group. But beyond that, I think if you have specifics about what we are seeing in this California data, we could come back with answers to those.

Does anyone on Yale have anything to add?

DR. BERNHEIM: No, but I agree

with Suzanne. I think the one thing that we could do, that we have done for the other populations, is evaluate how much the baseline risk of admission is up in the 30-day period after a pneumonia admission, because this is what we have done in the older populations.

You know, the trauma and accidents has nothing to do with follow-up care. You would expect the rate of -- the sort of baseline rate of admission to go down immediately, and we haven't done that with different cutoffs, and we certainly could.

CO-CHAIR WEISS: So I see that response as a really strong response for validity. I think you have done your work here. There is no question that this is a measure with a risk adjustment that seems valid.

It is the usability issue that I am thinking about, and that is: So we have an 18-40-year-old readmission for pneumonia all-cause readmission, and in an older population

I am thinking, well, that is probably related to comorbidity and probably a higher degree of repeat pneumonia. But what does it mean to have this usability? What I do with that information in younger populations for an all-cause, and how am I going to intervene with it, if I was in Steve's shoes where he is trying to change a whole hospital system around it?

So I don't know. It is kind of an interesting question. When you drop the age group, it opened those issues up in mind, at least.

CO-CHAIR GROSSBART: But at some level, addressing it from a hospital perspective, these are rare and random, and the rare and randomness isn't driven by the hospital demographic. So it is not making a difference in overall hospital rates, because it is just random noise.

It is like being hit by the truck or falling when you are in the parking lot

walking out. Of course, a social worker might have been able to help on some transportation home.

CO-CHAIR WEISS: But why introduce noise into a measure system when you don't have to?

CO-CHAIR GROSSBART: Well, because you do have to, because you can't -- See, you are saying we'll make the age older, but it is not -- Does it truly negatively impact the usefulness of the measure from a provider standpoint? I am not going to focus on my 18-26-year-old readmissions, you know. I am going to focus on those one out of every two patients that we don't connect with a doctor in the first 30 days after they go home.

MEMBER STEMPLE: And I think, you know, as we move into ACOs and other accountable organizations, they are stratified for their global readmission rate, and it is not broken out to age-specific categories, and I think assessing -- there may be a lot of

noise and interference, but whether it is a contributor of one-half of one percent to the overall, as we are looking at organizational performance globally across the country, we are not age stratifying outside of, quote, "commercial Medicare."

So I think every other measure that I am aware of, basically, starts at age 18 out of the pediatric age group and goes up to the adult age group. So I think that is just in concert with other ways we are looking at performance measures.

Admittedly, the background noise of the trauma should equalize across, as we have said for other measures. So even if it is a small contributor, I think to try to take other measures now and define the age population where it may have more a critical element is not how we are in this country looking at performance measures. To substratify into age range just complicates the whole system.

So I think the standardization of the methodology and using age 18, as we do for the vast majority of things, seems to make sense, just from a methodological effect. Makes sense to me from my world of managed care where I don't stratify my physician risk group by -- I look at their readmit rate globally. It is not cut out to different categories, different ages, and the complexity, at least of me, to measure that if I was only looking at 40 years and above would be a very difficult thing to do.

CO-CHAIR GROSSBART: Brendle, you had a comment?

MEMBER GLOMB: Yes, Stephen. You said in safe group, rare and random, except in hospitals that specialize in taking care of cystic fibrosis patients, it is a group that is going to be not only not rare or random but somewhat expected. So both three admissions and mortality will adversely affect those hospitals' numbers who do specialize, and

there are very few who are willing to take care of these patients.

CO-CHAIR GROSSBART: And does the risk adjustment model adjust for them?

MEMBER COHEN: CF would be counted as -- Cystic fibrosis would be counted as a cystic fibrosis related exacerbation, not as pneumonia. At least, that is how we call them, because we have a very big CF center.

MEMBER BURGESS: I would like to speak to the 18-year and older thing. Because it is tradition, does that make it right, because we have talked about COPD an including 18 to 40. You know, NCQA has some data around that. Their data was very muddy in that space.

I know we are not talking COPD right now, but I am struggling a little bit with this, that we are saying it is okay, because it is traditionally how we do this. This is the committee's -- This is our responsibility to think about is it the right

direction that we are going in.

So we have an opportunity to speak to that now versus to say, you know, it is okay. I don't know. It does not feel quite right to me in this space. So for the record.

CO-CHAIR GROSSBART: Helen, do you have a comment?

DR. BURSTIN: I was just going to make the comment that it has actually not been the tradition. The tradition has been these measures have only been limited to 65 and up and, in fact, it is through the encouragement of private purchasers and plans and others who said they want to be able to have a measure that works like this.

We have actually encouraged Yale to do the analysis to show the risk models work. The measure, as it is specified, is still -- the data available to run these measures remains 65 and up. The key was saying does the risk model work? Is there something different about the under 65

population?

MEMBER ALMENOFF: We have actually been running a risk model for anybody over the age of 18 for the last six years within the VA. So we basically look at 7-800,000 admissions a year in a risk adjusted outcomes model, put the data out quarterly, and it is anybody 18 and over. Usually, it is 19, because it is hard to get into the military and get out that quickly.

One thing we did do is we risk stratified the categories. So we have five categories of severity of illness, and patients with less than a 2.5 percent of dying. That is a very low risk, and so we actually categorize that out and give it to site. So it usually is misadventure. I think that shouldn't happen, because low risk patients shouldn't die.

So address like 18-year-olds that die in a hospital, that would probably pop up in our lowest group. So they actually will

look and review all the deaths of patients who died who were low risk, who shouldn't have died.

So there are a lot of ways to sort of adjust this, but to just look at an isolated 65 and older population isn't that useful either, because we have that whole major group in the fifties and forties that have very high death rates, and you can't sort of say I am going to cut it at 30. So you just -- I mean, 18 is probably arbitrary, but you have to start at some level.

MEMBER HAECKER: Pediatricians in the room feel that way. It is arbitrary .

MEMBER ALMENOFF: Yes, but you guys need to take care of stuff up to 18, for some reason. I don't know.

MEMBER HAECKER: Actually, we don't. It goes up beyond that.

MEMBER ALMENOFF: No? You go up to -- Cystic fibrosis, you go to the thirties. Right? Yes. It is still your workload.

CO-CHAIR GROSSBART: I would like to move the conversation on, and move to our usability vote. So unless there is a critical urgent comment that needs to be made, let's -- and we didn't make our 15 minute timeline there. So let's move on with the voting, usability, one to four scale.

Nine, High; six, Moderate; three, Low; two, Insufficient.

Then finally validity. Charles?

MEMBER STEMPLE: I think we discussed that. Thank you.

CO-CHAIR GROSSBART: All right. Feasibility. Feasibility, I'm sorry. Any questions, comments? All right, let's move on to voting. One to four scale.

Everyone vote one more time, see if we can register.

On feasibility, we have 17 High; two, Moderate; one, Low; no Insufficient.

Finally, our overall endorsement: One, yes; two, no.

We have 18 in favor and two opposed.

We now have two new measures that we are evaluating, the COPD Risk-Standardized Readmission and COPD Risk-Standardized Mortality Rate. As in the case of pneumonia, we are going to try to create some economies of scale by giving a brief overview. Jointly, Norm Edelman and I will discuss these measures, and then we will let Norm -- We are actually going to vote on mortality rate first.

Actually, Norm, do you want to kick it off?

MEMBER EDELMAN: Yes. We have had a lot of discussion already that is relevant. These models are very carefully done and very, very well described. They are very strong models. They look at very, very important variables, and they look at measures and variables that will be used very robustly. That is to say, as Mitch pointed out, they

will be used in a punitive fashion.

So it is important, I think, that they get rigorous scrutiny. You know, I feel very positively about much of the work that has been put in here. I have reservations about the risk adjustment, and it applies particularly to COPD mortality. It applies somewhat less to readmission, and even less but not zero, to the pneumonia groups that we just voted on.

So risk adjustments come in two flavors. You risk adjust for the patient, and you risk adjust for the hospital. With regard to the patient, I think there is one omission, and it is an understandable omission, because it is based on very recently accepted concepts in COPD. That is, there is no risk adjustment for previous frequent exacerbations.

Now I understand why the developer wouldn't want to do that, because they consider an exacerbation a bad outcome, but in fact, recent data -- the article by Hurst in

the New England Journal about a year ago, but more importantly, a consensus statement by the people who put together the GOLD guidelines in 2011 -- accept the fact that recent exacerbations is a phenotype of COPD.

That is, there are a certain group of patients, even those that don't have bad pulmonary functions, that get a lot of frequent exacerbations. So if you don't risk adjust for that and you have a hospital with a very strong pulmonary group that attracts people with difficult to manage COPD, then you are treating the hospital prejudicially.

So that is my problem with risk adjustment for subjects.

I have a significant problem with risk adjustment for hospitals and a less -- and a more complicated one. Now with regard to risk adjustment for hospitals, I reiterate the issue of air pollution.

I think the evidence that air pollution is an important cause of excess

mortality for cardiopulmonary disease -- and we are measuring all-cause mortality -- is strong. It has been with us for 20 years. It is significant, and I don't understand why it is hard to do. You just have to go to another dataset, and you have to go to another dataset to estimate SES.

So I don't understand why this is a difficult thing to do. I feel strongly that the concept should be tested. If it proves to be wrong, fine.

The other thing that troubles me -- it is a little more subtle -- is the fact that in the developer's analysis, SES doesn't fall out as a risk factor for hospitals. Now in this meeting, we have had lots of applications referencing lots of papers which show that SES is an important outcome for -- is an important measure for bad outcome in COPD.

I am a little surprised that in this dataset it is not, and I worry there may

be a countervening bias. That is to say, poor people in urban settings get their care in clinic systems and teaching hospitals where they are likely to get a follow-up visit when they are discharged. Poor people in rural areas may not.

So there may be an offsetting issue, right? So a teaching hospital may actually do a better job, because they have clinics, but that is offset by the fact that people in low SES are more likely to have bad outcomes. That is a more subtle issue, but I would be happy if the developer could look into it.

So my concern is -- My concern is the risk adjustment for patients it is not up to date, and the risk adjustment for institutions may have unintended bias.

CO-CHAIR GROSSBART: At this point, I actually ask Helen on the SES question, because it is actually relevant to NQF policy.

DR. BURSTIN: Yes. So to date NQF has encouraged developers not to include race, ethnicity or SES in risk adjustment models, but instead to actually allow to see the effects of those differences so we can see where there are disparities.

So for risk-adjusted outcomes, we actually do not, as part of our evaluation criteria, which you will see, ask developers to include those in, but we would prefer actually to see stratified results, as we saw, in fact, with some of the COPD measures that we talked about yesterday that were process measures.

There were differences, and they were talked about, the difficulties of trying to get the data, but that they should be stratified rather than adjusting away those differences and not being able to see them.

CO-CHAIR GROSSBART: Okay. And then just in my role, kind of high level overview, I think there was in the Work Group

a recognition that readmissions and mortality were both opportunities for improvement.

COPD is a major source of readmissions in the Medicare age population. That was based on the article that came in the New England Journal by Stephen Jencks and others. So there was clearly a sense that this was important, but again, as Norm has noted, there was some concern about the risk adjustment model for both measures, and the Committee was split on some of these areas.

We will go through that in detail in the next few minutes. So with that, are there any questions for either me or Norm from others on the Work Group, any comments that you would like to add? Any Committee questions? If not, then I will ask the developer to respond to the comments that were raised. Elizabeth?

DR. DRYE: I just want to confirm, because I stepped out for a minute. The main comments were on the adjustment for

particulate exposure at the patient level, and then SES. Did I miss anything else? Okay.

The challenge, again, of right now trying to modify this model -- we will look to bring in an environmental factor like particulate exposure, everything, and the county level data. Is that -- It sounds good. If we put it in the model, it wouldn't surprise me if it is significant, but we really need to understand what information that variable would be carrying, and anything we put in our models usually is specifically significant almost, because we have so much data.

So we really want to think about how to use environmental information in a way that is really linked to patient risk factors or to -- I appreciate what you are saying, but factors that are beyond the hospital level control, and we need to usually incorporate those, but we are not -- We really haven't been able to start that process, to get very

far in that process yet.

I can just say that is something that our group is looking at, but it is not straightforward, because that variable will be correlated with a lot of other factors that probably affect risk. So we want to think about that more before we go down that path and understand the data a lot better.

I think Helen already spoke to SES. You do see, and we have reported in the NQF application, that we will get race in and SES by medium income and the patient's ZIP Code, but there are slight differences in the distribution. But there is a lot of overlap.

Many hospitals with higher proportions of low SES patients, as designated that way, do really well on the measure. We really agree with NQF guidelines not to adjust those potential differences out of the measure, because we want to be able to see those differences where they exist. But it is a complex issue. It is another area where we

are looking at different variables and ways to separate potentially the hospital and patient level factors, but we are early on in our work there, too.

CO-CHAIR GROSSBART: Rubin, go ahead.

MEMBER COHEN: Just wondering. Looking at the risk adjustment, there is a lot on mechanical ventilation. Is there anything on noninvasive ventilation, because actually a lot less COPD patients are being intubated, and a lot of them now are carried on noninvasive ventilation. Is that part of ventilation, because all I see is mechanical?

DR. DRYE: Yes. That is a good point. So just to speak to the other comment about history of admission, we don't usually adjust for that, but we do adjust here for history of mechanical ventilation, and we did capture -- and, Laura, if you are there, I might need to confirm -- CPAP codes, for example, in that set of codes that indicate

mechanical ventilation, for that exact reason.

DR. GROSSO: Yes. Yes, we did account for invasive and noninvasive.

CO-CHAIR GROSSBART: I would like to take the Chair's prerogative. With regard to the risk adjustment, clearly, the Work Group raised questions, and I think it would be much easier for many of us to endorse this measure if we had a firm commitment from the measure developer to, one, do a thorough literature review and, two, to test the hypothesis.

I realize how many million patients do you have in your database, that you throw anything in there, you are going to get a positive p value. That said, you can test hypothesis and either reject the null hypothesis or fail to accept that. I can't keep track -- you know. Just test the hypothesis.

DR. DRYE: Sorry. Are you speaking specifically about the use of the

county level particulate data? I think we can look at that. I don't know how easy it is to get that data, but I think it is probably not too hard.

Let me just -- I don't know if I can confirm that on the spot. I don't know if CMS is on the line, but we just need to confirm that that is doable in a reasonable time frame.

CO-CHAIR WEISS: Just as a person who is experienced in working with that kind of data, it is messy, because for anyone who has been in that environment, it depends upon the monitors are; and even though they get county data, it really is an average.

It is just -- It is not a clean data. So you will find significance because of the size. It is really the impact, and if the impact of it is small, you don't know if it is because of the lack of factors, because of the lack of measurement capacity.

It worked in the Six Cities Study,

because they had put in monitors and measured. I just don't know that -- I think we don't want to send them too down the primrose path here, but it would be great to see it.

DR. BERNHEIM: This is Susannah. Can I just add one other thought? This is Susannah from the Yale team.

I think you hinted at this, but I would say that our other concern is not only about the ability to actually get this data, but also how well it might travel with other risk factors that we wouldn't want to risk adjust for.

So I think we would have some difficulty disentangling those. So I think we would need to feel pretty confident that it was likely to overwhelm the signal of the preventable or potentially preventable deaths due to the illness and the hospital environment and the hospital care before it worth CMS embarking on an expensive study.

I think we do need to take some

time to think carefully about whether we are going to be able to come up with an answer that is important and meaningful, and really likely to change the results of our measure.

CO-CHAIR GROSSBART: Thank you for the comment. As Dr. Weiss pointed out, it is not whether you find a significant relationship, but it is the magnitude of the relationship that counts. I hope you will commit to doing at least the pilot study to get some sense of whether this is an important issue or not.

CO-CHAIR WEISS: I think NIEHS would be very interested in taking a look at these sort of things, and this is not around these two measures of pneumonia and COPD. This is all-cause mortality, and it really reflects the whole measurement suite you are building.

So there is a real opportunity here. You have got some wonderful environmental scientists at Yale, and probably

gives them a whole new research trajectory to go onto. So this is a growth industry. We strongly encourage it.

CO-CHAIR GROSSBART: With that, I think we should move on. We are in our voting right now. So this is for the mortality measure, impact. Do you have anything else to add, Norm, or should we move on?

MEMBER EDELMAN: Well, I just want to say I am delighted to have generated a growth industry. We really need this in our current economic time. The impact is high.

MEMBER BURGESS: Can I ask Norm a question right quick before we start?

Norm, yesterday you raised a question around appropriate diagnosis of COPD, less than age 40.

MEMBER EDELMAN: I think Dr. Weiss raised that, but --

MEMBER BURGESS: Weiss? Did you raise that? Anyway, I would asked either of you to speak to that. They have looked at it

in California, the data. I can feel comfortable with that, if you all who have expertise in this feel like that is the right.

MEMBER EDELMAN: There is a spectrum of airways disease, starting with asthma, ending up with honest to goodness COPD, with a whole bunch of stuff in between. The British have a term, asthmatic bronchitis, which we don't like to use, because it confuses everything, but it is real.

That is a problem we have, and it is not just a classification problem. It is a pathophysiologic problem. The only point that I made yesterday was I don't think age changes that problem.

CO-CHAIR WEISS: Probably the only little bit I could add to that: I spent a few years at the National Center for Health Statistics asking why. It was a great opportunity, but they had work done on a comparability study around mortality records.

This is where they actually go in and they look and see. There is a huge amount of confusion in the under 40 when it comes to asthma/COPD, in terms of what was reported as mortality versus what was on the death certificate and how that death certificate rattled up to actually say underlying cause of death. You get into some technical space here.

They are using all-cause mortality. So it kind of washes that problem out here, but it does bring in the other problem. That is, when you get the underagers, you got this competing interest problem in terms of usability, which we talked about, and I don't want to open up again.

CO-CHAIR GROSSBART: And Dianne?

MEMBER JEWELL: This issue that you have raised is going to come up again when we look at the competing measures or related measures. The National Center for Health Statistics indicates that between the ages of

18 and 44, about four percent of the adult population has COPD exclusive of asthma, which translates into about a million people.

So the potential for impact, while it might be hard to find them in individual centers and practices, it is not a small number, I would say.

CO-CHAIR WEISS: Now that comes from self-reported information or from -- Is it from NAMSI or from NHIS? Do you know? Is it a health interview survey or is it the ambulatory care documented records of doctor's diagnosis, because that will make a huge difference? I would guess it is from the NHIS.

MEMBER JEWELL: Source is NCHS, Health Data Interactive and National Health Interview Service.

CO-CHAIR WEISS: So that is self-reported. So that is where people think they have got COPD.

MEMBER JEWELL: Okay. Thank you.

CO-CHAIR WEISS: Four percent, which means it is probably much less.

MEMBER JEWELL: Okay. Thank you. That helps.

CO-CHAIR WEISS: That is the whole thing of it.

CO-CHAIR GROSSBART: Okay. Let's get our first vote done on the impact, one to four scale again.

We have 18 votes for High and two votes for Moderate. No other votes.

The next question for us is the performance gap. Any additional comments?

MEMBER EDELMAN: Yes. The range of in-hospital mortality is two to five percent. So I think the performance gap is, at best, moderate.

CO-CHAIR GROSSBART: That is in-hospital and 30-day mortality.

MEMBER EDELMAN: I'm sorry. Thirty-day mortality. Great.

CO-CHAIR GROSSBART: Any questions

or comments from the Committee? All right. Then let's move to voting.

We have three votes for High, 13 for Moderate, four for Low.

Then the assessment of the evidence. Again. this is an outcomes measure. One for Yes and two for No. Any questions or comments, Norm, from the Committee? All right, let's move to the voting.

DR. DRYE: Can I just correct that -- You have already voted, but the range of mortality that we have in the Medicare data -- the risk adjusted range even goes from six to 13.5 percent across hospitals.

CO-CHAIR GROSSBART: I appreciate that, but at this point I think we just want to move on.

We have 18 Yes; one No; one Insufficient.

Moving on to our next set of questions, reliability and validity.

MEMBER EDELMAN: It is easy to measure.

CO-CHAIR GROSSBART: Okay. Any questions, comments from the Committee? Reliability: Is the measure reliable? So moving forward, again any questions, comments? Moving forward, let's vote on a scale of one to four.

We have 17 High and two Moderate -- three Moderate, I'm sorry.

Then validity questions coming up next?

MEMBER EDELMAN: For the reasons discussed, I don't think the model as presented is valid.

CO-CHAIR GROSSBART: And the overall Work Group, comments? Others from the Work Group, and again as one Work Group member, I agreed with the need for further testing, but I thought the validity of the model was much stronger than Norm. Here we go. One to four.

We have two High; 10 Moderate; five, Low; and three Insufficient.

So we get to move on to the next vote, which is usability. Norm?

MEMBER EDELMAN: No comment.

CO-CHAIR GROSSBART: Any questions? Scale of one to four.

Eight, High; nine, Moderate; three Low.

Finally, feasibility. Any comments? Any questions. Scale of one to four.

We have 12 High; seven, Moderate; one, Low.

Our final question is on endorsement, one for Yes, two for No.

We have 17 in favor of endorsement, three opposed.

The last measure for this part of the agenda is the 30-day, all-cause, risk-standardized readmission. I know we have gone through a lot of information and summarized

this very thoroughly. So I would like to just move into the voting.

