NATIONAL QUALITY FORUM

+ + + + +

PEDIATRIC MEASURES STEERING COMMITTEE

+ + + + +

TUESDAY
DECEMBER 1, 2015

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:30 a.m., John Brookey and Jeffrey Susman, Co-Chairs, presiding.

PRESENT:
JOHN BROOKEY, MD, FAAP, Co-Chair
JEFFREY SUSMAN, MD, Co-Chair
LAUREN AGORATUS, MA, Family Voices NJ*
MARTHA BERGREN, DNS, RN, NCSN, APHN-BC, FNASN, FASHA, FAAN, College of Nursing, University of Illinois Chicago
JAMES BOST, MS, PhD, Children's Healthcare of Atlanta
TARA BRISTOL-ROUSE, MA, Patient and Family Centered Care Partners
KAREN DORSEY, MD, PhD, Yale University School of Medicine
JAMES DUNCAN, MD, PhD, Washington University School of Medicine
MAUREEN EDIGER, Children's Hospital Colorado
DAVID EINZIG, MD, Children's Hospital and Clinics of Minnesota
DEBORAH FATTORI, MSN, RN, PPCNP-BC, Nemours Alfred I DuPont Hospital for Children
KERRI FEI, MSN, RN, Blue Cross Blue Shield Association
JONATHAN FINKELSTEIN, MD, MPH, Boston Children's Hospital

KAREN HARPSTER, PhD, OTR/L, Cincinnati Children's Hospital Medical Center
AMY HOUTROW, MD, PhD, MPH, University of Pittsburgh, Children's Hospital of Pittsburgh
DAVID KELLER, MD, University of Colorado School of Medicine
KRAIG KNUDSEN, MD, Ohio Department of Mental Health and Addiction Services
SUSAN KONEK, MA, RD, CSP, FAND, Academy of Nutrition and Dietetics
MARLENE MILLER, MD, MSc, Johns Hopkins Children's Center at JHHS*
JILL MORROW-GORTON, MD, University of Massachusetts Medical School
VIRGINIA MOYER, MD, MPH, American Board of Pediatrics
RICARDO QUINONEZ, MD, FAAP, Children's Hospital of San Antonio
JEFF SCHIFF, MD, MBA, Minnesota Department of Human Services
KEVIN SLAVIN, MD, FAAP, Hackensack University Hospital/Joseph M. Sanzari Children's Hospital
CAROL STANLEY, MS, CPHQ, Commonwealth of Virginia, Department of Medical Assistance Services
JONATHAN THACKERAY, MA, FAAP, Ohio Department of Medicaid

NQF STAFF:

HELEN BURSTIN, MD, MPH, Chief Scientific Officer
ELISA MUNTHALI, MPH, Vice President, Quality Management
MARCIA WILSON, PhD, MBA, Senior Vice President, Quality Measurement
NADINE ALLEN, MEd, Project Manager
SEVERA CHAVEZ, Project Analyst
KAREN JOHNSON, Senior Director
ROBYN NISHIMI, PhD, Senior Consultant

SUZANNE THEBERGE, MPH, Senior Project Manager*

ALSO PRESENT:

NAOMI BARDACH, MD, University of California, San Francisco
CASEY LION, MD, MPH, Seattle Children's Research Institute
RITA MANGIONE-SMITH, MD, MPH, Seattle Children's Research Institute
GREGORY SAWICKI, MD, MPH, Boston Children's Hospital
RAMESH SACHDEVA, JD, MD, PhD, MBA, AHRQ
MARK SCHUSTER, MD, MPH, Boston Children's Hospital
SARA TOOMEY, MD, MPH, MPhil, MSc, Boston Children's Hospital
DONNA WOODS, EdM, PhD, AHRQ

*** Present by Teleconference**

TABLE OF CONTENTS

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

(8:36 a.m.)

CO-CHAIR SUSMAN: Well, good morning. I'm Jeff Susman one of the co-chairs.