The first question we have is impact.

MEMBER EDELMAN: This is the mortality one. Right?

CO-CHAIR GROSSBART: We are doing readmission now. We flipped them. So the readmission measure, the importance of the measure and the impact. It is an important measure. Readmissions are a major source -- COPD is a major source of readmissions.

First of all, any questions or any comments from the Work Group? Any Committee questions? Then a scale of one to four, the impact of the measure.

Seventeen rating of High, and one rating of Moderate.

Moving to the next part, the performance scale, the performance gap is significant, nearly a 23 percent readmission rate for this patient population; represents a

total of four percent of all 30-day readmissions. I am looking for the actual range, high to low, and I don't see it -- Here it is. I don't see it off the top.

DR. DRYE: Do you want me to give you the range?

CO-CHAIR GROSSBART: Yes, that would be fantastic.

DR. DRYE: Unadjusted, at the hospital level the range is -- I am going to give you the fifth and 95th percentile, 11 to 32. That is just the 95th and then adjusted, our range is 18.3 to 25.3. that is with a median of about 22.

CO-CHAIR GROSSBART: So a fairly large performance gap even after risk adjustment. Any questions or comments for me or any comments from the Work Group? Committee? Okay, let's move to voting, one to four scale again.

The vote was 15 to 3. So that was correct.

The next question is evidence. I don't have anything to add to what has already been said in the conversation. So unless there are questions for me or the Work Group, let's move forward with voting. It is a 2. It is a Yes/No question, one for Yes, two for No.

The vote was 18 to one on the evidence question.

Moving on to reliability of the measure, again I think we have gone through this. The issues are the same or similar for this as well as the mortality measure. Looking back at how the committee rated it -- the Work Group rated it, we rated it as highly reliable as a Work Group, and so are there questions or comments from the Work Group? Questions from the Committee? Moving on, then it is a one to four vote.

Fifteen ratings of High; two of Moderate; no other votes.

Validity, we have discussed this

extensively already. Clearly, some difference of opinion and some opportunity for the developer to strengthen the model or investigate if the model could be strengthened, I should say. So in terms of validity, are there any questions for -- or any comments for the Work Group to be shared? Any questions from the full Committee? With that, let's move to voting on validity of the model, one to four scale.

We have three votes for High; 10 for Moderate; five for Low; and one for Insufficient.

So we will now move on to usability. In terms of usability, this has been -- Similar measures have been out there on public reporting sites for the age 65 and older -- I should say for the Medicare age population.

It is stimulating a good deal of performance improvement work at the hospital level, and similar measures are stimulating a

good deal of hospital performance work, and the committee, somewhat split, but tending toward high usability, moderate to moderate high -- high to moderate high. There we go, more high than low. Anyway, usability. Any comments from the Work Group? Any questions from the committee? Let's move on to voting.

We have seven votes, High; 11, Moderate; one, Low.

Then feasibility. Again, this is administrative data from Medicare and all-payer databases, and it is a very feasible measure to collect and report. Any other comments from the Work Group or questions from the Committee? All right, let's move to voting, one to four scale again.

A score of 14 for High, and five for Moderate; no other Votes.

Finally, our overall endorsement of the measure: One is Yes, two is No.

Seventeen in favor of endorsement; two opposed.

Now we are -- Where are we on the agenda? Now we are to move on to related and competing measures, and we are only 10 minutes behind schedule, which is pretty good, and yet we still were able to have a very robust conversation. So shall we move into that?

DR. WINKLER: We have had a chance to talk about related and competing tangentially for the last couple of days, and this is where we have to make some initial decisions on determining what is and is not related or competing. That may seem simple, but actually, it is not.

This is a deceptively simple 2 x 2 table of trying to understand related versus competing measures. The top row is the measure focus, that which is being measured, typically in the numerator. It is the same concept in the first column. It is a different concept in the second column. On the columns on the left, you are talking about the same target population or a different

target population and, of course, you have got the different combinations.

So when they have the same target population being measured by the same concept, that is a competing measure. If you have a different concept or a different target population, they are related; and if neither of those things are true, it is not an issue. We don't have enough that go in that lower righthand bucket.

What I did is I went through in the next slide, and for each of the topic areas we have discussed heretofore looked to see the decisions you have made, and let's look at the measures that you have deemed to be suitable for endorsement.

Asthma: there are six measures. All of these are process measures. We did not have any outcome measures for asthma. I put the Joint Commission measures for inpatient asthma treatment, which you recommended for reserve status, kind of at the bottom.

So what we have left are four measures of medication management of asthma. In this particular case, we have three measures from NCQA -- that is 0036, 1799 and 1800 -- which are a suite of measures from the same developer. They have the same denominator. They are inherently harmonized, but they approach the idea of medication management differently. Nonetheless, we are still talking about medication management.

So that is one level of potential competing. Are these all trying to measure the same thing or albeit differently? That is a decision point for you.

Sort of a sub-question of that is, if you look at measure 0047 and measure 0036, these two measures are measures of sort of the single prescription or dispensing of a medication, as opposed to the newer NCQA measures which were about proportions of days counted and the medication ratios.

So, clearly, measures 0047 and

0036 are really clearly competing measures. They are measuring the same patients and the same thing about the patient. So there are really two issues here for asthma around competing measures.

I just want to lay this out for you, and I want to go through the three topic areas, because each one has a different nuance that is sort of interesting, to sort of set the stage so we can determine what decisions around competing and related we need to make.

The one on asthma are really two questions: Are the four measures to be looked at as competing, and we see which ones among them really are best suited to go forward or do we look at them differently, and the competing measures are really the first two?

Going to the next slide for pneumonia: The pneumonia measures are a completely different animal, because as we in great detail discovered yesterday, the measures from PCPI -- or the measures of

patients with community acquired pneumonia fall into two buckets, those that get admitted to the hospital and everybody else that gets treated in some outpatient facility.

So the PCPI measures are all about that second group who are treated in the outpatient world and are not admitted to the hospital. So the group splits into two. Because those are different target populations, these measures are related, and we have two that are very specifically related in that their numerators are very similar, and that is 96 and 147.

They are both talking about empiric antibiotics therapy or initial antibiotic therapy. So there are some opportunities for harmonization, because they are related, that are relatively -- I don't want to put a quality on it, but they are pretty straightforward, as they both adhere to the same guidelines. So pneumonia is a completely different question than asthma.

Then we go to COPD, which is completely muddied, in which we have two measures of the spirometry. We talked a little bit about their differences. Clearly, 91 and 577, same focus of measurement, same target population of patients.

The next two measures are around medication therapy, although they are asking a slightly different question. So perhaps they are more related measures, and that will be a decision. Of course, the conversation we have had all along around harmonization of all the measures for COPD, because they are related by virtue of addressing COPD, is the issue around age.

So there are a fair number of questions for us to try to determine what the issues are for competing and related that we need to tackle. So it is kind of like an iterative process that we will have to really determine.

We have got about 45 minutes to

try and perhaps go back to sort of answer the basic questions of which ones do we consider competing, and what decisions will we need to make. Whether we can make them today or we will need to postpone that for later is to be determined.

Can we go back to asthma?

MEMBER JEWELL: Does the level of analysis play a role in any of this decision making?

DR. WINKLER: Yes, it does.

MEMBER JEWELL: Relative to who is being measured?

DR. WINKLER: Well, yes, it does. When you get into the algorithm around decision making of two competing -- or competing measures, actually, yes, it leads you through it. In fact, I have got the examples for you. But I think the question -- We start with the asthma group.

The first question is: You have got four measures about medication management.

Are all four, in your mind, competing measures?

MEMBER LANG: Yes. Having reviewed 0036 and 0047 and presented these yesterday to kick off the session, they are competing measures. I was admiring the metrics proposed for COPD and pneumonia from the standpoint of feasibility.

In terms of asthma, we can't -- It is not a dichotomous outcome in terms of if it is for pneumonia or COPD, in terms of alive or dead, fortunately, as mortality is rare. Nonetheless, this is what we have in terms of it is a process outcome.

So I think, for medication, I would say these are all -- I mean, to try to make what could be a long story short, in the interest of time, because I know we need to move on to the other areas, I would say that there are validity concerns that I have expressed regarding 0036 and 0047, although these metrics did pass. But I think the more

recent metrics, the 1789 and 1800, are more sophisticated, elegant. It is more, I think, the way we should be going, although there are some issues with these as well, and I don't know where we would go regarding some sort of, I want to say, composite metric that we could put together that would more closely approximate one of the more recently proposed metrics, which has also been approved.

MEMBER GLOMB: I want to echo David. I think that the 0036 and 0047 are definitely competing measures, but I did think that -- You asked, is it feasible to look at these today and knock them out. I think this one we could probably look at and knock out.

I had a question as well. Can we use our composite scores as we have graded the individual measures, because these two, more than probably any of the others, are head to head competing measures. Can we look at our cumulative scores on how we graded the two to help us decide? I love it.

DR. WINKLER: Yes, you may. I decided to print them, because there is a lot of information that would look messy on a production.

MEMBER GLOMB: Can I kick off the discussion about the two?

DR. WINKLER: Sure. I guess I was also asking, do we want to look at 0047 against 0036 as the only competing issue or there was some -- David sort of indicated that perhaps the whole question of the single prescription or dispensing that 0047 and 0036 measure versus the newer 1799, 1800, and perhaps you might think that 0047 and 0036 have been both been superseded by newer measures. That was something, as David indicated.

So that is another discussion point and decision point for you all around competing.

MEMBER GLOMB: I think it depends how and where it is used. I think

practicality -- these are fairly simplistic measures, perhaps easier to use than the more complex, and it might be a reason to keep one of the two of these in place, between 0036 and 0047.

DR. BURSTIN: And also they have both been retooled, one example as well to keep in mind for something for meaningful use.

MEMBER GLOMB: That is very good. Just to kick off the discussion between them, I like 0047 better than 0036. I know 0036 is part of a suite there, but it was a little bit tighter in its definitions. I felt that there was less -- There are fewer opportunities for questions of both validity and -- the one that precedes -- reliability, thank you. Brain is tired.

I thought that this measure was going to be a lot cleaner ultimately of the two, head to head.

MEMBER LANG: Just to echo Brendle's comments, I recall that there were

some issues of validity pertaining to some of the medications that were listed, in terms of alternative to inhaled corticosteroid medications that may not be consistent or aren't consistent, I should say, with optimal care.

As you may recall, there was -- I think his name is Mark who was over there responding to my comments, who clarified the issue of the inhaled long acting beta agonists, the inhaled short acting beta agonists from the alternative list. Although the alternative list is kind of apples and oranges in a variety of ways, nonetheless, there is a separate metric for inhaled corticosteroid alone, and it is one dispensing event during the measurement period.

So it is not an ideal measurement, but I would say, head to head, probably 0047 would have the edge over 0036, because with 0036 there are more validity issues.

DR. WINKLER: Any other thoughts?

I just want to kind of go over this table with you just a little bit in terms of the decision making around competing that has been outlined. We presented this to you in several of your briefing memos.

Certainly, the first thing is to compare your evaluations on the different criteria and the suitability for endorsement and how they compare, and no one criteria is it. it is kind of the gestalt of all of them, because all measures have strengths and weaknesses.

Certainly, you can make your decision based on that. There are other criteria that have been identified as being important in terms of looking at measures, and so you if you look at the lower down rows on the table, measures specified for the broadest application in terms of the target population -- in this case, they are fairly similar -- but settings of care -- these tend to be ambulatory measures -- and level of analysis.

So, Dianne, to your question, yes, level of analysis should play into it in terms of some usability, assuming the validity issues, David, you raised are acceptable tradeoffs. So none of this is a black and white, easy to sort through.

Measures that address disparities in care where appropriate, measures with the widest use -- you know, these measures are in use. Both of them are retooled for ERHs and are in the Meaningful Use Program. They have other uses as well.

Measures that are publicly reported: That is a priority at NQF, or for other accountability purposes, and moving toward an EHR type world.

So those are some of the other criteria we are asking you to consider and factor in when you are looking at the two measure side by side.

Helen, did you want to add anything?

DR. BURSTIN: That summarizes it well. I think that the one consideration is that the number 0047 doesn't apply to health plans. Only the 0036 does, and 0036 is also on the CHIPRA list, is one of the considerations.

MEMBER ALMENOFF: I am still trying to understand this. So the goal is to approve one and disapprove the second one? I am trying to figure out how you are -- what you are trying to get us to do.

DR. WINKLER: Yes, because these are competing measures, we would like to select one. If you cannot, we have to have very clear and compelling reasons to keep two that are essentially the same measure on the books.

MEMBER LEVY: You mean select one and combine them?

MEMBER ALMENOFF: No. Get rid of one and keep one.

DR. WINKLER: Pick one.

MEMBER ALMENOFF: Couldn't you even have a scale of -- I mean, it just kind of odd that we are kind of stuck in this situation now when we could have maybe dealt with this earlier on.

DR. BURSTIN: One of the considerations that we have had as we have gone through this process is we don't want to make committees go through a discussion of competing measures until they have actually passed the criteria.

So you have now deemed that both of these are, in fact, suitable for NQF endorsement. They have passed all the criteria, and now you need to get into the discussion of is there one that is best in class and, if there is not, why not, and how do we justify having two.

MEMBER HAECKER: I have another clarifying question. Is the issue of the public reporting the difference? What implications does that have down the road? If

we endorse something that is not publicly reported, what does that mean?

DR. WINKLER: Well, again, as I said, public reporting and other accountability uses are really sort of the cornerstone of what NQF is looking to do, and our stakeholders very much are looking for measures that are publicly reported. That is a high priority for, particularly, the consumers and purchasers and other folks that want to use that information.

CO-CHAIR GROSSBART: Reva, a point of clarification: PQRS is not publicly reported, but for how many more months is it not publicly reported? It's 2014-2015, we expect it to be up live, and it is a pay for reporting right now. If you don't submit PQRS -- You do get a bonus for submitting that data.

Then if we were to vote down 0047, what does that do to CMS? Are they obligated to use NQF endorsed measures for their public

reporting work?

DR. BURSTIN: In general, NQF endorsed measures are the measures they tend to go to first. They are certainly allowed to use any measures if they are not NQF endorsed with justification in the Federal Register. So that is certainly possible.

At times, they will also switch to the endorsed measure when the opportunity arises, but certainly not 100 percent of the time.

MR. HAMLIN: 0036 is also in the PQRS's list.

DR. BURSTIN: Oh, yes. Thank you. Hi, Ben. I was wondering if you were on.

MEMBER YEALY: I had one question for the folks that were on this --

DR. BURSTIN: He said that, just to keep in mind, 0036 is also on PQRS's list, not just the 0047.

MEMBER YEALY: The question for the folks that were on this Work Group,

particularly, and it seems to be a preference now for 0047, but what I recall, and if you look at the scoring, looks like 0036 scored slightly higher. It is not dramatically different.

What I remember from the conversation is that, although there was more granularity in the denominator statement for 0047, there were some definitional concerns, and 0036 was much more simplistic, harder to change.

Make the case now why -- What I recall is yesterday we actually struggled more with 0047. Do you think that it is going to be transformed and, therefore, perform better? I am having a problem getting to the preference.

MEMBER GLOMB: David, correct me if I am wrong. I think that a lot of the discussion that was around 0047 was truly related to what was written here in our submission versus what we were told by Mark

and AMA yesterday.

he cleared up a lot of that stuff, so that as we went through those concerns, they all literally disappeared,s because they had cleaned up the descriptors here. Yes, having asked for the composite score, now that I have got it, you know, it goes against -- it seems to go against what my personal gut feeling and what I had thought was the feeling of the group overall.

Still, I like the -- I think the danger is in the lack of detail of 0036 and the preferred therapy statements versus some specifics. I still have a problem with 0047 not including the ICS/LABA combo with the ICS. I think that is where it belongs, not in the alternative therapies, but other than that --

I just think there is just a lot of wiggle room, so that the actual scoring will be more imprecise ultimately with 0036.

MEMBER YEALY: I view the scores as essentially the same. That is biologic

variation. There is nothing dramatic in the score differences that I see.

MEMBER STEARNS: Is there is something dramatically different in the outcome? My focus is, of course, on the public reporting, as someone who sits here as a representative of consumers and purchasers, and I went through this process -- this is my second time sitting on the Steering Committee, the Cardiovascular Steering Committee last year.

In the last week I have had the opportunity to see those measures used and sort of picked up in the press and used in reports in different contexts. So I do think that we shouldn't gloss over that point, that public reporting is very important.

So if it is a close call on the two measures, I would just really ask why folks are leaning toward, if folks are, toward 0047 versus 0036, since that seems like a very key difference between the two measures.

MEMBER LANG: I think, if you look at the -- Although the scores are kind of mixed, if you look at the validity issue, there were more individuals who voted low for validity on 0036 than 0047, although again you could point to another vote and say something else, but I think validity is a major issue.

The reason, I think, for the prolonged conversation was that there were some errors in how one of the metrics was described in terms of medications being listed there that didn't belong there, and that was cleared up. So I think that accounts for some of the scrutiny on 0047 as opposed to 0036.

Neither is ideal, but I think, of the two, I think 0047 is preferred, because it is closer to an appropriate asthma outcome that would track appropriate asthma care behavior.

MEMBER STEARNS: I think that my concern is -- and I don't know if there is any way to go back to the -- and I don't want to

complicate the process -- to see if -- Is there any opportunity?

DR. BURSTIN: None.

MEMBER STEARNS: Okay, fine. So improve 0036 now.

DR. BURSTIN: And that is actually the exact question I was going to raise, because I think a lot of the points raised yesterday about 0036 in some ways related to the lack of stratification of ICS versus the others, which is the hallmark of the PCPI measure.

Ben, I know you are on the phone, above us here from NCQA. I just wonder whether that is a possibility. Could we actually potentially see if they could kind of bring these measures closer together so that, in fact, the health plan data that is publicly reported is completely aligned with the measure that is used potentially for PQRS. Just a question for you, Ben.

MR. HAMLIN: Yes. We would

certainly be willing to consider it and take it back to our pulmonary panel.

MEMBER GLOMB: You know, I guess just from a scientific standpoint, I would rather have cleaner data that there is some access to than random public access to data that may not be as ultimately meaningful and may be misused, misinterpreted.

DR. BURSTIN: So I guess one question might be to PCPI, and I see them lined up behind me, is whether we may actually want to ask PCPI and NCQA to put their heads together perhaps and bring this back to you. At times, we have had developers truly combine their measures. It is not an easy or quick process, but it certainly does, we think, the public good.

Unfortunately, it could take a while, as we learned when combining the CDC and the American College of Surgeons surgical site infection measures. It took eight months, but I think at the day having one

national standard is preferable.

I don't know, Mark, if you would be willing to entertain talking to NCQA.

DR. ANTMAN: We are happy to work together.

DR. BURSTIN: Okay, good. So why don't we perhaps -- Are you okay with that, Reva?

DR. WINKLER: Oh, yes.

DR. BURSTIN: Great. Okay. Are there any other issues, as long as we have your brain power collectively here, that you think would potentially be important ones to consider, if there were some efforts to bring them closer together?

MEMBER HAECKER: I would just add that the combo issue is one that has to be addressed, and if you are going to combine the effort, because we are going to more combinations of long acting -- a lot of this with inhaled corticosteroids, and to exclude that would be a mistake; and given that the

National Heart, Blood, and Lung guidelines are going to be revised shortly, that will take years as well, obviously. I do think that would be an important combination.

I want to echo -- deviate a little bit from my partners here, that public reporting is very, very important for all of us, and the nuance of the combination drugs being missed in PQRS is a big problem for some of us. So to be dinged on a measure that would have us not using a drug that is very important, and many severe persistent asthmatics really need that drug, is a problem for me.

DR. BURSTIN: Is there a preference for where it lives, in the ICS versus the other strata? Sounds like that was an issue discussed yesterday as well.

MEMBER HAECKER: I am a generalist, not a specialist, obviously.

MEMBER LANG: What was the question?

DR. BURSTIN: My specific question was -- I think both of you raised this issue yesterday, of whether the combo with ICS is more appropriate under the ICS strata?

MEMBER LANG: Yes. I think the issue to track appropriate therapy, and particularly if you are targeting one matter to severe persistent asthmatics. There is no question that the combination of inhaled steroid long activated agonists is frequently what is prescribed and is consistent with evidence based therapy.

So combining inhaled steroid with the inhaled steroid combination, just tracking whether inhaled steroid is prescribed, I think, is really where to go.

Also, as long as you are entertaining suggestions, I would also focus on the one dispensing event and work on that, because one dispensing event in a period of time is a little short of optimal therapy. So you might want to focus on that as well in

terms of more regular exposure to the inhaled steroid or inhaled steroid combination, and eliminating some of the other agents that you are calling alternative medications, which are not -- for which there is not as much evidence supporting their use in patients, particularly with persistent, let alone moderate to persistent, asthma. And thank you for considering the suggestions.

DR. WINKLER: Any other thoughts on competing/related for asthma? I just want to mention that Mark has already told us that, in terms of the age range for measure 0047, that is already sort of in the works to align it with the NCQA measures, the five to 65. So I didn't bring it up, because it was already happening.

Anything else on asthma?

MEMBER LANG: Is there discussion regarding 1799 or 1800? I understand there are other issues coming up here, but I don't want asthma to monopolize the time we have for

harmonization.

DR. WINKLER: The question is what issues do you think are there, David? I think it would be important to at least put them on the table so we can figure out the best way to address them. 1799 and 1800 are the same developer. So there is sort of inherent harmonization in the description of the denominator

MEMBER LANG: Yes.

DR. WINKLER: And the measure focus is just different ways of looking at medication adherence. So I guess I would really like to know what your questions are about them.

MEMBER LANG: Right. Just to remind everyone, both of them had the same age range, five to 64. Both of them focus on persistent asthma. 1799 tracks the percentage who are on asthma controller therapy for at least 50 percent of the treatment period, and also 75 percent of the treatment period; and

1800 -- it is a ratio of controllers to total asthma medications of .5 or greater.

There are validity issues with regard to each in terms of the lack of evidence showing that 50 or 75 percent or, for the 1800, that a .5 ratio is associated with desirable or improved outcomes as opposed to a ratio of something less than .5 or, you know, an asthma controller medication for 40 percent of the treatment period.

So we don't have the precise data to support either of these, but again I think it is more consistent with the way we would conceptualize optimal asthma treatment.

DR. WINKLER: Other comments? Ben, This is Reva. To David's point, does NCQA actually look at your data such that patients -- you can look at the performance on measure 1799 or 1800 and then correlate with ED visits and hospitalizations and other potential outcomes to try and answer this question?

MR. HAMLIN: We can't through our normal HEDIS reporting process, because we only received aggregate data on an annual basis from the plans. When we generate large field testing databases is when we usually have access to member level data.

So we will either create another field test to test measure concepts or we will ask individual sites to run specific calculations or analyses for us and then provide us with the results, neither of which is easy to do nor cheap. So we try and get as much bang for the buck as we can out of our databases created for when we test and validate these processes.

DR. WINKLER: There seems to be great interest around the room for that sort of data to support the specifications for these new measures.

MEMBER GLOMB: These are the ones we really loved the spirit of what was being attempted, but just had to cringe at the fact

that there just wasn't a background to suggest these numbers would be ideal or optimal numbers.

DR. WINKLER: Let's go on to pneumonia. As I mentioned, pneumonia -- The measure is split into two discrete buckets. either you were admitted or you weren't. So the PCPI measures are the group that weren't. There are three measures from your assessment yesterday on vital signs, mental status, and then the empiric antibiotic therapy.

Measure 0147 is the hospital version of the initial antibiotic selection, but again these are only for inpatients, as are the mortality measures, whether inpatient or 30 days. Those target populations are those patients who are admitted.

So I guess, in terms of competing measures and related measures, the group from PCPI are related to the hospital measures, but they are not competing, because those are two separate target populations.

You raised this morning the issue around the mortality rates, inpatient versus 30-day. What do you all feel about those as competing measures, related measures, the utility of both? You had started having the conversation earlier. So now is the time to continue with it.

MEMBER RHEW: IN terms of the mortality rate, I would definitely say they are related. They are not competing. They are tied to the hip. You have to include the inpatient mortality and at the same time, then you have to talk about the 30-day.

Again to Don's point earlier, just talking about one area, you can really miss the boat in terms of what is happening in clinical practice. So clearly related, but I would not in any way say they are competing.

MEMBER YEALY: I would agree with Dave. If I had to pick one, I would pick 30-day, but I don't think we have to pick one.

MEMBER LEVY: Yes, I agree with

that. I think, if we just pick hospital, it could have unintended consequences of driving people out of the hospital to skilled nursing facilities to die there. So I don't think we are serving the field well by doing that.

On the other hand, I don't think we should ignore hospital mortality. So I don't see them as competing.

MEMBER ALMENOFF: I agree. They are not competing. They are related.

DR. WINKLER: All right. Given that you determined that they are related, Peter, you brought up earlier the issue of different risk models. So the question of harmonization now becomes important.

MEMBER ALMENOFF: No, I agree. I mean, models are different, and I guess we would have to really look at the details of what is in each model to know how different they will be. I mean, they might be much closer than we think. I just don't know the answer, because I don't have the models.

DR. WINKLER: Elizabeth, did you want to comment on that, because I know Patrick -- She and Patrick were talking about that earlier.

DR. DRYE: Yes. I can just highlight a couple of differences. I am looking at my screen at -- I guess I could email, and you could put it up. But the cohorts who we capture are really pretty close. AHRQ's is a bit more expansive, and they include some histoplasmosis and some other really rare and more regional pneumonias that we don't include, but they are a very small percentage of the cases.

So I think our cohorts are well aligned. They both include a vast majority of viral and bacterial pneumonia, and they have up to date codes.

Our models are different in that we use hierarchical modeling. Our is logistic progression, and I don't know if they are finding new names or not in it.

Then we looked at an episode of care. So we include transfers, as I mentioned before, and we attribute the outcome mortality to the first admitting acute care hospital. AHRQ, because the users can apply AHRQ however they want, I think our people could include or not include transfers, but I looked quickly at the specs before, and I think they essentially do transfers.

There are those differences, but at least on the cohort I would say we are really well aligned already.

MEMBER ALMENOFF: And you used three years of rolling data?

DR. DRYE: Yes. We have, in public reporting.

MEMBER ALMENOFF: I'm not sure how -- the other model, how much --

DR. DRYE: They work really different. We have to be looking at -- We have a risk-standardized model. So we have to take a whole national dataset, and then we can

give you your rate, but what AHRQ does is it uses -- and I hate to speak for Patrick; he left me an email, but I think there are probably other people here that can correct me if I am wrong.

They build a model. They estimate coefficients then in a nationally representative HPEP data, and then you can apply it locally within your hospital. That is an advantage of AHRQ.

So you can use -- They don't -- You can use whatever data you want to use to estimate your rate, and then they don't have the same -- I don't know how their uncertainty estimates work and their reliability and whether they have a minimum, but I don't think that they do.

MEMBER ALMENOFF: Do you both report out RSMRs?

DR. DRYE: Yes. Both the rates are with the regression standardized rates. Ours is just a two-level model to account for

clustering. So that the numbers are -- The models just work a little bit differently.

MEMBER ALMENOFF: Right.

MR. DREFFORD: This is Jeff Drefford from AHRQ, just to clarify that last point. The AHRQ model also includes the hierarchical. It just does it a little bit differently.

DR. DRYE: Oh, okay.

CO-CHAIR GROSSBART: Norm?

MEMBER EDELMAN: I think there is an important point here. As was point out, not only are they related; they are importantly related, and they may move together or reciprocally. So interpreting outcomes would be best if they can be interpreted together.

That suggests that, if possible, the developers should try to come up with a single model. That would be much more useful for the field.

CO-CHAIR GROSSBART: I would like

to echo what Norm just said. It drives us nuts in the hospital world if one model has two or three different diagnosis codes than the other one.

It just -- Usually tangential ICD-9 codes, it just -- you know. So standardizing the definition of the denominators, the included populations -- Obviously, when you are 30-day, you have got different considerations, episode of care versus encounter, things like that; but if there are areas, standardizing the models, standardizing the populations would really help hospitals improve, and that is really what the -- At the end of the day, that is what this is all about, and just the frustration and the noise because of the inconsistencies is a distraction for us.

MEMBER ALMENOFF: We do have about a five-year experience with this, and we originally reported out in-house mortality rates. We found that the in-house mortality

rate dramatically drops so quick, nobody can drop that fast.

So then we added a 30-day, and what we found was, as many people have just described, people were moved to the palliative care unit or they were transferred -- I mean, lots of things happened. So it is really kind of important to have both numbers.

A 30-day -- If your 30-day is high, you don't know if it happened in the hospital or you don't know if it happened on discharge. So, to me, it gives you lots of information and provided lots of information to know where their problems might be.

DR. WINKLER: I think I hear very clearly that you feel that they are both important. They are not competing. They actually work well together.

I guess I would ask the developers if there is an opportunity for you to work a little bit more closely together to further harmonization to try and address some of the

issues that Steve raises.

This is the feedback we get from the field, that it makes it very difficult to implement measures, and the reason we push so hard for the harmonization, the competing measures issues, is because the implementation is just very, very difficult when measures are just slightly different.

So we would really, really think that that could be extremely helpful, if there are opportunities for further harmonization with these two measures.

MEMBER ALMENOFF: Can I just bring up one last point? One of the other problems with the measures are the time limits. Everybody complains about that. So these are very good models. The data goes up, but they are old.

So we are talking about data that sometimes is two years old, and based on three years of data retrospective to that. So I think there is an issue about timeliness and

value. We are reporting this on a public website, but I am not sure how valuable it is to know what the death rate of a hospital was three years ago.

I think we need to know what it is. They are pushing us to transactional -- you know, yesterday, and I can't do that. But we at least need to do every six months or something a little better than what we are doing. So I think that is another consideration.

DR. DRYE: The challenge we have, as you know, is that outcomes measures are noisy. So we have to accumulate cases, and there are a lot of hospitals with relatively few cases, although pneumonia -- All hospitals face pneumonia.

So I think it is hard for us to -- There are tradeoffs between seeing differences among hospitals versus having sort of maybe just the year of data. I think people understand that is why we use three years. I

wish that tradeoff wasn't there, but I don't think you can get rid of it.

I would just say that, at least in the area of readmission, -- I think CMS has been on the line but having a hard time getting an open line -- they are looking at ways of getting more frequent information out to hospitals on their way.

MR. DREFFORD: Just from the AHRQ side on that question, I think there are methodological approaches that you can use to deal with the reliability issue and still get more current data. So we would be glad to talk with our CMS colleagues about that.

Similarly, I know AHRQ has several initiatives with their state partners with whom they work to get this data. I know they get quarterly feeds at the moment, and so they are looking to make that data publicly available and available to us as the developer in a similar time frame. So I think that will also help with the timeliness issue.

MEMBER STOCKWELL: I have a question for you guys at the NQF, actually. When the recommendations from the Committee comes to you with competing measures actually for somebody like NCQA and then the PCPI group to go back and attempt to create one uniform measure, are there any processes in place to help ensure that that actually happens?

Then the same basic question for the harmonization approach that we just talked about.

DR. WINKLER: Yes. Again, yes, this is something that we are ramping up very quickly now. Because all these measures are -- we check in with the developers every year on an annual update. We are kind of keeping track of these things so that those questions will be part of the responses expected as part of annual updates.

MEMBER STOCKWELL: Are the endorsements somehow made conditional? I mean, is that a consideration that you guys

have raised?

DR. WINKLER: Not really. What the endorsement is, is whatever the measure is, and that does not -- This is a rapidly evolving world, and so we are looking to try and push things along. But measure development does not happen overnight or over lunch.

So it is important to understand those realities, but gently push. It is the best way I would describe it.

CO-CHAIR GROSSBART: Reva, but in the case of competing measures, we do have -- I am assume we are not going to vote today. Maybe I am wrong, but if we don't vote today, we will have a conference call.

So, for example, the two asthma measures: If both developers have made no headway, we can de-endorse one by only selecting one. Correct ?

DR. WINKLER: Well, yeah. I think we have to walk ourselves through how exactly

that might work. Let's just say we would like to be optimistic that they could bring those measures together.

DR. BURSTIN: Although I do think it is fair to say that they will have discussions in this interim period and bring us back at least an initial assessment of what they think they can do. They can't complete the work, obviously, in a short time period, but we will at least get assurances, yes, we can walk down that path. If the answer is no, then I think we need to revisit your decision.

CO-CHAIR GROSSBART: Alternatively, to say we can't even schedule lunch. We do have to pick one.

DR. WINKLER: Yes. Steve, the opportunity, actually -- The work you have done -- you are acting as a proxy for our membership, and you are making decisions on their behalf.

Once we are done with today's work, we are going to compile a report that is

going to go out for comment. They are going to respond, critique, and give you feedback on how well you have made decisions on their behalf.

A very important conference call that we will have after that is to look at that and listen to it and evaluate it. This can certainly be a part of that follow-up, to see where we are in terms of the efforts toward bringing these measures together. At that point, you can make a different decision.

All right. So we talked about the pneumonia outcome. The other things on pneumonia is I just want to ask if Ben is still on -- I'm sorry, Mark. The 0096 and 0047 -- One is hospital, one is outpatient, empiric antibiotic therapy and initial antibiotic selection.

Mark, what efforts are made to keep those measures harmonized? You are both supposed to be aligned with the IDSA/ATS guidelines, but has there been efforts on your

behalf to be sure those measures are aligned?

DR. ANTMAN: Not having been closely associated with work on the pneumonia measures recently, I am not certain of this, but I believe that we have been in contact with the appropriate folks at CMS about their measure and about what may be aligned or misaligned with ours. We can certainly talk to them further, but I can't say for certain what discussions there have been before.

DR. WINKLER: There certainly is. They are related, and there is definitely a need for harmonization on the measure focus of the numerator aspect of the measures.

Also, just a harmonization opportunity or request in the title. Your measure is called Empiric Antibiotic, and the CMS measure is called Initial Antibiotic, and in our implementation comments feedback we get, using those two different words is misleading or hard for people to understand.

So there is just an opportunity to

harmonize just the title and the words you use, so that we know that we are measuring the same thing. So harmonization can sometimes be as little as that, and it has a huge impact in understandability out there in the field.

MEMBER RHEW: I would even echo that and go beyond the empiric antibiotic to say for all the PCPI's and all the ones beneath that that are ambulatory versus ED and inpatient, that there is somewhere in the title that it specified that. I mean, I could look at vital signs and say, oh, this is the ICU, or on the ambulatory side.

I would really encourage that we in the title make it clear, not only the area but the population, because I have seen others where they don't even mention the disease. So I would encourage that as a universal standard.

MEMBER EDELMAN: Another title issue, which is far more than the title issue, is whether or not they are truly talking about

bacterial pneumonia or pneumonia. That is a very substantive issue.

DR. WINKLER: Great. Is there anything else on this list of pneumonia measures that you would highlight as relating and competing? Okay. Then why don't we move on to COPD.

Again, I think it is very clear, because we started talking yesterday, there are the two spirometry evaluation measures. If you were looking at the table I gave you on asthma, turn it over, and COPD is on the back, for the same, side by side.

Also, I think this is an area where we have to look at the whole group of COPD measures, and this is, I think, where the age issue comes into play as the predominant you guys talked about.

So given that, do you want to look at the spirometry head to head initially as a first step?

MEMBER JEWELL: So the principal

differences between these two measures: Age is the biggest difference. As you can see, the PCPI measure has the broader age range, down to 18. We heard from the NCQA that they purposely start at 40, because of their concern about noise relative to appropriate classification below that age group.

It is also clear that the NCQA is looking specifically at initial diagnosis and trying to capture whether that diagnosis was verified with spirometry.

The PCPI measure -- We got some clarification that there would be some exclusion -- additional exclusion language, potentially, that would clarify that, if there was already spirometry on the books, this wasn't a monitoring measure.

So that helped considerably, but it also -- As I understood the discussion yesterday, the evolution of or the potential that somebody -- many people are walking around with "COPD" who have never had

spirometry, and that that is really the nuance difference, I would say, between the two.

The other area of difference is the level of analysis. NCQA covers more places. The PCPI measure is at the level of the individual clinician or practice, but still at the clinician level. And the public reporting issue that we discussed earlier -- it is broader for the NCQA at the moment.

So it is really -- I guess I would say it is really -- First and foremost, the question is how concerned are people that the lower age group clouds the picture?

MEMBER EDELMAN: Since we got the developers to agree, this really isn't a method to find COPD, but much more to find misdiagnosed COPD. I think it is an advantage to start at age 18. So if there are a lot of people running around at a younger age who don't have COPD and spirometry will clarify it, that is just fine.

MEMBER STEMPLE: I would agree.

You know, it is a confirmatory diagnosis, and it is a big Medicare Stars outcome, and I think taking it down if somebody at 18 has that diagnosis, I would think you would want to confirm it. So I would appreciate having the lower age range, because if we are concerned about the appropriate diagnosis, it would make sense, if they are going to have the diagnosis to do the spirometry for validation. So I think the lower age range is totally appropriate, because this is a confirmatory, not a treatment, etcetera.

DR. BURSTIN: And there are differences in the reliability and validity scores that you gave to each of those measures.

MEMBER LEVY: We are arguing on behalf -- in favor of including the 18-year-olds, but the reliability and validity was rated lower, it looks, than the other measure, because it is cleaner. Yes, I think that is right.

MEMBER EDELMAN: It kind of depends on what you are trying to do, whether you are trying to find COPD or find misdiagnosed COPD. It is different.

MEMBER LEVY: Yes, I appreciate your point. If someone is carrying a diagnosis of COPD at the age of 30, they should have spirometry. Yes.

MEMBER CANTINE: Well, I have to say, as somebody who is in the lab day in and day out, I see frequently individuals between 18 and 40 where I am looking at results with -- you have air trapping, you have -- you know, you can see that dramatically, and these individuals don't have that diagnosis yet.

I am not a physician. I can't make a diagnosis, but I know it when I see it, and I think it is important to include this group.

DR. BURSTIN: Ben, are you still on the line?

MR. HAMLIN: Yes, I am.

DR. BURSTIN: Any thoughts about this discussion about potentially the need for greater precision of the COPD diagnosis in those younger? Have you considered at all having perhaps another strata that goes 18 to forty?

MR. HAMLIN: You know, again, we would consider it, but again our concerns are really in the higher false positive rate in the 18 to 40 group for spirometry confirmation diagnosis, and the fact that the confounding effects of asthma in the younger population make it really hard for us to get to a comfortable level. But we can certainly take it back for additional consideration and look at the data again, if possible.

DR. WINKLER: Thoughts from the Committee?

CO-CHAIR GROSSBART: One question: Is this another opportunity for NCQA and AMA to get together and to create a single, more robust measure that addresses all these

concerns? They already got the reservation for lunch.

DR. BURSTIN: All right. Another lunch engagement for PCPI and NCQA here to see if they can bring these together.

DR. WINKLER: A long lunch.

DR. BURSTIN: A lot of good in both of these, but I don't think it makes sense to have competing measures that are not harmonized. This doesn't help the broader universe.

MEMBER JEWELL: And if I understood the measure developer's comment a moment ago, if there are data to evaluate the false positive rates, that would help. Right? As a place to start, if that was a concern.

DR. BURSTIN: I think he was saying, and I was hoping some of the pulmonologists would pipe in -- I think what Ben was actually saying, this is a higher false positive rate on spirometry for the younger population. So it is actually about

the test itself. Is there a higher false positive rate, people being misdiagnosed?

MEMBER ALMENOFF: But the test is not making a diagnosis of COPD. It is making a diagnosis of obstruction. That is all it is doing, and then with the other pieces you make a diagnosis of COPD. So it is just showing obstructive versus nonobstructive disease.

So having a spirometry that shows -- We'll just say it is an obstructive lung disease. We are not going to say it is COPD. So having a spirometry that is positive in a 25-year-old or a 22-year-old can mean he might have asthma or something else, but you are picking up disease. So I am not sure about the false positives.

MR. HAMLIN: Right. It is the combination, though, of the -- You know, we have a couple of articles that had questioned the false positive screening for identification of COPD or the ability of spirometry to predict COPD in the younger

patient, along with the issue of asthma versus COPD in the younger population.

So it is the confounding factors altogether are why we decided to go with a higher age range.

MEMBER EDELMAN: But the point, though, is spirometry, simple spirometry especially, that is not done post-bronchodilator does not make a diagnosis of COPD.

MEMBER COHEN: From a practical point of view, speaking from a clinical point down in the trenches, we just see so many patients who have been diagnosed with COPD or asthma who have never had a pulmonary function test, and I think that is what is really important here.

To me, the age is not such a big issue personally. It is just, you know, they go to the doctor. They have smoked before or they are short of breath. They get an inhaler; you have asthma, you have COPD, and

we see them in the office and we see them in the hospital. They have never had a pulmonary function test, never.

If you had disease, if you had heart disease, you get an EKG. If you have anemia, you have a hemoglobin level. You get a hemoglobin A1c if you have diabetes. But if you have lung disease, you don't get a PFT. From a practical point of view, that is why you need to do spirometry on patients who supposedly have lung disease.

MR. HAMLIN: I think require us to expand the denominator, because we would then be creating prospectively an obstructive airway disease measure as opposed to just a COPD measure.

DR. WINKLER: Ben, can you repeat that?

MR. HAMLIN: I said that -- I mean, I actually agree with the comment. However, the denominator is limited to COPD at this point. So from an admin data

perspective, you know, those are the decisions that were made.

We would certainly be willing to take these comments back to our pulmonary panel for consideration, but also the fact that, if you do feel that there is a recommendation that we should look at an obstructive disease, you know, spirometry for obstructive disease, COPD and asthma, in the younger population, we certainly make that one as well, if I understood that comment.

DR. WINKLER: There are nods. Some are nodding. You might want to speak up and let Ben know that, yes, you would like that.

MEMBER HAECKER: I would love it. Thank you.

DR. WINKLER: Okay. So we have left with the two spirometry measures, NCQA and PCPI having another lunch, but we will want to hear what exactly the plan is when we do the post-comment call in terms of what your

intentions are going forward, given the feedback that you have gotten in terms of bringing these two measures together.

In terms of COPD, we have got six measures of COPD. One of the issues that came up a lot was age, and as it turns out, the age range for all the measures except 0091 is 18 and above.

So despite the fact there seem to be a big problem as we went through them, as it turns out, we have just kind of discussed it as the one problem on the age. They are aligned at least on that factor.

From your discussions on the other measures for the bronchodilator or the management of poorly controlled COPD, I know this is a stretch, but I did not hear you all raise particular harmonization kinds of questions.

Is the focus of those two measures somewhat different and, therefore, the denominator populations are necessarily

perhaps different? Norman, I know you did 1825.

MEMBER EDELMAN: Yes. I mean, one is poorly controlled, and the other is everybody, and one is long acting bronchodilator, and the other is bronchodilator. So they are different in the numerator and the denominator.

The one that I think is relevant to care is poorly controlled. The other one, I can't find anything wrong with, but I am not very excited about how it is going to help us take care of patients.

DR. WINKLER: Okay. So any other input into that in terms of other harmonization issues or are the measures sufficiently different that we can live with where we are right now, since the ages are harmonized with the one exception?

Okay. We just finished the discussion of competing and related measures.

CO-CHAIR GROSSBART: And we picked

up another five minutes on the agenda. We are only 10 minutes behind. We are 10 minutes behind schedule. We had a break scheduled for 10:45. So we are going to take a -- We are scheduled for a 15-minute break. Do we want to drop that down to 10 so we can stay on track for airports? So 10 minutes. So at 11:05 we will reconvene here.

DR. WINKLER: For the folks on the phone, we are just taking a 10-minute break, and we will resume at 11:05 to begin the agenda at the eleven o'clock spot. So I know we do have measure developers calling in. So we will be with you shortly.

(Whereupon, the above-entitled matter went off the record at 10:55 a.m. and resumed at 11:07 a.m.)

CO-CHAIR GROSSBART: Well, so just to tee this up, we have CDC calling.

DR. WINKLER: Do we have anybody from CDC on the line?

MEMBER LEVY: I can do this. If

Shelley is not on the line, I could talk about it.

DR. WINKLER: Okay.

DR. MAGILL: And I am on the line.

MEMBER LEVY: All right, Shelley. It's Mitchell, but as long as you are on, you should do it. You could take the heat.

DR. MAGILL: Oh, I'm sorry. Hi, Mitchell.

MEMBER LEVY: Reva had just told me I shouldn't vote. I told her I would make noise, but I wouldn't vote.

DR. MAGILL: Fair enough.

CO-CHAIR GROSSBART: Okay. If someone would grab his counter, just to make sure. Seriously now, we have a representative from CDC, and what about the other measures with CMS and VPS? Do we have those developers on? Okay, we will do one at a time. So CDC, we are asking for about a two-minute update/ overview/introduction of the measure, and then we will take it into the Committee. So who do

we have from CDC?

DR. MAGILL: This is Shelley Magill from the Division of Healthcare Quality Promotion.

CO-CHAIR GROSSBART: All right, Shelley. Steve Grossbart here, Co-Chair of the NQF Pulmonary Critical Care Committee. Two minutes. Please introduce the measure.

DR. MAGILL: Sure. This is the Ventilator-Associated Events Outcome Measure, and this measure is the result of work done over the past few months by a working group consisting of representatives from several key societies, including a number of critical care societies, and is intended to replace the current ventilator-associated pneumonia surveillance definition that exists currently in the National Healthcare Safety Network.

The focus for the ventilator-associated event algorithm has really been on utilizing objective, streamlined criteria, criteria that can be assessed across the

spectrum of mechanically ventilated patients, because as most of you know, the current definitions are limited by their subjectivity and the possibility for manipulation.

So our focus is really on enhancing reliability while maintaining clinical credibility, and the new algorithms intended to capture a broad range of conditions or complications that could occur in patients on mechanical ventilation. It is not really specifically a ventilator-associated pneumonia definition, although it will capture some patients with VAP.

The definitions that you will have a chance to review really have as a foundation a period of worsening oxygenation following a period of stability or improvement on the ventilator, and following meeting that particular criterion, if patients meet some additional criteria related to signs of infection or inflammation, these are objectively defined and include changes in

white count, temperature, and antibiotic starts. Then the patient would be defined as having infection-related ventilator-associated complications.

If there are any further details that anyone would like me to speak to, I can certainly do that as well.

CO-CHAIR GROSSBART: I will ask if there are any questions for the developer from the Committee? If not, then Peter, this is your measure. Can you update us on the Work Group's analysis?

MEMBER ALMENOFF: Sure. It has a very nice synopsis. This is one of the HAI initiatives where sort of you develop an infection while in the hospital, which is probably inappropriate. So the idea is to try to prevent things, things like central line infections, MRSA infections, was VAP and now they are sort of redoing VAP, CA-UTI which is urinary tract infections by catheter, I think that we should be able to prevent if we do

certain bundles of care.

VAP has been an issue for a couple of years, even though within our system we actually still do it, and reported. There have been a lot of issues about the validity of the diagnosis. If you sort of do sampling for the way people make these diagnoses, it is very variable. So there has been a lot of standardization issues regarding VAP.

I want to applaud the group for really trying to tackle this, because this is a very difficult area to tackle, and it has already a very interesting history; because people are just not doing this measure, for the fact that it is very hard to reproduce, and there is too much variability throughout the system.

So with that, I will sort of go through some of the data points One interesting thing that I don't think they mentioned on the call is that in the past we have been reporting out ventilator-associated

pneumonia rates in rates per thousand. In this new concept, they are now doing SIRs, which are standardized incident rates.

So not only is it a new measure, but it is a new way of looking at infection rates, which is very similar to what we do for standardized mortality. So it is something that has been established in other areas, but yet this is kind of a new concept, and we will probably go to that for CLAB and for other things in the future. So just to note there is some other additional changes.

Regarding the impact on the system, it is -- At least on the inpatient side, it is a very deadly thing to get. I think, at least in their write-up, probably more than 50,000 cases a year, but the mortality rate can be 50-60 percent.

So in an ICU setting this is something that you definitely don't want, and the mortality rate is very high. So there is a lot of issues regarding that piece.

When you look at the performance gap, one of the issues is this is a brand new metric with new definitions. I know she mentions it has just gone through the societies for a couple of months, but we really don't have any data.

The problem with VAP is that VAP has been a very inconsistent diagnosis. So what they are trying to do is sort of extrapolate some of the VAP information into the new measure. The problem is that the VAP measure was probably not that accurate, to start with.

So we do have some issues regarding that. So, to me, the impact on the system is extremely important, especially for the ICU setting, and they also have sort of chronic rehab facilities that have ventilators. So those are all sort of appropriate settings. We are not doing this on the outpatient side, which was my big issue yesterday.

So we do know the designated sites. The impact is high. The performance gap: As I mentioned, they even write in their statement they really don't have any performance gap data, because it is a new measure, but were trying to extrapolate a little from the VAP data. So there is really very little there.

Should I stop there until we go into the next one?

CO-CHAIR GROSSBART: Yes. Let's move to the voting, since you just covered the first set of -- the first voting block. I guess the first thing I would like to do is ask if the Work Group has comments that they would like to contribute; and if not, the Work Group. Does anyone from the Work Group have a comment, first of all? Mitchell?

MEMBER LEVY: Shelley, can you clarify what you are going to do. Is CDC going to give this to NHSN to start collecting data on these metrics?

DR. MAGILL: Yes. We are moving ahead with all the various steps that need to happen to implement this for use in NHSN, and our anticipated start date is January 2013.

MEMBER LEVY: I think that is important for this group to understand that. I think Peter summarized the struggle with VAP that has gone on with as long as people have been using the term, which is it is an invalid definition, which is why CMS dropped it as a performance metric, and it was HHS and Don Wright who motivated this consensus group to try to get a surveillance definition.

I think the point that is important for us to recognize is this is moving forward, whether or not we recommend it. I am not saying that should make us recommend it, but it is going to appear as the surveillance definition for NHSN and, therefore, public reporting, without or without our support. I don't think that should color us, but I don't think we should -

- I think we should be clear about that.

MEMBER ALMENOFF: I think one of my concerns -- I know this is going to go through the network -- is VAP already has a tainted past, and this has not really been validated yet. IF this winds up being like VAP was, we are going to never be able to do a measure on this at all.

So at least part of our discussion was this is important. It needs to go forward, but is this really too early yet, because we really don't have any data. So we are sort of wanting to approve things on things that don't have any data yet, which may be a little concerning. But I agree with you, it is an important area, and I think the definitions are interesting, but I want to see then some kind of validated process.

CO-CHAIR GROSSBART: And I would just like to remind the Committee that we did reject some measures for lack of evidence yesterday, and we need to be internally

consistent. Norm?

DR. EDELMAN: To the same point, just so I understand, so there is no retrospective look? thee is no attempt to validate this metric. We don't know how it performs relative to VAP in a retrospective look. We know nothing about this?

MEMBER LEVY: Shelley, you piloted this in some hospitals. Didn't CDC do that?

DR. MAGILL: Yes. What I will say is that some of this work -- a lot of this work is based on the work that Michael Klompas and the CDC Prevention Epicenters has done.

So there are published data on variations of what this algorithm is that you have been presented with. So Mike has papers in clinical infectious diseases and in the PLoS ONE Journal on looking at similar definitions.

There are differences. His definitions to date and his investigations to date have not included antimicrobial use

requirement that we have in the algorithm, but the other components are very, very similar, particularly the period of worsening oxygenation after a period of stability improvement, which is kind of the foundation of the algorithm.

Our published data -- and we have done a small, kind of a pilot study within the last year or two looking at, again, a variation of this definition algorithm compared with the current VAP definition, and looking at basically how the event determinations compare between the two.

One is not a subset of the other. If you look at the work that Mike has done, rates of VAEs will be quite a bit higher than rates of VAP, which is perhaps not surprising, and he has also looked at outcomes.

So we found that patients with events detected by the new or similar definition algorithm to VAE do tend to have longer length of stay, even higher mortality

than patients who do not meet the definition.

So there is some work out there, albeit not with the identical definition algorithm.

CO-CHAIR GROSSBART: Dianne, a question?

MEMBER JEWELL: I am not familiar with the literature that was just referenced, but I guess the question is: Is what is out there relative to the potential performance of this as a quality discriminator, because that is the question of reliability and validity that we are after, more precisely? Right? Okay.

MEMBER STOCKWELL: Would it be possible to show what the actual numerator and denominator are, because I think it is important for everyone to get a real sense of what we are talking about. This is vastly different than the previous definition of ventilator-associated pneumonia.

So I think, even though it

probably will be voted on at a later time, I think it will color the conversation that we are having, if people are seeing what we are talking about.

DR. MAGILL: This is Shelley. Would you like me to comment on that?

CO-CHAIR GROSSBART: Yes, please.

DR. MAGILL: Okay. For VAE, the numerator -- you know, if you are talking about rates, the numerator is the ventilator-associated event, just as currently the numerator is the VAP event, and the denominator is identical. It is ventilator days, determined in the exact same way as it is determined now.

The standardized incidence ratio is the same thing as the standardized infection ratio, which is where NHSN has gone in terms of presenting this rate information in publicly available reports.

So the SIR is no different than what is being used currently for reporting

HAIs from NHSN. That is an identical measure.

CO-CHAIR GROSSBART: Shelley, this may be a stupid question, but how do you calculate an expected rate without a database?

DR. MAGILL: Right. So that is where the issue of implementation and experience comes in. So, obviously,. thee is going to have to be a baseline period where these data are being reported to the system, and we would anticipate -- If we do succeed in implementing this in January 2013, probably we are talking about the first couple of years of reporting.

So, yes, it is true, we do need to have events reported to the system for a period of time in order to have that baseline, but that would be true no matter what definition we were to move to in the system.

CO-CHAIR GROSSBART: I am going to recommend that we go through our voting process. So the first question was impact. Peter has already given us an update. I

guess, real quickly, any other outstanding questions before we vote?

DR. BURSTIN: Just one quick comment. I just want us to again separate. A lot of the discussion we had was about reliability and validity, which is the second criterion. So this is really just about those first three subcriteria, just to keep it clean.

CO-CHAIR GROSSBART: Is this an important clinical condition to be looking at?

DR. BURSTIN: Yes.

CO-CHAIR GROSSBART: In terms of impact, 14 voted High; three, Moderate; and one, Insufficient.

The next question is performance gap. This is, again, a four-number, high to low on a scale of one to four. Any questions? We don't have any evidence on the performance gap.

MEMBER ALMENOFF: We don't know what performance is. The write-up says there

is no performance gap, but --

DR. WINKLER: This is a new measure, and so we don't necessarily require data from that specific measure, that you can use other surveillance data or whatever else might be around to help support the argument that there is a performance issue in this topic area.

DR. BURSTIN: Especially, evidence citations of a performance gap are acceptable as well for a new measure.

CO-CHAIR GROSSBART: So is there a serious gap in terms of ventilator-associated events, complications? Okay. That is very different. Thank you. Again, on a scale of one to four.

We have five Highs; six Moderates; and seven Insufficient Evidence. So it passes.

Then the final area in this section of the voting is the evidence, and again clarification. Evidence?

DR. WINKLER: This is an outcome measure. So as long as there is a rationale for process of care that can impact that rationale, that is the kind of evidence. It is not the same detail and quality/quantity consistency as you see in process measures.

CO-CHAIR GROSSBART: So these events are reducible through process changes, is the question. Thank you. This is a Yes or No question, one Yes, two No, and three Insufficient.

So the evidence vote was 13 Yes; one No; four Insufficient Evidence.

Now we move on to reliability and validity. We have already had some conversation around that. Let's go to reliability. Peter, any additional comments?

MEMBER ALMENOFF: Let's see. Some of the people on the group -- I guess in the write-up they felt it was sort of a valid method of data collection, but it did not really discuss the data reviewed or to suggest

levels of validity. I think that is about all I was going to say.

DR. WINKLER: And the fundamental question here is has this measure been tested for reliability and validity, and where the results demonstrate reliable and valid measuring?

MEMBER STOCKWELL: At least my impression of that is the answer is no, that this is a brand new measure, that there are some corollary outcome measures that have been looked at, but as this is defined, there is zero experience with it whatsoever.

CO-CHAIR GROSSBART: Any other questions, comments by the Committee?

DR. BURSTIN: From CDC, any comments? Anything you can add there? Or the timing of when you might have data?

DR. MAGILL: Sure. Again, I think, we do have these investigations of very, very similar definitions that, I think, are useful and helpful and can be extrapolated

to what we expect with the definition that has been proposed, but there is a lot of work that is going on right now. I think the hope would be, in the next year or two, we would have additional evidence in these areas.

CO-CHAIR GROSSBART: Okay. Dianne?

MEMBER JEWELL: Just in reference to the earlier disclosure that this would be moving on, irrespective of our decision, sort of to balance out the perspective that whatever we decide, it is clear that that is an independent decision, and not that that was in doubt, but I just think it is important to be able to say for the purposes of the transcript that, while things might in use out there, the NQF and its member groups are deciding whether or not to put the NQF seal of approval, if you will, on something as a quality measure -- as a quality measure.

So I just wanted to be clear that that is what we are doing.

MEMBER ALMENOFF: Yes, but when they say it is going out, it means it is going out to the network. It is not being endorsed by CMS and saying the country is doing it. It is a very select network that runs this kind of data analysis.

MEMBER JEWELL: Yes, not everybody is going to know what that is out in the -- who reads the transcript. So that helps.

MEMBER ALMENOFF: Right.

DR. BURSTIN: And since the issue has been brought up in the past, those measures have occasionally come -- have come forward to NQF untested for time limited endorsement.

We don't allow that for what we consider complex measures like composites or outcomes.

CO-CHAIR GROSSBART: Okay. Any further questions from the Committee? So this is a one through four, one being High, two being Moderate, three being Low, and four

being Insufficient. Let's vote.

There is zero votes for High, zero votes for Moderate, two votes and Low, and 16 for Insufficient. I believe we are done, and we are not able to endorse the measure at this time or go through the process to endorse.

I do want to emphasize to the measure developer that there is a lot of enthusiasm about this measure, and we are looking forward to the opportunity to evaluate it with a little bit stronger evidence base.

DR. MAGILL: Thank you very much. We appreciate your consideration of it.

MEMBER ALMENOFF: I think it is really important that you get a resubmission, and maybe in the next six months or a year when you have a little data available, because I think this is something very important that we really need to put forward. But with our experience with VAP, and we thought that was a good validated measure, and that wound up being a disaster, I want this to have some

good validation to it. Then I think it would be important to push through the system.

DR. MAGILL: Thank you. I hope we will have some additional information in the relatively near future.

MEMBER ALMENOFF: Thank you.

CO-CHAIR GROSSBART: Next on our agenda is measure 0356, which is blood cultures performed within 24 hours prior to or 24 hours after hospital arrival for patients who were transferred or admitted to the ICU within 24 hours of hospital arrival, which also, I believe, is the longest measure name in the NQF list.

DR. BRATZLER: We did it that way, Steve.

CO-CHAIR GROSSBART: Dale, welcome back. Can you give us a few minutes update -- introduction to this measure?

DR. BRATZLER: Yes. This should be relatively short. This particular performance measure -- In the United States in

the CMS database for pneumonia, we see about -- I am going to say about 800,000 cases per year that come into the clinical warehouse that are collected by hospitals, and somewhere between 10 to 15 percent of those patients go to the intensive care unit.

This particular performance measure simply says, if the patient is admitted to the intensive care unit within 24 hours of hospital arrival, is a blood culture performed within 24 hours of arrival?

Also, the denominator is limited to those patients whose admission to the intensive care unit is because of pneumonia. In other words, once in a while, not uncommonly, we will see patients that come into an emergency room with pneumonia but have some other unrelated reason to be placed in a monitored or an ICU bed.

So perhaps the patient has an arrhythmia or a GI bleed or something else that is unrelated to the pneumonia. So the

denominator is restricted to those patients who go to the ICU because of pneumonia, and the numerator is did they have a blood culture performed.

We had done studies in the past, very large studies, that demonstrated that sicker patients -- the yield that the -- the true positive yield of blood cultures is substantially higher in patients who are sick, and the IDSA and ATS specifically do recommend that a blood culture be obtained in the patient that is admitted to the intensive care unit for pneumonia.

CO-CHAIR GROSSBART: Thank you, Dale. Are there questions for the developer from the Committee?

MEMBER RHEW: Hi, Dale. This is Dave Rhew. At the risk of adding extra length to the title -- we talked about this earlier -- could we just add "pneumonia patients" instead of just "patients" to the title? Thanks.

DR. BRATZLER: Yes, I don't think that would be a problem.

CO-CHAIR GROSSBART: Any other questions for the developer from the Committee? With that, Mitchell, I believe you are up to walk us through this measure.

MEMBER LEVY: You just heard the description by CMS. It is reported already on Compare and has been reported for a while. It is specific to ICU patients who are admitted from the emergency department within 24 hours with pneumonia, and there is no qualification on pneumonia. It is a large percentage of the ICU population, as you also heard.

The evidence comes from a number of sources. There is one RCT that was published in, I think, 2005. There is a couple of systematic reviews and a number of the guidelines from -- one guideline from IDSA ATS, and another one on the sepsis guidelines, which is about 25 societies.

All recommend this -- at least

recommend blood cultures before antibiotics. So this is a metric that reflects the scientific evidence, and also the opinion in the field.

CO-CHAIR GROSSBART: Thank you. Any other comments from the Work Group? Any questions to the Work Group from the Committee, full Committee?

MEMBER YEALY: I have no question, but as the strongest opponent of the blood culture version yesterday, this is much more targeted in a much more high yield population, and I don't have anywhere near the same concerns.

This actually allows any ICU transfer. Doesn't have to be directly from the emergency department. What it excludes is transfers from other institutions, but this one makes a lot more sense to me, in that it is targeted, focused, and verbose at the same time.

CO-CHAIR GROSSBART: With that,

let's move into our voting process. So the first thing is the impact of the measure. Mitchell, any additional comments?

MEMBER LEVY: No.

CO-CHAIR GROSSBART: Any question or comment from the Work Group or the Committee? With that, let's vote, a one to four scale again.

The vote is 16 with a rating of High and three with a rating of Moderate.

Let's move on to the performance gap.

MEMBER LEVY: This is, I think, where probably the biggest question is with this metric. The report is 96.4 percent or so on Compare. So there is a question of what the bang for the buck is here in terms of the performance gap, although it does in the submission look like there is a certain percentage of hospitals that are under 80 percent, but that is just mentioned, and I don't see it anywhere else. So that is the

biggest question about this metric.

CO-CHAIR GROSSBART: Can someone jog my memory? Is this part of the Value Based Purchasing program?

DR. BRATZLER: Yes, it is. It is, I believe. I need to double check, but I am pretty sure it is one of the VBP measures.

CO-CHAIR GROSSBART: I thought so. and I asked that, because CMS has done a pretty good job of removing from their program topped out measures, and through their analysis found that the gap was sufficient to justify rating hospitals. So I just wanted to bring that tidbit to the Committee's attention.

Dianne, a question?

MEMBER JEWELL: Actually, just on page 22 of the application, they reference that performance rates of 18 percent of hospitals, nearly one out of five are still below 90 percent.

CO-CHAIR GROSSBART: Thank you.

Any additional questions or comments?

MEMBER RHEW: Just a question from a historical perspective. I know with beta blockers, they were removed at one point. Is there a certain threshold that we have seen in the past where traditionally we have thought that the threshold -- or the performance gap was too small? Maybe you can just give some historical perspective on that.

DR. WINKLER: Yes. You did it yesterday with the asthma measures. That is sort of the whole discussion around reserve status. Have they really reached the limit of opportunity for improvement, but it is not as if there is a numerical number, because it often has to do with the target population at risk.

Then the other issue around it is maybe a national average won't show it, but are there disparities questions? So when you look at subpopulation analyses, perhaps that is where you find your disparities. Are you

getting data from a large number of entities?

For instance, we have seen measures that come in from a single state, now done very well in one state, but you have no idea what is going on in the other 49 states. So perhaps it is great in one place, but you don't know enough about everybody else.

So these are the issues you have to weigh. That is why there is not an absolute cutoff on it.

DR. BURSTIN: Dale, I know you have done a fair amount of work looking at sort of some formulaic ways of trying to assess measures to be retired. I forget what it is -- 75th percentile, X number of years or something? Is that something you could share with the group in terms of whether your assessment would be that this measure meets that topped out?

DR. BRATZLER: I was afraid you were going to ask that, Helen.

DR. BURSTIN: Sorry. You have

taught me this before.

DR. BRATZLER: Yes. So this is a formulaic way that CMS looks at these individual measures to determine if there is no statistically significant difference between the 75th and the 90th or 95th -- I don't have the methodology laid out in front of me, whether there is no statistically significant difference in those performance rates.

So they do periodically look at the individual measures to see whether or not they are topped out, and currently this measure had not been topped out yet, but that is -- As Steve was pointing out, they do look at these measures periodically, and through the rulemaking process. That is the only way they can put in measures or take measures out of the Value Based Purchasing Program.

CO-CHAIR GROSSBART: And that is in the Inpatient Perspective Payment System rule, if the staff could quickly print us up a

couple of copies.

DR. BRATZLER: All 1,000 pages. I am sure they would like that.

CO-CHAIR GROSSBART: But seriously -- The CMS methodology for topped out measures might be something the NQF wants to look at in terms of making a decision to move measures to reserve. I didn't think of that yesterday.

DR. WINKLER: Actually, in the cardiovascular project we did that, and if CMS determines a measure to be topped out, they don't include it in Value Based Purchasing, because the math doesn't work.

CO-CHAIR GROSSBART: Right. Mitch.

MEMBER LEVY: We have a small database with this Value Based Purchasing campaign of 30,000 patients, and that percentage is running about 60 percent blood cultures before antibiotics for patients with severe sepsis and septic shock who are admitted to the intensive care unit, and that

is within six hours. I'm sorry?

DR. BURSTIN: Why so much lower than the CMS number?

MEMBER LEVY: I don't know.

DR. BRATZLER: Well, I am betting -- because we are actually in the process right now to, unrelated to pneumonia, building sepsis performance measures for -- through another contract for CMS, and it just hasn't been focused on. It hasn't -- and there are some challenges around identifying those patients.

Pneumonia -- hospitals now have been working on improving quality of care around pneumonia for many years. So defining the denominator for hospitals is pretty easy now.

CO-CHAIR GROSSBART: All right. Yes?

MEMBER PELLICONE: Point of clarification. This measure is looking for the blood cultures performed within 24 hours,

irrespective of the timing regarding antibiotics?

DR. BRATZLER: That is correct, because many of these patients have already received the first dose of antibiotics in the emergency department. So this is irrespective of antibiotic timing.

CO-CHAIR GROSSBART: All right. If there are no further questions, let's move to voting on the performance gap.

The scoring was eight votes for High, 10 votes for Moderate, one for Low.

Now for the evidence, quality of the evidence. Mitchell?

MEMBER LEVY: I don't have, really, anything to add to what we have already discussed.

CO-CHAIR GROSSBART: Any questions for Mitchell or the Work Group? If not, let's move on to the voting, and this again is a Yes/No, one/two or three for Insufficient.

The results are 18 Yes and one No.

Now we move into our reliability and validity section. Mitchell, any comments about reliability?

MEMBER LEVY: Not really. I think John just brought it out that the real goal of this is to get blood cultures before antibiotics, which actually made me realize why our number is probably lower, because it is automatically calculated, because people enter what time patients get the antibiotics and what time they get the blood culture. So they don't really self-report.

So I think, from that point of view, the reliability and validity is not what we would like to see ideally for a metric, but I don't think it challenges the metric that much.

CO-CHAIR GROSSBART: Any questions or comments?

DR. BRATZLER: Yes. I would highlight that this measure includes those patients who get admitted to the floor and

then subsequently transferred to the intensive care unit. So that is why we don't look at timing of the antibiotic. It is not limited to those patients that go from the emergency department to the ICU.

MEMBER YEALY: One quick question. If you had blood cultures drawn 48 hours before going to the intensive care unit, you are admitted to the floor, and you have adequate biologic sampling and then deteriorate, how is that handled for this metric? There really wouldn't be a whole lot of reason to redraw them again only to meet the criteria, but it looks like it says only 24 in either direction.

DR. BRATZLER: That is correct. It only looks at those patients admitted to the ICU within 24 hours of arrival.

MEMBER YEALY: So if they had been drawn 48 hours earlier, were already positive at the time they entered the ICU, to meet the metric you would have to draw them again. Is

that correct?

DR. BRATZLER: I guess, but I think that would come up very uncommonly that blood cultures would have been drawn pre-admission, because we are only looking at a window from 24 hours after hospital arrival.

MEMBER EDELMAN: It is not hospital arrival. It is intensive care unit arrival. The point is there are patients up on the floors of the hospital who have blood cultures and then get sicker and are transferred, and the sicker may not be that they think there is a new organism. It may be something else.

DR. BRATZLER: Right, but gain, the denominator for this patient -- for this measure only includes those patients that are admitted to the intensive care unit within 24 hours of hospital arrival.

MEMBER YEALY: Okay. That is the clarification. Thanks. Then the blood culture has to have been within 24 hours

either side. Okay, thanks.

CO-CHAIR GROSSBART: Any other questions? Well, then let's vote on the reliability question, one to four scale.

The reliability came with a vote of 15 High, four Moderate, no other votes.

Then the validity. Mitchell, anything?

MEMBER LEVY: Really, nothing.

CO-CHAIR GROSSBART: Any questions from the Committee regarding validity? Hearing none, let's move to a vote, again a one to four scale.

The votes are 17 High, one Moderate, and one Insufficient Evidence.

Now we move on to the usability and feasibility question. So usability.

MEMBER LEVY: Really, nothing to comment. It has commonly collected and reported in Compare. So I think the committee thought it was a high factor for usability, high support.

CO-CHAIR GROSSBART: Any comments or questions from the Committee? All right, let's move to voting, again a one to four scale.

The results are 16 High, and three Moderate. No other votes.

Then feasibility.

MEMBER LEVY: I don't have anything to add.

CO-CHAIR GROSSBART: Are there any questions or comments by the Committee? Let's move on to voting then, again a one to four scale.

On feasibility, 16 votes High, three votes Moderate, no other votes cast.

Now we come into our overall suitability for endorsement. This is a Yes/No question, one Yes, two No.

It was unanimous endorsement, 19 votes in favor. Thank you, Dale.

DR. BRATZLER: Thank you.

CO-CHAIR GROSSBART: And our next

measure is 0334, PICU Severity-adjusted Length of Stay.

DR. WINKLER: Do we have developers from VPS on the line?

DR. SCANLON: Hi, yes. This is Matt Scanlon. I duly identified myself as being with the Medical College of Wisconsin for the purpose of the call, and I have Chris Gall who is my VPS counterpart.

CO-CHAIR GROSSBART: Reva, just a point of order. Should we have them introduce all of their measures right now? Matt and Chris, what we are going to ask you to do is invest about three minutes of your time in just giving an overall summary and introduction to the -- what is it? -- six measures that you have submitted for -- that you have developed.

DR. SCANLON: That would be great. Thank you. So let me start, actually, with the last one, 0343, which is PICU Standardized Mortality Ratio. I suspect the panel is very

familiar with the concept of SMR, and I would like to thank Dr. Winkler and Katie Streeter for their assistance in helping us navigate the language and terminology of NQF.

This is a complex measure, because it uses a proprietary risk adjustment scheme, the PRISM III algorithm, which is currently the only validated and calibrated severity of illness tool for pediatric use available in the States.

I think, in full disclosure, there is an international tool called PIM2 that has not had published validation in the U.S. but has been validated overseas, but for that reason PRISM III is used. That is used for SMR and also the complex measure of PICU Severity Adjusted Length of Stay.

This is based on work by Murray Pollack who actually is the intensivist who created the PRISM algorithm, and was published in a Journal of Pediatrics article back, I believe, in '96, describing the methodology

for risk adjusting length of stay to account for variation attributable by the severity of the patient, independent of the care provided.

Having said that, I think the adjusted length of stay is always subject to potentially gaming in the eyes of reviewers. So when these measures were originally put forth, and we still feel strongly that to look at severity adjusted length of stay in absence of an unplanned readmission rate is probably a mistake.

At least anecdotally, there are rumors that there are centers that keep their length of stay down by prematurely transferring kids, but the thought was, if you identify those kids who bounce back because they were sent out prematurely, by looking at unplanned readmission within 24 hours, that is almost a balancing measure.

So that touches on 0334 and 0335. 0336 is an attempt to actually add quality learning to the process, which is essentially

are you systematically reviewing your unplanned readmissions in the attempt of trying to reduce those. So that if you identify factors that could be addressed organizationally to reduce the likelihood of that happening in the future, that would be ideal, and we felt just to track a number of unplanned readmissions without learning from them would really be a missed opportunity.

Finally, the last two measures 0341 and 0342, are measures that at the time were Joint Commission elements. I will say The Joint Commission no longer requires these, which was a surprise to us as we were going back through these, but address the fact that we felt that pain was a very important aspect of care in the ICU.

For that reason, creating an expectation that pain be assessed at the time of admission and then in an ongoing fashion during the stay in the ICU was something that was important, and the pediatric critical care

community felt should be tracked and reported in a public manner.

So I don't know how I did in your three minutes, but I would be happy to clarify any of that.

CO-CHAIR GROSSBART: Are there questions from the Committee for the developer? Mitchell?

MEMBER LEVY: Yes. I have a question about 0336, the numerator and denominator. I read this. I keep reading it over and over again, and I can't tell the difference between the numerator and the denominator.

So I am not sure if it is just it is written -- I understand that the intention, just as you described, is to look at whether or not there is review of unplanned admissions, but the way it is written, it looks like the denominator is are there reviews of unplanned admissions, and the numerator is the number of nonreviews of

unplanned admissions. I can't tell the difference.

DR. SCANLON: You know, I don't have my copy in front of me. So I am at a bit of a disadvantage, but you are correct in your confusion, in that it should be that the denominator be all unplanned readmissions within 24 hours to ICU X within time frame Y, and the numerator would be, of those in the denominator, how many were reviewed?

MEMBER LEVY: Yes, it is not quite written like that, but that is sort of --

DR. SCANLON: I apologize for that and, certainly, that is the intent, spirit, and thought behind the measure.

CO-CHAIR GROSSBART: Any other questions for the developer?

MEMBER ALMENOFF: Just one. We have been having a lot of discussion about standardized mortality for adult measures, and part of the discussion had to do with not doing an in-house SMR but doing a 30-day SMR.

I noticed that you guys only do an in-house SMR. Was there any thought of maybe eventually developing a 30-day, because the information is much more useful to have both as opposed to just in-house where lots of gaming occurs?

That also will affect your length of stay risk adjusted model, because you said one of the co-factor is the rebound back into the hospital, but the other point is, if they go home and die, you won't see that piece of it. So it is just a thought of whether you thought of going to a 30-day model.

DR. SCANLON: So it is a very good question, and the answer is I don't think that -- Well, let me answer it a couple of ways.

First, these were the measures that were developed by a national task force, a kind of a self-formed and then nationally organized task force of pediatric critical care providers a number of years ago, before they first were submitted.

Actually, we put those together with the hope of getting Joint Commission endorsement before it had even struck us that we might be even ready for NQF prime time.

So having said that, there has not been a reconvening of that group, and I think that that would be a reasonable charge to us to try and pull together.

So when were contacted about resubmitting the measures, a lot of this has fallen to myself and Ms. Gall as I was one of the clinician leads on the measures a long time ago.

The VPS has essentially assumed the reins of stewardship for these, in absence of the prior organizations who had been doing that.

The 30-day measure is again provocative. I think the challenge is how to track those, and it is not that we are not willing. I think we would need to explore what is the feasibility of tracking.

One of the reasons is that a lot of these are admissions to tertiary, quaternary referral centers that are regional, and so our patient distribution often crosses many states. I am not offering that as an excuse, but I think instead one of the challenges to capturing that information.

So you are right. I don't have any doubt that a 30-day SMR would improve our understanding of care delivery. I always looked at these measures as our crawling before we walked before we run, and not ever claiming they were truth in the universe of all quality in pediatric critical care.

So I don't know that I actually answered your question, but I guess that is how I would respond.

CO-CHAIR GROSSBART: Again, any further questions?

DR. SCANLON: Oh, I would add one other thing. It is -- Chris was scribbling a note to me. It is really rare that children

die at home, for better or worse. So the part of it is tracking them across hospitals, but at least I think they are seen in mortality figures in a state.

Whether we could link those back to a given ICU, I think, is the challenge, and some of that is due to HIPAA limitations of tracking patients across institutions currently.

CO-CHAIR GROSSBART: Thank you, Matt. I want to do a time check. We have got -- Lunch has arrived, I believe, and maybe this would be a logical time to do public comments, and then lunch, rather than trying to squeeze the length of stay measure in before lunch. We want to make it a working lunch as well, so we can stay on track.

I think we can do public comments now. WE have 10 minutes for public comment on our morning work, and then we can adjourn for lunch, sit down, get back here, get reconvened, and start going in a few minutes

after we have all settled in.

DR. WINKLER: Matt and Chris, can you live with that? WE have been working all morning. We need to take a break.

DR. SCANLON: Oh, no, not at all. That would be perfectly fine. You know, I can just hang on the line here. I think the key is just knowing when would be -- if you can give us a time of when you would like us available again, so we can plan accordingly on this end. So even if it is a 10-minute break or whatever.

DR. WINKLER: Probably 15 minutes.

CO-CHAIR GROSSBART: Yes, 15 minutes.

DR. WINKLER: Thanks. Anthony, the operator, is there anyone on the line, audience who may want to make comments? We do? Then could we ask if anybody wants to make a comment or ask a question for public comment?

OPERATOR: If you would like to

ask a question or have a statement for public comment, please *1 on your touchtone telephone. Once again, that is *1 if you have a question or comment at this time. We will pause for just a moment to give everyone a chance to signal.

It appears we have no questions or comments at this time.

DR. WINKLER: Great. Does anybody in the room, in the audience? All right, I think we have done public comment.

Lunch is served, and so if we could grab lunch and get ourselves back here so we could get rolling again in about 15 minutes. Thanks, everybody.

(Whereupon, the above-entitled matter went off the record at 12:04 p.m. and resumed at 12:19 p.m.)

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

(12:19 p.m.)

MEMBER STOCKWELL: All right. I guess the first one on the agenda here is the PICU Severity-Adjusted Length of Stay. The introduction by the folks there in Wisconsin, I think, was great, and the length of stay calculations pretty straightforward.

Well, I guess I should do this in order. The impact of this metric, I think, is pretty self-explanatory. The performance gap is also noted within the VPS data that is within the application, that there is a fairly wide performance gap with a range from 1.something to 4.something average length of stay, and the evidence for using the severity-adjusted model that they used, the PRISM III score, is very well validated. It is on its third iteration, it sounds like. For the adult folks in the room, it is very similar to the APACHE score, and it has gone through the same kind of evolution as that has. Murray

would tell you that he very proudly ripped the idea off of the APACHE score, too.

So I think, at least for the first three parts, I think it is all pretty sound. We will talk about in a little bit, as I would understand what we have been doing, the feasibility piece comes up, because is not a publicly reported measure. This is a private group of pediatric ICUs that participate with VPS. We talked about that a little bit in our Work Group.

So I think one of our overall Work Group questions for NQF was: How does that skew our decision making about this metric? Does it weigh into it? Does it not? If it does, how so?

DR. WINKLER: In terms of just responding to that question, as Matt mentioned, we did clarify this is a complex measure that contains a proprietary risk adjustment methodology, which is allowable. However, one of the very important

considerations is that you look at whatever fees are associated with it under the criterion of feasibility.

So the details of belonging to VPS to have access to the risk model for this measure and the Standardized Mortality measure has been provided to you. Actually, Katie is projecting it.

So we want you to be aware of this information so that you can incorporate it into your assessment of feasibility.

CO-CHAIR GROSSBART: Reva, I have a question. So how did these measures get into the mix? I know that developers, particularly involved with health policy, CMS, Joint Commission and so on, submit measures so they can use them for the accountability/public reporting. Does NQF solicit these measures as part of a larger contract or is this just the developer has just submitted them?

DR. WINKLER: The history of this,

as Matt alluded to, is there was a task force of several national entities around child health and pediatric intensive care units who got together, including NACHRI, several children's hospitals, and VPS, I think, was sort of the data manager.

So that was the group that came together to develop these measures around 2005, something like that, with the intention of wanting to put these on the national stage for use, and they did. I think he mentioned perhaps the Joint Mention, and they did bring them to NQF.

DR. SCANLON: Dr. Winkler, this is Matt. If I could just clarify, too. At the time this was done, VPS was in existence but really had no direct role in the measure development. I think the important thing to appreciate, to the questioner's question is that at the time Dr. Pollack actually was in ownership of the PRISM III algorithm.

So this wasn't an effort to kind

of drive business to VPS. This was actually -- - VPS was licensing the algorithm from him for our use, but he had his own database that was also available at the time, the PICUEs database.

The issue is that Dr. Pollack has -- with updating the PICUEs database, and sold the license for the algorithm to Dr. Randall Wetzel who is affiliated with VPS, but his purchasing of it was a separate endeavor.

So I don't want to suggest that -- I think it is important to understand the historical context, so that this wasn't a proprietary company trying to guaranty business for life. This really fell out a different way.

MEMBER HAECKER: How many children's hospitals participate in this?

DR. SCANLON: One hundred seventeen children's hospitals -- or I'm sorry, 117 PICUs, and there's just over 100 children's hospitals in the U.S. and Saudi

Arabia currently. Canada is -- The Canadian PICUs are looking on joining.

MEMBER STOCKWELL: Matt, that sounds -- That is VPS membership, right? How many of those groups submit PRISM data?

DR. SCANLON: Eighty-five percent.

MEMBER STOCKWELL: Okay.

MEMBER STEARNS: Could you clarify for me whether the measure is publicly reported in terms of the data or whether it is not either as well -- Okay.

DR. WINKLER: Individual hospitals may choose to report different measures, and I have seen public reports of the mortality measure from several children's hospitals that they do on their own website. So it is an independent kind of thing, but there isn't a single entity that does it for a whole bunch of sites.

MEMBER HAECKER: It doesn't go to NACHRI or CHCA at all?

MEMBER STOCKWELL: No. No, not at

all. In fact, what you get is a standardized report with your hospital's data, and then you go through a fairly extensive process with VPS to select who you feel like your comparator or peer hospitals are within that group of 100 different children's hospitals, and you are blinded to what the results are from those other places, but you get whatever length of list as you are looking for.

DR. SCANLON: I would also add that the California Children's Services, which is the funding body for California pediatric health care, has mandated public reporting through VPS.

So, actually, there is one example where these measures are being reported, at least to the state. I don't know that it is in the public domain. I can't speak to that.

I think that is the direction they are going. But they are already mandating reporting of all these measures currently to the state.

DR. BURSTIN: And just one more context setting point, since this is the first proprietary measure you will have talked about over the last couple of days. A few years ago, the NQF Board specifically allowed a corridor for proprietary measures to come forward. We had never had any before.

The idea was that there was a fair amount of innovation in that community, and we wanted to make sure we were, in fact, getting a chance to have the full transparency to see under the hood of some of these, for example, that are quite proprietary and not very transparent. So that is a requirement, that it be fully transparent to the committees reviewing them to, in fact, see what is inside.

Secondly, if there are fees associated with the use of the measure, that they needed to be shared with the committee, shared with the public as part of the review of the measure, and that we would incorporate

the consideration of the fees under feasibility. So as we go through that, we will pop that slide up again with the fees. So back to you, David.

CO-CHAIR GROSSBART: Can I just make one more comment, because I am still trying to get my head around. So every other measure we have looked at has had a compelling policy reason for evaluation, in PQRS, in Value Based Purchasing, in being used for accreditation of health plans, even being used for accreditation of hospitals, and so on.

I am just still asking the question. I mean, I know NQF may choose to endorse these measures, but at the end of the day, I mean with the possible exception of the California public reporting, it is like what difference does it make?

DR. BURSTIN: Again, I think it gets to the fact that we really consider measures appropriate for both quality improvement and accountability. So there is

no question, many of these that are registry type measures or measures along these lines have been incredibly useful and often demonstrated results in terms of improvement.

The accountability functions, again, are quite broad. So they may, in fact, be used for other purposes, benchmarking with improvements, state based issues, health plans, pay for performance. So not everything needs to rise to the level, for example, of a Federal program, use in a Federal program, or pay for performance necessarily.

Public reporting still is the end goal, certainly given the preponderance of consumers and purchasers in the leadership of NQF. That is still a very strong goal. I think the hope is, over time, as we have seen, for example, with the STS database probably being the best example, the CABG database now being publicly reported as part of consumer reports and on the STS website -- they were endorsed through NQF for many years, with

continued sort of pushing to move in that direction. So I think the hope is that we bring in some of these innovative, important tools, and perhaps -- and measures, over time they can move in that direction.

DR. SCANLON: I would add, again, during the original meetings that led to the development of these measures, one of the criteria we used for measure development was would we be as a group comfortable, based on the state of knowledge, with this used for public reporting; and the answer was yes.

So I think -- My personal opinion is it would be a mistake to fault the measure, because nobody cares to force us to publicly report it. You know, that is outside of our bailiwick. There are lots of centers that are actually volunteering this information readily.

Parenthetically, the VPS is a database -- a registry. It is not in the reporting business. So I don't think that

that is where it would be, plus who really cares what the VPS has to say at a certain level. The absence of the joining to our NQF efforts, somebody -- or the public body saying this has to be reported can't make it happen.

CO-CHAIR GROSSBART: Matt, appreciate the feedback. So we are going to move on with our process here. So let's get back on track. So I think we are ready to start voting on sections of this. So, David, if you want to -- a question of impact for PICU length of stay?

MEMBER STOCKWELL: I think I already covered it. The impact is pretty clear. The performance gap is very reasonable to consider, and the evidence for the use of this metric, albeit with the caveats that we have just mentioned, is fairly sound.

CO-CHAIR GROSSBART: So it let's take those three in a bundle. So for all three of those, just to keep this moving, impact, performance gap, and evidence, are

there any questions or comments from the Committee? Norm?

MEMBER EDELMAN: I don't understand exactly what the impact is. Is length of stay in the PICU a measure of the quality of the final outcome or is it only a measure of resource utilization?

MEMBER STOCKWELL: I think it is severity-adjusted length of stay.

MEMBER EDELMAN: Is severity-adjusted length of stay a measure of the quality of the final clinical outcome or is it just a measure of resource utilization?

MEMBER STOCKWELL: Yes, it probably is a combination of both, but it certainly, I think, allows the individual ICUs to be able to assess themselves in terms of what can be done to get back into the norm within comparative groups.

MEMBER EDELMAN: I understand norms and comparative groups, but is this in the patient's best interest or is this in the

interest of the bottom line of the institution?

MEMBER STOCKWELL: I think that Matt Scanlon mentioned that a little bit, that there is a balancing measure that is sort of part of this package where there is the unplanned readmission rate for the ICU, to help to address that.

MEMBER EDELMAN: No, this measure. Is this going to provide healthier babies? It is not clear to me. It is not clear to me that an extra day makes a worse clinical outcome, and --

DR. SCANLON: Well, I would argue, from the standpoint of hospital acquired infections, every moment in the ICU that he doesn't need to be there, increases your risk of mortality.

MEMBER EDELMAN: And you have data to support that.

CO-CHAIR GROSSBART: Well, and that definition of -- Your definition of

quality is inconsistent with the IOM's, which would use efficiency as a measure of quality and overutilization or waste is --

MEMBER EDELMAN: Okay. So you think resource utilization is a sufficient rationale?

CO-CHAIR GROSSBART: My opinion is irrelevant. The IOM has said so.

DR. BURSTIN: It is intended to be a combination of resource use with quality, and that is why it is risk adjusted, and it is has got outcomes associated with it.

CO-CHAIR GROSSBART: Any other comments on the first three items in the voting ritual here? Ritual is not the right word -- process. So let's move on. So then we are going to vote on impact, a one to four scale again.

The vote is nine votes High, seven votes Moderate, one vote Low, one vote Insufficient Evidence.

The next item we will vote on is

the performance gap, and David has already addressed that. So unless there are any specific questions around performance gap, let's move on with the voting, one to four scale again.

The vote is eight with a score of High, nine with a score of Moderate, and one with Insufficient data.

Then the final question is evidence, and again this is an outcomes measure. So one for yes, two for No. We are still not getting everyone to vote, so try voting one more time.

We have 15 Yes and three Insufficient Evidence.

That moves us to the next phase of our voting, which is reliability and validity. David?

MEMBER STOCKWELL: The reliability, I think that the group felt comfortable with. Again, this is -- The approach that is used to generate this data

has been shown to be reliable, and has been validated in several different articles. I think the recommendations were both high on those items.

CO-CHAIR GROSSBART: Any questions or comments from the Committee? Let's move on to the reliability vote, on a one to four scale.

The voting was 12 High, six Moderate.

The next question is validity. Again, any additional comments, David? Any comments regarding validity? Let's move on with the vote then, one to four scale.

The validity results are eight with a score of High, nine with a score of Moderate, and one with Insufficient Evidence.

That moves us to the usability discussion. David, I know you have touched on some of these points. Anything to add? Any questions about usability? Then let's move on with our voting, again a one to four scale.

The vote was eight with a score of High, nine with a score of Moderate, and one with a score of Low.

Then feasibility, again a one to four scale.

MEMBER STOCKWELL: This is the big question, I think, and you can see up on the screen this is a decent chunk of change to participate in this, and then your payment to the company is only the first step, obviously. I am sure many of you participate in registries like this. The big chunk really comes in the manpower that it takes to generate this data and submit it.

I am not sure how to guide the conversation in terms of NQF standards for this question any further than that. So if you guys have any other recommendations, I would love to hear them.

DR. BURSTIN: We don't really have any standards for this. Actually, to date we have not endorsed any measures with fees. We

have evaluated some, but they have usually failed for other reasons. Well, I take that back. We have endorsed a couple of new Ingenix measures. So we actually have recently endorsed our first two.

MEMBER LEVY: So what percentage -- I know you said 85 percent of the hospitals in that system are reporting this, but what percentage of PICUs in the U.S. are already doing this, would you say?

CO-CHAIR GROSSBART: Matt, do you know the answer? I don't know the answer to that.

DR. SCANLON: I'm sorry. Could you repeat the question? I wasn't sure.

MEMBER LEVY: What percentage of PICUs in the country are already reporting this?

DR. SCANLON: You mean through VPS or outside of VPS?

MEMBER LEVY: Either way.

DR. SCANLON: To my knowledge, no

one is reporting it outside of VPS, because again --

MEMBER LEVY: Right.

DR. SCANLON: -- they need the PRISM III algorithm, and while Dr. Pollack had previously offered an alternative method for doing that, no longer supporting that project. so to my knowledge, the only groups that are reporting this are through the VPS use.

CO-CHAIR GROSSBART: Yes. So I am asking the question. What percentage of PICUs in the country are reporting it through you? Do you know what percentage of PICUs are --

DR. SCANLON: There's two different questions there, as I am hearing it. One, we represent about one-third to one-fourth -- One-third of the ICUs in the country, pediatric ICUs, are using VPS currently. There is a very large number of small -- of six-bed community hospital ICUs that aren't represented in this.

Having said that, again, I cannot speak to the reporting aspect, if you mean publicly reporting. Reporting to VPS, we can speak to. Reporting beyond that is up to the individual institution.

CO-CHAIR GROSSBART: Yes, Matt. What percentage of the total PICUs in the U.S. are in your database? We are asking feasibility.

DR. SCANLON: Right now, as we understand it, about a third.

CO-CHAIR GROSSBART: About one-third are in there. Would it be different if you counted PICU days or PICU admissions? What percentage of PICU admissions? I assume you have the larger facilities.

DR. SCANLON: I would guess that it is a much larger percentage from that standpoint. That is an excellent question, because this disproportionately represents large tertiary, quaternary centers. The problem is there is no place to go to get that

comparative data, that I am aware of.

CO-CHAIR GROSSBART: To get paid a premium on the top of billed charges.

DR. SCANLON: I mean, the KID database or HPEP data, theoretically, we could back into percentage of PICU days, but I think there would be some fuzzy math there. It is a great question. We can try and dig into that, but I don't know that that is even answerable.

CO-CHAIR GROSSBART: Okay. That was very helpful. Let's move on with the feasibility.

MEMBER LARSON: I have a question. is this a new measure or a renewal, because based on the number, I thought it would be a renewal. Somebody said we never had approved these before.

DR. WINKLER: It was approved in the past. It wasn't part of a proprietary -- It was at a time of transfer of ownership that Matt told you about.

CO-CHAIR GROSSBART: So are still

looking for some guidance on the feasibility vote?

DR. WINKLER: If this helps you, feasibility is not a must pass criteria. So you are willing to assess -- you know, is this a problem for the feasibility of the measure? You will then vote on whether you would recommend the measure for endorsement, but unlike importance and unlike scientific acceptability, a measure does not have to pass feasibility in order to be recommended.

CO-CHAIR GROSSBART: Okay. With that said, are there any other further questions before we vote? So it is a one to four scale, again. Let's vote.

Zero, High; seven, Moderate; eight, Low; and three, Insufficient Information.

Then our final vote will be on endorsement of the measure, one Yes, two No.

The endorsement carries by a vote of 11 to seven.

CO-CHAIR GROSSBART: Our next measure up is going to be -- Actually,. I am going to ask if Mitchell and Peter would be willing to kind of tag team on the 0335 and 0336, just in the introductory parts of this, because these are similar measures: PICU Unplanned Readmission Rate and Review of Unplanned PICU Readmissions. Are you comfortable? Great.

Mitch, do you want to lead?

MEMBER LEVY: This measure got mixed reviews by the committee, and I think in part it is because of the confusion between these two, the two metrics, which we had clarified a little bit. This metric has a more clear numerator and denominator, in that it is clearly measuring the incidence of unplanned admissions back into PICUs.

It is being proposed as a balancing measure with the one that was presented previously, which is length of stay. So it has the potential for high impact. The

problem is, although it is being presented as a balancing measure, the reason for readmission is confounded by factors on the wards, factors in the hospital, and then factors that led to the discharge in the first place.

So although it is being presented as a balancing measure to ensure that hospitals aren't driven to discharge kids from the ICU more quickly because of the length of stay, the readmission rate may not really reflect that. So there is a question about that.

The scientific evidence is confounded also in that a lot of the evidence that is cited in this, first of all, appears to be from adults and, second, appears to be measuring the impact of rapid response teams on wards to reduce readmission rates. So it is very confusing what data are being presented in support of the metric altogether. I will stop there.

CO-CHAIR GROSSBART: Any additional comments?

MEMBER ALMENOFF: I agree with the comments that were made. The only piece to the second portion is of the -- and I think that was clarified today -- of the patients who get readmitted back to the unit within 24 hours. The second portion of this metric was that they would review all those charts. I mean, that is basically all it is doing which, to me, almost seems like, why not just make that part of the 0335.

Of course, if you are going to bother to do this at all, why are you writing a separate metric to look at what you should be doing. So, to me, it just seems a little different, but basically I agree with most of the comments that Mitchell made.

CO-CHAIR GROSSBART: Can the developer briefly comment on the relevance of 0336, and why you -- your logic behind including it?

DR. SCANLON: Well, I think, actually, in a way that was answered by the presenter. What I mean by that is 0335 is subject to confounding factors, such s the capability and resources of the acute care unit that the patient was transferred to.

I can speak to our own data. We have seen children who have come back as an unplanned readmission for new unpredicted problems that were not foreseen and could not have been foreseen.

So that I think, 0335 in and of itself, is necessary but not sufficient.

To the second commenter, you could argue, and I think the case could be made, that they could be merged, and I at a certain level have no problem with that. I think, again, the time these were developed, they were thought of us different pieces -- different legs of the stool, if you will -- and whether it would be enhanced by combining these or not is, I think, open to discussion.

But it is to understand exactly where there are areas to improve and what was stuff that was beyond the control of the ICU, but may even point out the hospital system changes that could be addressed.

CO-CHAIR GROSSBART: Thank you for that feedback. At this point, I think it would be appropriate to move to our systematic voting, and the first measure up is 0335, PICU Unplanned Readmission Rate. We have touched on some of the high level -- Hayley, a question?

MEMBER BURGESS: Yes. How is unplanned readmission defined? I was looking through. I couldn't find exactly what that definition is.

CO-CHAIR GROSSBART: Actually, wouldn't all readmissions be unplanned, because you wouldn't have --

DR. SCANLON: That is actually incorrect. For better or worse, because of resource issues, it is not uncommon in ICUs

for patients to be transferred out for a time period and then come back after a subsequent procedure. We could debate whether that is ideal care or not, and I would concede that point readily, but it is a known phenomenon that there are predicted readmissions to ICUs, even within a 24-hour time period.

CO-CHAIR GROSSBART: All right. My apologies.

DR. SCANLON: Oh, not at all.

CO-CHAIR GROSSBART: Hayley, did your question get answered? How is it defined, though?

MEMBER LEVY: Yes, I didn't talk about it, because I think that is under reliability, but clearly, that is -- and it is mentioned even in the submission that it is a very subjective definition.

CO-CHAIR GROSSBART: Okay. So let's step through the voting process.

DR. SCANLON: Would it help to clarify? There is a standard definition at

least for this, which is did the ICU know within 12 hours of readmission that the child was coming back?

CO-CHAIR GROSSBART: Okay, thank you for that feedback.

Again, the importance of the measure, one to four scale. We are going to have to vote separately on each one. We can't combine the two votes together, can we, Reva?

DR. WINKLER: No, you can't really.

CO-CHAIR GROSSBART: So we just go straight through 0335, then straight through 0336 with just briefer comments by the Committee since we have already --

So the results are six High, nine Moderate, four Low.

Then moving on to the next piece, the performance gap.

MEMBER LEVY: In the submission, there are two aspects. One is referring to the value of an outreach service, which I

assume is like a rapid response team, in reducing readmissions, and the other was the analysis from VPS of the variation of readmissions between zero and 3.14 percent of discharged patients. So there is some variation across hospitals of about three percent.

CO-CHAIR GROSSBART: Any comments or questions? Let's move on to our voting, a one to four scale for the performance gap.

The results are one score of High, 11 Moderate, seven Low.

DR. BURSTIN: Katie, could you scroll the screen up? The other way, I'm sorry. Thank you.

CO-CHAIR GROSSBART: Then the final question is -- and this is an outcomes -- the evidence, and it is a Yes/No question. Any comments about the evidence, before we move on? All right, let's move to the voting.

The results are 15 Yes, one No, three Insufficient.

Now we will move to reliability and validity.

MEMBER LEVY: Well, so it sounds -- If there is a standardized definition of unplanned, it makes it more reliable. The committee was -- I think you can see here -- split on the validity. I said this already. So I won't repeat it, but it is presented as a balancing measure, but what this metric actually reflects is not clearly balancing length of stay appropriately.

MEMBER EDELMAN: Is the plan to readmit easily documented?

DR. SCANLON: I'm sorry, are you asking --

MEMBER EDELMAN: The definition is unplanned readmission, and I am asking how hard it is to document that the readmission was planned.

DR. SCANLON: Actually, that is pretty easily found. Again, children may be transferred out, because, for example, the bed

is needed for another more acutely ill child, but there is clearly reservations made in an organizational system to readmit the child, which makes tracking of this actually pretty clean.

MEMBER LEVY: Yes, this doesn't happen in adult ICUs.

DR. SCANLON: You've got a lot more ICU beds than we do.

CO-CHAIR GROSSBART: So there further questions about reliability?

MEMBER STOCKWELL: I would just offer our experience. We track this number. It is something that is essentially in the -- that the definition is in the absence of any plan to have the kid come back, then it meets the criteria. Our experience, if it helps to clarify who these kids are, many times they are seizure kids. They are asthmatics. They are respiratory kids that you anticipate are on the right trajectory, and they are just -- they need more care than was anticipated. I'm

not sure if that is helpful, but I thought I would offer it.

MEMBER WHETSELL: To me, that kind of sounds unplanned.

MEMBER STOCKWELL: That is what I meant. I meant to say that. If I didn't, that was a mistake.

MEMBER WHETSELL: Okay.

CO-CHAIR GROSSBART: All right. Seeing no body language suggesting other questions, let's go to our voting on reliability, a one to four scale again.

The results are two votes for High, 12 for Moderate, four for Low, and one for Insufficient Data.

Now validity. Any comments from -- Mitchell, anymore comments? So validity, any questions, comments from the Committee? If not, let's vote, a one to four scale again.

No votes for High, 12 Moderate, six Low, and one Insufficient.

So we can now move on to the

usability and feasibility questions. Mitchell, usability, any additional comments?

MEMBER LEVY: No, not really.

CO-CHAIR GROSSBART: Any comments or questions from the Committee? Christine.

MEMBER STEARNS: I know we discussed that at length previously, the highlights that concern that I have . It is not with the measures, but with the lack of public reporting. It is hard for me to understand the usefulness of our endorsement of the measure, given the proprietary nature.

DR. WINKLER: This measure, I believe, does not include the proprietary aspect. So all of the specifications are laid out here and could be picked up and used. In fact, two of these measures, the pain assessment measures, have been retooled for EHR use. It is only the two measures that use the PRISM that fall into that proprietary.

MEMBER STEARNS: Okay. This one does not include the complication with the

PRISM proprietary?

DR. WINKLER: No.

MEMBER STEARNS: Okay. But right now the end use of this, the measure results are not publicly reported in any way. So it is -- The measure itself could be picked up and used by someone else, because there is no part of it that is proprietary. However, it is not a publicly reported measure. Okay, thank you.

CO-CHAIR GROSSBART: Okay. So usability, a one to four scale. Any other questions? Let's move on and vote.

The final results are four votes High, 12 Moderate, three Low.

Then feasibility. Any comments, questions? All right, let's move on to the voting, a one to four scale.

One vote for High, 15 Moderate, two Low, one Insufficient.

Then overall endorsement of the measure, one, Yes, two, No.

Sixteen Yes and three No. The measure is endorsed by the Committee.

Then Peter, if we can step through -- Questions?

DR. RHEW: I just wanted to make a comment. We talked earlier -- this is more about the overall process -- that this is, I guess, the sister measure to the length of stay, and we have identified these that are tied to the hip. If it had turned out that we had voted no on this and we kept the length of stay, it might have created some issues whether the length of stay should have been valid.

I am just wondering from a process standpoint, when you identify those issues, is there a way to somehow address that; because to tell you the truth, that was a factor in terms of my decision, whether or not length of stay was actually there. The fact that it was there, I felt that you had to have a sister to that.

So I don't know if we want to address that, but I just thought that was an issue that we at least should be aware of.

CO-CHAIR GROSSBART: That is a great comment. Is that something staff can work up for the next meeting. Right?

DR. WINKLER: If you all would like to, we do have a concept of pairing the measures. What that does is bind them at the hip, and so they travel together as a dual entity. That is a decision that you as a Steering Committee can make that recommendation that these measures be paired.

CO-CHAIR GROSSBART: So moved. So we are going to have a vote that 0335 and 0334 be paired measures? It has been moved. Is there a second? Do we use Robert's Rules?

MEMBER ALMENOFF: Second.

CO-CHAIR GROSSBART: Then do we get to use this? Why don't we just do a show of hands? Those in favor of pairing the measures? Those opposed? The vote was

unanimous.

MEMBER RHEW: I would also add for the pneumonia one, we talked about the in-hospital mortality and the 30-day. That also would be one that I would consider pairing as well.

CO-CHAIR GROSSBART: Well, let's move on to 0336. So, Peter. We want to move as quickly as we can, right?

MEMBER ALMENOFF: This is just an extension of the one we just saw, but this now requests that all the unplanned admissions actually get reviewed or documented that they are reviewed, and they are looking for 100 percent of the number.

Quite honestly, if you are doing 0335 and you are not doing 0336, then you shouldn't be doing 0335, because I can't see anything more ridiculous than tracking the data and not actually looking and seeing what the issues are.

So I am glad that Mitchell finally

got clarification. I think this was probably a Microsoft mistake, but if you look at the numerator, that makes sense. The denominator basically, the back of that, the clinical review is documented with the piece that talks about the same exact thing as the numerator needs to be pulled out.

So with that, if you look at the Committee's report -- I mean, we were kind of all wondering why we were looking at this, but we thought the impact -- we were sort of mixed, either high or medium, but if we think that 0335 needs to be done, then looking at the data and making some determination would probably be sort of important to do as a second piece to that.

CO-CHAIR GROSSBART: Any questions about impact?

MEMBER ALMENOFF: They should be all three together, actually.

CO-CHAIR GROSSBART: Performance gap. Any questions about impact, performance

gap or the evidence?

MEMBER YEALY: Yes. I guess I am still lost. We don't really have a definition of review, and I am not sure what we are achieving exactly here. This could be as cursory as possible, and I am not sure what outcome we would improve.

MEMBER STOCKWELL: I would agree with that, and we talked about that in our Work Group, and by definition, going through and making the determination whether or not something was planned or unplanned, you have reviewed the chart. So it is almost like you have satisfied this just by doing 0335.

MEMBER YEALY: Yes.

MEMBER STOCKWELL: So I am not sure how much value it is adding to the process.

MEMBER YEALY: We have a fake process over a hard outcome, and it is hard for me to figure -- as a numerator and denominator, hard for me to figure out who

wins in this.

MEMBER GLOMB: Yes, I would go along with that. All of the children's hospitals where I have been peds ICU attending, these came up as part of routine M&M for 15-20 years.

MEMBER ALMENOFF: I would think the gap would be zero, that probably 100 percent get reviewed.

MEMBER GLOMB: At least in my experience.

CO-CHAIR GROSSBART: What concerns me is: So if review, the unplanned readmission -- there is no clear definition of that. Since this is only being used by hospitals self-motivated for performance improvement, probably not a big thing. But if this were to hypothetically become a publicly reported measure that parents were using to make decisions about where they send their children, would we find this a reliable measure, or meaningful? Could it be gamed?

How easily gamed? How easily could you game it?

MEMBERS STOCKWELL: I think the corollary is, if a hospital reports a CLABSI rate, is it meaningful to have some kind of measure that says, oh, and we also review our CLABSIs? I am not sure where the value is there.

MEMBER ALMENOFF: Well, because you should be dong it. It should be part of the process. That is why this doesn't --

CO-CHAIR GROSSBART: Well, so let's go through our process then. Impact of what this measures?

DR. WINKLER: You can make whatever recommendations you want to make, but it seems like some of the questions that were on the table were more fundamental than that.

CO-CHAIR GROSSBART: So impact.

DR. SCANLON: This is Matt Scanlon, if I might weigh in a second. I think the Committee's comments are very

appropriate, and I can tell you this is the dilemma of being a steward for the measure versus how we have tried to handle this in the VPS.

In the VPS system, we have a series of structured questions to drive a systematic review with the goal of identifying system problems. At the time the measure was developed, there wasn't support to embrace a common framework.

So that is where, I think the Committee is dead on correct in that this could be incredibly superficial and cursory or it could be very meaningful and discover system level problems. That really is in the hands of the reviewer.

CO-CHAIR GROSSBART: Thank you for those candid comments. Are we still -- Is our voting still active here on impact? Okay.

The vote on impact is five High, six Moderate, seven Low, and one Insufficient. So that passes.

Now the next question is performance gap. High, moderate, Low -- or it is a one to four scale. Any additional questions about performance gap? Do we have evidence of a performance gap?

MEMBER ALMENOFF: We don't.

CO-CHAIR GROSSBART: What is that?

MEMBER ALMENOFF: I don't think we do.

CO-CHAIR GROSSBART: Okay. Any questions? The Work Group has said they do not see any evidence of a performance gap.

MEMBER ALMENOFF: Did anybody else see one? I didn't notice one.

CO-CHAIR GROSSBART: Does the measure developer want to add a comment before we move forward with this vote?

DR. SCANLON: No, I don't have any objective evidence. We do have anecdotal reports that a lot of these smaller ICUs that, depending on your perspective, one could argue dabble in critical care do not do this sort of

review systematically, but again we don't have any hard data to put to that.

CO-CHAIR GROSSBART: Thank you. We are voting now, one to four scale again.

MEMBER ALMENOFF: Doing the ritual?

CO-CHAIR GROSSBART: Doing the ritual. Going to bring incense next time.

So the vote is one Moderate and 18 Insufficient Evidence. That means we are done with this measure, and we will not move to an endorsement vote at this point.

Stepping through, Janet, I believe you are up.

MEMBER ALMENOFF: Could we add one piece, that maybe this needs to be incorporated in 0335? I think we need to make that suggestion, even though it is obvious.

CO-CHAIR GROSSBART: That is a great suggesting and, Matt, I don't know if you heard that, but the Committee feels that aspects of this could be merged with the

measure for 0335.

DR. SCANLON: I think that, knowing the original discussions that led to the development of the measure, we would be -- if I ever dare speak on behalf of the pediatric critical care community, there would be support for that.

CO-CHAIR GROSSBART: Thank you. Again, thanks for your very candid and helpful feedback as well.

Janet, you have the next two measures, and they are very similar. So if we can expedite this somewhat by merging some of the discussions, where appropriate, that would be helpful.

MEMBER LARSON: Sure. 0341 is the percentage of PICU patients receiving pain assessment on admission, and 0342 is the percentage of PICU patients receiving a periodic pain assessment, which is defined as every six hours during their PICU stay.

In the application, it is

presented that assessment of pain is important, and they cite a few references, some more current, and there is a statement, the American Academy of Pediatrics and the Canadian Pediatric Society both suggest that it is important. So that is basically the evidence.

CO-CHAIR GROSSBART: So let's ask the Committee and the Work Group if there are any additional comments about the pain management questions.

CO-CHAIR WEISS: Was there any discussion in the Work Group? It seems tome that this could be brought together very easily saying that the definition of continued being that there was one done initially at a certain time period and continued on, and to have two measures when you can just do one makes less sense. Was that discussed at all?

MEMBER LARSON: You know, I as sick. So I wasn't on the Work Group. So I am not really sure who was there. I don't see it

written up in the notes, but it does make sense

CO-CHAIR GROSSBART: Members of the Work Group, any comments or feedback?

MEMBER STOCKWELL: Yes, it came up that day. It would seem like it would be a reasonable thing to include. One is not necessarily more important than the other.

MEMBER YEALY: And the second one, if you did it every six hours, you would probably be covering the initial admission period. So it is hard for me to see why we would need the admission, part one.

MEMBER LARSON: Right. The admission is just defined using hospital policy, so whatever they consider on admission. So, yes, I think it could easily be combined.

CO-CHAIR GROSSBART: I think it might be appropriate to ask the measure developer. matt, I don't know if you heard the conversation, but the body language in the

room is kind of scratching their head over why two separate measures, that the measures should be --

DR. SCANLON: So you are testing the length of my memory here, but as I recall, part of this goes back to the original structure that we were advised by -- and I don't recall who it was -- on developing the measures originally, that they thought that melding these were problematic.

Again, I think I would be very comfortable, speaking on behalf of the measure development team, saying that we would have no problem combining the two, perhaps just changing the language: Should be every six hours, beginning at admission, so that they can't start the clock whenever they want.

CO-CHAIR GROSSBART: That said, Reva, what do we do now?

DR. WINKLER: I think that what we could do is say that 0341 is -- you are basically saying you don't see any need to

have it as a stand-alone, and then that 0342 would be the measure you would probably want to go forward with, sort of the understanding that they might -- that they would do whatever wording adjustment to add that factor in, because it seems to be a relatively minor change, as Don said.

It may already really sort of be there. It is just you want to be more explicit about it. So it looks like it really can wrap under 0342 with them maybe changing the wording to make it crystal clear.

CO-CHAIR GROSSBART: Then my next question is can we do one vote on a combined 0342-0341 or can the measure developer withdraw 0341 right at this moment?

DR. SCANLON: Whatever makes you guys happy.

DR. WINKLER: Matt, would you be willing to withdraw 0341 and then make whatever wording adjustments to 0342 to be sure that the periodic assessment starts at

admission?

DR. SCANLON: Yes.

DR. WINKLER: Then I think you can do one vote.

CO-CHAIR GROSSBART: Okay. Well, let's go through the process then. Will we be voting on a combined measure and formally endorsing 0342 with amendments? So impact. Janet, any additional comments?

MEMBER LARSON: No.

MEMBER EDELMAN: I have a question. Are there any exclusions for age?

MEMBER LARSON: Oh, under 18.

MEMBER EDELMAN: No minimum?

MEMBER LARSON: No, no minimum, and it does cover neonates.

MEMBER EDELMAN: I guess this is a little later on for validity, but are you comfortable that methodology exists for all age groups?

MEMBER LARSON: That, I don't know.

CO-CHAIR GROSSBART: All right. So impact, one to four scale again.

The impact scores were 12 votes for High, six for Moderate, one for Low, zero for Insufficient.

Performance gap?

MEMBER LARSON: So for performance gap, they have 14 units reporting to the VPS database, and in the last quarter results ranged from: For admission, it was 83 percent to 100 percent completion of the assessment; and for the every six hours, it was 77 to 100 percent. You don't have a sense -- that is just the range. You don't know how many or any sense of that, and that is the evidence.

When they implemented this in one unit, k there was a 10 percent increase in the assessment.

CO-CHAIR GROSSBART: Any questions or comments from others in the Work Group or on the Committee? So performance gap, we will vote on it. One to four is the range.

We've got five votes for High and 14 votes for Moderate on performance gap.

Now we move on to the evidence, and this is a Yes/No question, one, two or insufficient. Any comments, Janet?

MEMBER LARSON: No.

CO-CHAIR GROSSBART: The Committee rated the evidence fairly low -- or no, actually, they rated it -- well, they did write it low, but with a definite yes. So let's move on to the voting.

The results are 13 Yes, three No, three Insufficient.

Now we move on to our reliability and validity section. So reliability of the measure.

MEMBER LARSON: You know, it inherently sounds reliable. It is did they do it or didn't they do it, but they presented absolutely no evidence. They said that, because JCAHO endorsed it, they didn't need to present evidence, and they said that was true

for validity as well. I mean really nothing.

MEMBER BURGESS: Did we hear that JCAHO removed endorsement of this?

MEMBER LARSON: Yes, and then we heard that.

MEMBER YEALY: Well, and since the actual measure is not specified, it couldn't -- since almost anything would count for it, if that is your definition, not deciding whether that is useful or not, is a whole separate conversation. It almost can't be anything other than the reliability. It might not be valid, but --

CO-CHAIR GROSSBART: Matt, you heard the conversation. Any comment on it?

DR. SCANLON: Well, again, at the time these were put forth, this was absolutely consistent with -- and I don't remember the Joint Commission standard, but there was a standard that was on pain assessment that it was consistent with, and that was why we felt it was easy to create that as a publicly

reported measure for hospitals.

The challenge of specifying methodology is that it is an age and hospital based preference of scale. So the tool you use for a child of one age is different than another. It is confounded by developmental factors. It is confounded by issues of sedation and mechanical ventilation.

So there was a great resistance by the development committee, and even NQF acknowledged the first time we endorsed this that we would not try and dictate a single methodology for capturing this. That is what was done at the time.

Now, unfortunately, I think it has fallen by the wayside, but we could easily remedy that, is that NQF at the time asked that we publicly post on a website what our examples of validated aim assessment tools, and that was done.

My concern is I haven't been to that site in a while, and I don't know that

site is still up, but we have created a new site to post the measures, as is expected by NQF, and we could readily post that same information there.

I don't know if that helps at all, or not.

CO-CHAIR GROSSBART: Thank you. David?

MEMBER RHEW: I just wanted to comment, since the question came up whether the Joint Commission has actually endorsed these or not. The Joint Commission published in 2012 their pain management standards, and there are four standards that relate to assessment and reassessment in patients with pain, and they specify a variety of elements for performance, including when additional specialized, more in depth assessment should be performed, a whole thing around comprehensive pain assessment.

So the quick answer is that there is a very recent document, Joint Commission

2012. I can get you the actual reference if you want, but it is out there.

CO-CHAIR GROSSBART: While you have that up there, David, is it a performance measure or is it a Joint Commission standard, and there is a difference.

MEMBER RHEW: That is an excellent question, and all I can tell you is that the name of the document says 2012 Hospital Accreditation Standards, Elements of Performance Scoring Accreditation Policies, 2012. That is the title. I don't know.

CO-CHAIR GROSSBART: Again, I am blessed. I don't have to be accountable for Joint Commission, although someone on my team does, but I think that is basically the standards that the accrediting individual comes from Joint Commission accredit, and it is kind of subjective, isn't it? They assess your policies and your processes in real time.

So anyone that can help. Dianne was first.

MEMBER JEWELL: Well, I am not responding to that. So if you are responding to that, go ahead.

MEMBER WHETSELL: Yes. Going through Joint Commission, they do retro review. They do look at it in depth to see if there is documentation of a pain assessment, if there is an action performed, and if there is a reassessment completed from that action to see if there was good response or not.

So while it is a standard, let me tell you, when they are doing their tracing method and they are walking through your hospital from stem to stern and looking, they are watching to make sure that that happens in every single environment.

CO-CHAIR GROSSBART: But it is not a performance measure in the traditional sense.

MEMBER WHETSELL: You can get cited on it.

CO-CHAIR GROSSBART: Okay.

MEMBER WHETSELL: You can get cited on it.

CO-CHAIR GROSSBART: Right. Understood. Okay. Dianne?

MEMBER JEWELL: So recognizing what the measurement developer said earlier about pushback related to specifying a methodology, I have a memory that some of the other measures NQF has endorsed over the years have at least included the phrase "a standardized tool" or a standardized -- it leaves it up to the discretion of the user to pick which tool, but still is a little more directive than just do an assessment. Wouldn't that work here?

DR. WINKLER: Dianne is absolutely right. There is a preference for having a little bit greater specificity, so that any old thing that you may have created, you know, over lunch would not be acceptable, as opposed to when you ha a field where there are numerous tools or they are age based or

something like this, you can't overly specify. But you might want to specify to the degree that it is a standardized, validated instrument, dah, dah, dah, and then often "such as," and give examples that are not necessarily --

CO-CHAIR GROSSBART: The question comes to mind: So, you know, you have got a rate. I mean, ultimately, you get a rate on this. So what does 75 percent mean or 80 percent mean? Then from a practical standpoint, isn't the real test the HCAHPS score on quality of was your pain in control while you were in the hospital. Children's hospitals don't do HCAHPS?

DR. SCANLON: Oh, children's hospitals, to my knowledge, don't do HCAHPS.

CO-CHAIR GROSSBART: There is a project.

DR. SCANLON: From a developer's standpoint, I think -- My recollection is I thought we had the language, but clearly, I am

mistaken on that, about those standard tools. So I think the request for language saying a standard validated tool is actually absolutely consistent with the intention of the developer group, and we would readily be happy to address that.

MEMBER JEWELL: The language currently reads "a policy statement and compliance with Joint Commission expectations." But I think going for the more direct standardized tool would get the message loud and clear to the users.

DR. SCANLON: Very good.

CO-CHAIR GROSSBART: Well, with that in hand, we do need to vote on reliability. Does staff want to give us anymore guidance or are we on our own?

DR. WINKLER: The criterion asks you to evaluate the testing for reliability and validity of this measure in play, and then evaluate the results of that testing, and there isn't any.

CO-CHAIR GROSSBART: So on a scale of one to four, reliability of the measure.

The results are six Moderate, four Low, and nine Insufficient Evidence. That wraps it up. We will not be able to vote further on this measure.

The last one, David, I believe you up again on PICU Standardized Mortality Ratio.

MEMBER STOCKWELL: Yes. I think, basically, everything that was said about the one that I did previously about length of stay can be cut and pasted into this as well. Everybody understands what a mortality ratio is. It uses the PRISMS III method, which we have discussed.

So the impact, I think, is reasonable. The performance gap by internal VPS reporting is present, and it is an outcome measure. So the rationale is high.

CO-CHAIR GROSSBART: So questions for David or comments from the Work Group? Questions or comments from the full Committee?

Okay, let's go on to our voting, impact, again a one to four scale.

The vote was 13 High, five Moderate.

All right, let's move on to the next topic, which is the performance gap. David, any additional comments? All right, performance gap, any questions from the Committee?

MEMBER LEVY: Is there currently a different -- a similar for peds other than using the proprietary PRISM? Are there any mandated public reporting or any other -- I just want to make sure we are not going counter to something that is already in use out there. So PICUs don't routinely report SMR?

MEMBER STOCKWELL: This is the method. If there is a method that is most utilized, it is this approach, yes.

MEMBER LEVY: So that means -- So two-thirds of the -- and I understand the

difference between the actual ICUs versus the number of beds, but two-thirds of the PICUs in the country don't really report standardized mortality ratio?

MEMBER STOCKWELL: That's right. Mat can certainly speak to who is in it, but it is all the major players that are in the VPS dataset.

MEMBER LEVY: Okay.

DR. SCANLON: I think it gets to a very important distinction between what is publicly reportable and what is publicly reported. Again, in the absence of any stick or carrot to mandate it, some of us are doing it, and lots aren't.

CO-CHAIR GROSSBART: Okay. So back to performance gap, vote scale of one to four.

The results are 10 with a vote of High, six with a vote of Moderate, one with a vote of Low.

Then finally, the evidence. this

is an outcome measure. So any other comments, David? So this is a Yes/No question, or a three, Insufficient.

Our final vote was 17 Yes, one No.

We now need to move on to the reliability and validity questions.

MEMBER STOCKWELL: I will join those two like we did last time. The reliability of the severity of illness approach is, I think, sound, and the validity of the metric, in and of itself, is -- The recommendation for adoption would be high for those two things, from our Work Group.

CO-CHAIR GROSSBART: Any additional comments from the Work Group? Any questions from the full committee? Well, let's move on to voting for reliability, scale of one to four.

The results are 13 High, five Moderate, no other votes cast.

Validity, again a one to four scale. Let's vote.

The results are 12 High and six Moderate.

Now usability.

MEMBER STOCKWELL: I think we can tell from the general feeling in the room that it is probably not used enough. So I think the usability component evidence is fairly high.

CO-CHAIR GROSSBART: Any other comments from the Work Group? Questions or comments from the full Committee? Let's move to voting, again a one to four scale.

The results are 15 with a vote of High and three with a vote of Moderate.

Then the next question is feasibility.

MEMBER STOCKWELL: Yes. Again, this is the key point here, because this is where the only way you can get this number is if you participate with the payment of the fees that we showed before, same data for the entry fee, and then also for paying staff to

generate the chart review and the data submission. So the barrier to entry is not insignificant to participate with this metric.

CO-CHAIR GROSSBART: Same question we had with the first measure. So are there any additional comments or questions from the Work Group or from the full Committee? We will test our interrater reliability then. Let's vote. Not very good interrater reliability. We had much more optimistic -- What is that? People are tired.

Four High, six Moderate, five Low, and three Insufficient.

Of course, feasibility is not contingent -- does not impact the ability for us to move to endorsement. So it is a Yes/No question. One, Yes; two, No, for endorsing the measure.

Looks like we have 17 up there -- oh, here we go. The measure is endorsed by a vote of 16 to two by the Committee.

So we are now 55 minutes ahead of

schedule.

DR. WINKLER: Thank you all very much. Matt, Chris, thanks very much for being with us.

Before everybody kind of wants to gather up, take a deep breath. We do have a little bit of discussion left on some other issues that are broader, but not the evaluation of measures.

The first thing I want -- Katie, could you put up the spreadsheet of the measures that have been withdrawn. Just as part of the maintenance process, there were measures previously endorsed by NQF who, as we went out to request the maintenance review, the measure steward withdrew the measures from further consideration. So they are -- By the end of this process, they will no longer be endorsed.

So we wanted you to be aware of those measures. Many of the measures have been superseded perhaps, and so you see -- I

think you go up one more, right? Okay.

So the management plan for people with asthma, the severity standardized average length of stay, the VAP measure we talked about earlier, the hospital measure for initial antibiotic within six hours -- go ahead and scroll down -- and then a COPD assessment of oxygen saturation measure.

So these have been withdrawn by the measure developers. We wanted you to be aware. Do you have any comments on it? It will form the basis of the report, because as we do maintenance review of our measures, we have to account for all the measures.

MEMBER LEVY: I have a question about the antibiotics, because it is going to relate to what the critical care societies are recommending. Can someone remind me what the unintended consequences of -- I know that CMS withdrew the initial antibiotic, and that was for CAP in the emergency department, wasn't it, Don?

MEMBER YEALY: Yes, and some unintended consequences, antibiotics given out like water for anything that could remotely be considered an acute pneumonia case, and that the data that were being generalized for the metric (a) never defined six hours as a particular break point, and (b) were a very narrow group of plain chest X-ray generated pneumonia.

So one of the other things that drove this one nuts is that people who would get -- I will tell you a common scenario would be you get an abdominal scan, and on the scan see something in the lower lobes that may or not really be pneumonia, but once it was entered as an infiltrate, you then invoked this metric.

So the way to get around that was to give everybody broad coverage right away and just create a whole different set of problems. So it was because the six hours was never placed on these data, and you created a

-- opening the dam on antibiotic therapy.

MEMBER LEVY: Yes, that is helpful. So it doesn't preclude a future metric looking at timing of antibiotics. It is just the way that one was written. Okay.

DR. WINKLER: Any other comments about these measures? Like I say, it is part of our accounting system to make you aware and to see if there are any issues. Okay. Mark?

DR. ANTMAN: Thanks, Reva. I just wanted to mention that measure 00001, the asthma assessment measure -- we didn't submit that at this time, simply because we are in the midst of a pretty significant revision of that measure.

So it is not -- I just wanted to convey to the group that it is not as if we don't feel that it is important that there be a good measure of asthma assessment, but we didn't have it ready at this point. Once we do, we will plan to resubmit it.

DR. WINKLER: Thank you, Mark. So

there typically are these rationales behind the withdrawal that are usually evolution of measurement has occurred, and revisions and updates and new measures are typically the reason that measures are withdrawn.

So that one agenda item. Yesterday when we were having our technical difficulties, we particularly when we were talking about the last asthma measures, particularly the all or none composite measure from Minnesota, there was a real communication challenge, their hearing us, not hearing us.

The discussion that ultimately ended in the group voting down the measure was around evidence. That is where you voted it low, so that it did not meet the importance criteria.

That had been a relatively new focus of discussion and concern compared to the Work Group conversation. So Minnesota felt they had not had an opportunity to respond to your concerns.

Last night we received from them -- and I forwarded on to you all -- some additional information that they provided from the evidence review at the early -- when the measure was first evaluated. So it is important that we are transparent and fair to all of our players.

There is a lot of information in there. So I don't think it is realistic for you to be able to quick look at it, you know, right now and make any decisions. Would you be willing to look at it over the next week or so, and perhaps we could have a bit of an email exchange to comment on whether you feel the new information might change your feeling and your rating on how this measure is suitable for and meets the criteria, or not?

Is that a reasonable action? Would you all be willing to do that to give Minnesota their opportunity to inform your decision?

MEMBER YEALY: My only question

would be -- I think it is a good idea to make sure we know everything, and if there was a communication barrier. The only concern I would ever have is that this would become a torrent for any criteria that wasn't viewed positively, and we would just -- It would be a never ending loop.

So it would be incumbent on you to be able to build a firewall about why it was okay once and won't be okay in a different set of circumstances. But otherwise, I think we ought to be as transparent and open-minded as possible.

MEMBER ALMENOFF: Remember, we did vote twice, because at the end you asked us again, and we did get another vote.

DR. WINKLER: Right.

MEMBER ALMENOFF: It is okay with me, as long as you let us take our clickers home. Then we can vote at home.

DR. WINKLER: Don, you don't know how much I understand your comment. I would

say that in this particular case, given it really was our responsibility to establish good communications, and we really did not meet what we consider adequate performance on that score, that we really need to be a little bit flexible for this particular thing.

CO-CHAIR WEISS: I would agree a well. I would perhaps give us a little more than a week, though, but I am a little concerned that they would put in a whole bunch of new studies at the time of the review. It feels a little different than us not digesting what they gave us and clarifying what they gave us, which we did, and then they gave us a whole bunch of new stuff, which was beyond just in time. It was after the time.

That does set a precedent that I would be very wary of that has nothing to do with communications at all.

CO-CHAIR GROSSBART: You know, Reva, one thing that I would like -- One of the reasons that I voted the way I did was

because the measures in the composite, in and of themselves, were not endorsed or fully vetted. Unlike other measures submitted by the Minnesota group, these weren't endorsed measures, and I am wondering if a process that, when we go -- composite measures need to be endorsed both for each of the components then as an overall composite, just to --

DR. WINKLER: Actually, Steve, that already exists. The issue we are having is it is very easy to put that play when you are talking about a more traditional composite where you do combine the individuals, but these all or nones are a different kind of thing, and it is causing some challenges in that original one.

So I agree. We have not totally clarified our stance on that and our approach to that, and we know we need to. You are raising a very pertinent and timely issue we need to address, but composites are a multitude of things. They are not all one

thing.

MEMBER LEVY: Yes, I was just going to say that, you know, we are calling what was behind my vote is Kevin led us through this, and as you pointed out so clearly, it is a composite measure. There are issues with composite measures with each of the components here.

Just looking through some of what you sent around, I don't know that this new information totally addresses the reasons that we voted the way we did.

In fairness, I don't know that they understand the reasons we voted the way we did, and I am not sure, without a dialogue with them or their seeing a transcript of what was said, whether this attachment is going to fully address the situation.

DR. WINKLER: Well, what we plan on doing is we will have the transcript next week, and we will use that. We will use any feedback you have and communication with that,

and we will compile all that as sort of the bundled response.

MEMBER JEWELL: I have a suggestion and a question. The suggestion, which might be a pipedream on my part, but I will suggest it anyway, is that the process is that the measure developer present a synopsis of what the measure is and does, based on whatever highlights they wish to choose. That was the part that we missed because of the communication difficulty.

Then we asked clarifying questions where they chime in as we go. That feels different to me than we have gone out and gotten a bunch of evidence to defend our position, which is the precedent setting concern that I hear being talked about.

I guess my wish of a suggestion would be to somehow reorient the exercise to the way the process would actually look. That is going to be more burdensome to them, because that would mean typing a synopsis or

something, and I realize that may not be a reasonable thing.

For the record, that feels more consistent with the process that was missed, not a -- if by evidence, they mean send us a bunch of citations or a defense.

CO-CHAIR GROSSBART: Just a point of clarification. My recollection is that they did provide an overview. They did get their couple of minutes of introductory, and then we went to a clarifying question, and there was no --

MEMBER JEWELL: For the second vote?

CO-CHAIR GROSSBART: I thought they did kick off and say, you know, their two-minute introduction, but then they couldn't respond to a clarifying question, and we kept --

MEMBER JEWELL: Oh, did you? Okay. No, no, you may be right. I don't remember.

DR. WINKLER: I think I just want to bundle all of that information up for all of us so we can be sure that we have had a fair exchange.

MEMBER JEWELL: I'm sorry, Reva. My question was: Did you say you already forwarded their response to us?

DR. WINKLER: Yes, I did last night.

MEMBER JEWELL: Well, I didn't get it.

DR. WINKLER: Okay. Sorry.

MEMBER JEWELL: That is why I am asking.

CO-CHAIR WEISS: Just a question as we think about this an respond to whatever electronic process. There are some elements of their composite that include measures that are otherwise similar measures, competing measures, in the NQF measure library.

One example would be the ratio of short-acting to controller. For us to say yes

to a composite which has an element with a ratio like that as part of it when we have an endorsed measure, how do we want to treat that sort of issue?

DR. WINKLER: All right. If we are going to be launching into a completely different type of conversation, because we didn't really talk about those things yesterday, and you want to go there, we would really have to regroup on a conference call to allow you to have that conversation.

What we are proposing to do is not to go onto new ground, but to just capture everything that was said, to be sure that on both sides it was heard.

Let's start with just initial dialogue, and see where it ends up.

MEMBER GLOMB: If I can just mention a concern that we did voice yesterday, that we did discuss that came up a couple of times with regard to a composite measure and weighting of the various parts of the -- Oh,

you did that. Okay, I'm sorry. Looking across the spectrum of things that they were requiring at all or none, there was really not an appreciation of what was more important than other things.

MEMBER LEVY: I'm sorry. And in terms of just supporting the issue of insufficient evidence, as long as I am here looking at this, one of the issues that Kevin raised is that the asthma control test is validated in a specialty clinic population and, you know, when you take that out into a primary care population, it is not apples and apples, and the three citations they give here for the asthma control test by Bob Nathan and Mike Schatz, Andy Liu, are all in allergy immunology. Those are all allergy immunology authors, and these are all -- I am familiar with the references.

This doesn't adequately overturn your point regarding validity as one of the metrics -- on the aspects of the three

composite.

DR. WINKLER: Yes, we are compounding the problem of discussing things without having everybody at the table. So --

MEMBER LEVY: All right. No, I mean, I am just making the point while we are altogether and while I am looking at it.

DR. WINKLER: The last thing that we like to do when we've got a few minutes before we are racing out the door -- Do you guys have anything you would want to say before we have an opportunity just to talk about what measures would you have liked to have seen that did not come in? Anything from you guys?

It is typical that during the course of a project, during the course of your conversations, there will be, gee, you know, here is this measure; wouldn't it be great if we had a measure that did this or did that or, you know, we don't have anything that measures thus and such.

So we do like to take the opportunity to ask you how you perceive the -- where are we lacking in measurement, so that we can provide information to the measurement development field and try to encourage development of the kinds of measures that you al feel would be particularly helpful or fill gaps or make the portfolio more robust.

I would like to point out to you that we provided two documents that the American College of Chest Physicians -- I'm trying to find my element of it. They had done an exercise on gaps for both critical care and pulmonary -- I'm trying to find it -- a very nice, kind of elegant assessment of gaps, and they do identify areas in critical care and areas in pulmonary subjects that they would suggest for -- Kenny, can you go down? This is the one from critical care.

These are the kinds of things, again, measures for sepsis, and I think we heard Dale say they are working on measures of

sepsis, which I think would be good.

Blood transfusions. We have talked about ventilator-associated pneumonia; risk adjusted ICU outcome measure. Actually NQF has endorsed a risk-adjusted outcome measure and a risk-adjusted length of stay measure for ICU from the University of California at San Francisco. So we do have that in the portfolio.

Can you scroll down? So there is therapeutic hypothermia, daily chest radiographs in the ICU patients, and then screening of ARDS.

So this was an NQF member being proactive and making suggestions about this topic area where we are lacking in some measures. I would appreciate your thoughts. We will pull up the one for pneumonia in a minute.

We have given you these documents. They are nicely done, but comments on this, and then your own thoughts about measures we

should have had.

MEMBER COHEN: Have we ever had any measures regarding instructing patients on how to use their handheld inhalers prior to discharge from hospitals, let's say, because I am always amazed with the asthma patient who has had asthma for 10 years who doesn't know how to use their Ventolin inhaler.

CO-CHAIR WEISS: That counts for COPD too.

MEMBER COHEN: Oh, yes, all lung diseases. Before they leave the hospital, they are instructed on how to use their handheld inhaler, and they demonstrate appropriate use.

CO-CHAIR WEISS: Would you say that is folded into asthma education -- comprehensive asthma education at time of discharge kind of concept? I don't know if there is a measure for that, but it would be an asthma education, not isolated just to --

MEMBER COHEN: No, no, I

understand. I meant education, discharge instructions, but specifically how to use your handheld --

CO-CHAIR WEISS: Demonstrating, yes. Good. Dave?

MEMBER RHEW: Yes. I am a firm believer in outcomes, and I know a lot of these are process measures. I would suggest across the board for any high mortality conditions that in-hospital mortality severity-adjusted be included or proposed as a quality metric across the board. Combine that with 30-day mortality, and also include 30-day readmissions.

That may be already in place, but just as a general rule, I think those tend to be ones that we should always consider as sort of the gold standard.

The additional thing is a lot of what we are talking about here is underutilization, trying to get higher numbers. But I think looking at efficiency

metrics or ways that we can identify overutilization, make care more efficient, more streamlined.

I think right now every organization across the country is faced with rising health care costs. If there are ways that we could be proactive and identify metrics that can help them in those efforts, I think that would be really helpful.

CO-CHAIR WEISS: Great. Mitchell?

MEMBER LEVY: A couple of things. I strongly support what David just said about the most frequent diagnoses in critical care, mainly sepsis and ARDS, and some risk-adjusted 30 and hospital mortality outcome measure, I think is really important.

Two, I would really encourage in particular sepsis measures, because we don't really have good sepsis measures. We have pneumonia, but as we know, sepsis is -- overall, sepsis is more common.

Then finally, palliative care

measures. There are palliative care measures in critical care out there, Judy Nelson and some others, and we don't have anything in palliative care. I know the government, obviously -- HHS wants to stay away from it, but that doesn't mean NQF has to stay away from it.

CO-CHAIR WEISS: Great. David and then David.

MEMBER YEALY: Mitch would likely guess what I was going to say. So I think early identification of sepsis, including compensated sepsis in rooms, a measure around that. Right now, all that is identified commonly is decompensated septic shock and so constructing a measurement there.

Then as I see you have here, something about the initial resuscitative aspects, and specifically not to be any one. There could be a menu of choices. I am not here to advocate for a river style approach or anything like that, but there are some basic

things that still don't happen with people with sepsis, and it is still the most morbid or actually, the most mortal condition that I admit from the emergency department, bar none.

CO-CHAIR WEISS: Great. David, then Rubin.

MEMBER STOCKWELL: I would endorse the reading of this document. It is actually really well done, and appreciate you guys including it in there.

One thing that -- It is unclear to me how much of a problem it is in adult medicine, but one thing that is increasingly becoming apparent to me is the impact of unplanned extubations in pediatrics.

These are not just tubes that come out and there is no repercussions. There are actually chest compressions and resuscitative measures that have to be undertaken, especially the smaller age of the patient.

CO-CHAIR WEISS: Thanks. Rubin?

MEMBER COHEN: Regarding the point

of palliative care, New York state now requires as of July of 2011 any patient who is only expected to have six months to live, you must document a palliative care consult in the chart. But that is because the northeast has a very high rate of patients in the last days of life, the highest in the country of getting ICU consults. Most of the patients that die in the hospital die in the ICU. So that was the response of the state.

CO-CHAIR GROSSBART: Just a high level comment. What we are not doing is developing measures that really are tackling the realities of the Affordable Care Act. So do we have any measures that would take advantage of per capita costs over per capital, you know, over an episode of a pulmonary or critical care. What is that?

DR. WINKLER: We are getting there. NQF just completed a couple of phases of a project of resource use, and it is around resource use for specific conditions. So it

is happening.

CO-CHAIR GROSSBART: David touched on it, the overutilization efficiency measures. You know, I am looking at the roster. We are an acute care system group, even though we have a fair number of community based measures in this measure set, and you know, really keeping people out of the hospital. We are all going to go out of business if we don't. So might as well get going on it.

CO-CHAIR WEISS: Dianne and Norman.

MEMBER JEWELL: There were no rehabilitation measures in this set either. For those of us that served on the cardiovascular panel last year, there were a few that were brought forward that were definitely not ready for prime time, unfortunately, but particularly for the COPD population, measures to really look at outcomes of rehabilitation in the post-acute

setting.

DR. WINKLER: Yes. If you recall, we have endorsed two measures in rehab management of COPD patients. That is quality of life, improvement in quality of life and the improvement in the walking.

MEMBER JEWELL: Oh, I did not remember that. Thank you. Okay.

DR. WINKLER: But, certainly, there may be an opportunity for more in that area.

MEMBER EDELMAN: Well, I am glad to hear the last two comments, because I was going to say all we are talking about is hospital based medicine, and for the needs of patients and for needs of resource utilization, we really have to focus on ambulatory care medicine. It is much harder to do. I admit that, and that is why we need measures.

We need very simple things. As you pointed out, we don't have an ACT for the general population. We don't have, in my

opinion, something that really works for COPD that is comparable to the ACT, and on and on and on.

There are lots of measures of simple quality of life and quality of care that could be applied in the ambulatory setting, and I guess there is no motivation for a sponsor to step up and do it. Is that the problem? We really don't have anybody who has invested in it?

DR. WINKLER: I think there are some methodologic challenges as well, but the development community each has their own sort of reason for being in that space, and so there are going to be a variety of priorities and competing priorities out there, which is why these kinds of suggestions that kind of go along with our measure sets are helpful, and we do try and point to them to really bring that issue and help foster that dialogue.

MEMBER ALMENOFF: I had to walk out of the room for a couple of minutes. Are

there any measures on ambulatory sensitive conditions, which are outpatient measures?

DR. WINKLER: We certainly have reviewed and endorsed some of the ambulatory care sensitive measures, not all of them, but some of them. Off the top of my head, I couldn't run you the list.

MEMBER ALMENOFF: I mean, there is a journal of all conditions, of all 16 of them, there are. You could do 16 individual measures. So that would be -- Are there?

DR. WINKLER: She remembers.

DR. BURSTIN: We have endorsed almost all of the AHRQ prevention quality in pairs. Yes.

MEMBER ALMENOFF: How about any -- Are there any measures on functional status?

DR. BURSTIN: Few and far between, primarily in home health where they use OASIS to get -- Actually, you guys talked about this yesterday, so some of those kinds of measures, but not very much, although we have a new

project just beginning on the methodologic issues in using patient reported outcomes in function.

MEMBER ALMENOFF: Okay.

MEMBER HAECKER: Along that line would be the patient centered medical health from NCQA and any of those measures.

DR. BURSTIN: Being evaluated as we speak.

MEMBER HAECKER: Okay, thanks.

MEMBER CANTINE: One of the performance gaps that I see pretty routinely in my pulmonary lab and getting pulmonary functions or spirometries from other centers, and we base our diagnosis of COPD on spirometry, but I have to tell you, the quality of those spirometries are -- many times they are not meeting ATS standards for reproducibility and time.

I know, in terms of pediatrics, you can't have the same set of standards, but it is something I see quite a bit. So I would

just put that out there, because we are basing the diagnosis on that, and in many cases even the physicians making the diagnosis don't understand the quality aspect of spirometry or pulmonary function.

MEMBER ALMENOFF: And it is also getting worse, because everyone can get a spirometry.

MEMBER CANTINE: Exactly.

MEMBER ALMENOFF: When we owned the labs and controlled it, it was probably at a better quality.

MEMBER CANTINE: And there is no one looking at it.

MEMBER ALMENOFF: Right.

MEMBER STEARNS: I just wanted to add quickly that we didn't see a measure that was a composite, which is really sort of an outpatient setting for consumers to be able to look at to get a sort of better picture.

We saw the Minnesota measure, which I have loud and clear from folks has

some problems, and I don't know if it can be retooled so that it becomes something that is more useful, but I do think that having that kind of measure would be useful.

DR. WINKLER: Any other thoughts? Again, this is kind of a way of bringing the discussion to a conclusion about measures, what we are measuring, what the good measures are, and what we should be measuring. Yes?

MEMBER LEVY: What is the next process? When is the next submission, etcetera?

DR. BURSTIN: Just that there may be additional information flowing to you from a couple of developers who didn't feel like they were able to give you their full information.

This is quite common in our process that, on reflection, people feel like some of the information perhaps they provided was not as clear as it could be.

So coming soon to your email box,

we may ask you to consider whether you want to revote. Again, this is something quite common. It is not always clear for developers up front exactly what you are going to ask. So they often can provide some information post hoc.

So NCQA has indicated a desire to have you take a look at one of the COPD measures you looked at, at the eleventh hour yesterday, with some additional information, and that ought to be coming to you.

DR. WINKLER: Anthony, are you there? Do we have anybody in the audience who would like to make any public comment. Operator?

OPERATOR: This is Yvonne. I have taken over for Anthony.

DR. WINKLER: Hello, Yvonne, is there anybody there who wants to talk to us?

OPERATOR: I don't show that we have anyone on the line, actually.

DR. WINKLER: All right. Thanks

very much. Anybody in the room? Our small but loyal audience, thank you all for being here. So I think we have done the public comment.

Next steps: We have talked about the additional information we are going to send out to you, and to see if there is any additional activity. But for the most part, you have made all the decisions we had hoped you would make in this two days.

What we are going to be busily doing while you recuperate is compiling a summary report of all of the work that you have done and the recommendations you are making for endorsement.

We are planning to put that out in the middle of April, sort of four weeks from now, for a 30-day public comment. We will collect the comments. It is quite typical to expect 100-200 comments from 40-50 organizations. That is not rare. So it will be important for us to carefully consider

those comments.

It really is a critique of the work you are doing on behalf of all those people out there, and so having that feedback -- they often bring new ideas, new thoughts, different ways of looking at things. They will agree with you. They will disagree with you. They will do a lot of different -- provide a lot of different kinds of feedback for us to look at.

So we will be scheduling a conference call after we have all those comments. We will collate them and organize them into a discussion agenda for a conference call, which we will set up. I think, once we get the date set, we will be setting that up as soon as possible.

If an issue arises that, with discussions with our co-chairs, we think we need to pull you together on a conference call, that is possible. I try to avoid it, because I know you are very busy. You have

already given us a lot of time, but occasionally it becomes necessary. So that is possible.

We will be keeping you informed of all the steps we are undertaking as we go through by email. We will let you know when we go out for comment.

We will keep you informed of all the different milestones, when it goes out for comment, how you can read the comments as they are coming in, which you can, if there are particular measures you particularly want to follow, as well as all those comments stay on our public website. All the timing is up there. So we will keep you posted on all of that.

As always, at any point along the way we are always available. We would love to hear from you, especially now that we have gotten to know you. We are all friends, and please don't hesitate to get in touch with us with any questions, thoughts, ideas. I mean,

I get cartoons. I get, you know, articles from the literature. I get all sorts of stuff from people. So it is actually rather fun to have enlarged my group of friends outside the Internet.

Again, we thank you. Do you have any questions for us in terms of what we are going to be doing going forward? If anybody did not get the email that has the reimbursement form, the expense reimbursement form, or you may have put it in your outbox or whatever, just email one of us, and we will make sure you get it.

MEMBER LEVY: Reva, when is the next submission period, and how does that work?

DR. BURSTIN: The next for pulmonary critical care?

MEMBER LEVY: Yes.

DR. BURSTIN: So we are in the process of completely revamping this process, just so you know. I just saw this email. So

we are now probably moving forward with a pilot over the next couple of months to actually split the endorsement process into two stages.

So that you will do stage one on measure concepts, the importance issues, get that out of the way before developers go off and test them. By the time they come through fully baked at times, they don't want to make any changes, because they are fully baked.

So we are going to pilot that, but the expectation would be, as we move to that change probably sometime in the early winter, we will then go to the process of having committees like this meet twice a year for review of either concepts or measures.

So we will be moving to having twice a year submissions to all the different topic areas. So 2013 there should definitely be an opportunity for pulmonary critical care again and, hopefully, some of those measures we talked about that maybe weren't quite ready

like the VAE measure or other ones on that list from ACCP, hopefully, will be developed and brought forward.

CO-CHAIR GROSSBART: So the term of this committee is really just through this project?

DR. BURSTIN: We have been doing a lot of sort of lean process redesign as part -- - This CDP -- The consensus process has been in place for a decade. Time to sort of move it forward and get some changes done.

So one of the things we actually had the Board approve was moving to standing committees across each of these topical areas. So we are -- Probably about half the folks would -- Everybody would get probably a two-year term, staggered. You would get to stay on for two years, probably get a chance to be two or three of these, build expertise, and also have that continuity over time with about a half-turnover every year or so.

So we are just finalizing what

that will look like. My guess is you will be invited to say whether you would like to stay on to be on the standing committee for pulmonary critical care going forward, but hopefully, you have liked this and found it interesting.

You have also been -- We have been doing a lot of work on our criteria as well. So this is one of the first groups where we have actually been going through the detailed subcriteria on importance and reliability and validity. So we feel like it gives the end user a lot more information about these measures to be able to comment more effectively about the things you had concerns about, but it is definitely still a work in progress.

DR. WINKLER: Any last thoughts or closing? Good. Maybe we don't have to run to catch whatever transportation will take you away from us, and if you are able to stay a few hours, I have seen beautiful pictures of

the cherry blossoms from those of you who went out last night. They are at peak. It is a gorgeous day. If you have the opportunity to go see them, it is not that far away. They are pretty nifty.

Again, my thanks to all of you. We will be in touch, and travel safely.

(Whereupon, the above-entitled matter went off the record at 2:13 p.m.)