CO-CHAIR BROOKEY: And I'm John Brookey, the other co-chair. So, good morning, everyone.

CO-CHAIR SUSMAN: Just want to welcome everybody on this rainy day, but just think it could be snow, so this is the better part of evil. It is really a pleasure to see such a great group around the table.

We've got a lot of important work to do. And fortunately, the NQF staff is always has made this just a very easy task that can move us through the work as quickly as possible.

John and I are going to sort of break up our duties. I'm going to start off this morning, but let me see if John wants to add any words of welcome.

CO-CHAIR BROOKEY: No. Thank you very much. And we're going to move forward. Marcia

is here to do the introductions and disclosure of interest.

DR. WILSON: Good morning, everyone. My name is Marcia Wilson. I'm Senior Vice President of Quality Measurement here at NQF. And I'm joined by a number of my colleagues, and we'll make introductions of the NQF staff in just a moment.

But the first order of business for us is the disclosure of interest. And typically, Ann Hammersmith, our in-house counsel, would be here to do this, but I'm going to take care of this today. So I certainly don't have the script memorized as Ann does, but we shall get this job done.

So you did receive a disclosure of interest form when you were seated on the Committee, where we ask you a number of questions about your activities. But today we're going to ask you to orally disclose any information that you believe is relevant to the work of this Committee.

So we're specifically interested in a disclosure of interest about work directly related to what the Committee is going to do. So please do not feel you need to summarize your resume. We don't need to do that today. But we're particularly interested in grants or research or consulting, again, only if it relates to the work before the Committee.

And also, I would note, it doesn't need to be paid work. You may have served on a committee or done volunteer work that may be relevant and you may disclose that as well.

One reminder. You do sit on this Committee as an individual. You don't represent the interest of your employer or the person who may have nominated you.

Now, just because you disclose that does not mean you have a conflict of interest. We do oral disclosures in the spirit of transparency and openness, so we're going to do this as a way to disclose and also to introduce yourself, so we're going to start with the

Committee members.

And I would ask you to state your name, the organization you're with, and if you have anything to disclose.

And Jeff, if we might start with you?

CO-CHAIR SUSMAN: Sure. So, I'm Jeff Susman. I'm at Northeast Ohio Medical University. I serve as the Dean of College of Medicine. And I have no disclosures to make.

CO-CHAIR BROOKEY: Good morning. John Brookey, Kaiser Permanente Southern California. And I have no disclosures.

DR. NISHIMI: Robyn Nishimi. I'm a senior consultant to the National Quality Forum, and prior to that, I was the chief operating officer. So, thank you all for your efforts today and for today.

MEMBER HARPSTER: I'm Karen Harpster. I'm from Cincinnati Children's. I'm a researcher and occupational therapist there. And I have no disclosures.

MEMBER SLAVIN: I'm Kevin Slavin. I'm

at the Sanzari Children's Hospital at Hackensack University Medical Center in Northern New Jersey. I'm also with the Council of Children's Hospitals for the State of New Jersey. And I have no disclosures.

MEMBER FATTORI: Good morning. I'm Debbie Fattori. I work at the DuPont Hospital for Children where I serve as the Director of Advanced Practice and Ambulatory Nursing. And I have no disclosures.

MEMBER FEI: Hi. Good morning. My name is Kerri Fei. I work Blue Cross Blue Shield Association in Chicago, Illinois. Previous to working there, I was at the American Medical Association as a measure developer for the Physician Consortium for Performance Improvement.

However, last week, I crossed the five-year mark, so I haven't developed measures in five years. None of the measures that we will be discussing today are their measures, so thanks.

MEMBER QUINONEZ: Hi. I'm Ricardo

Quinonez from the Children's Hospital San Antonio where I am the chief of the Division of Hospital Medicine. And I have no disclosures.

MEMBER BERGREN: Hi. I'm Martha Bergren. I'm from the University of Illinois, Chicago College of Nursing. And I have no disclosures.

MEMBER KELLER: David Keller. I'm the vice chair of Clinical Affairs and Clinical Transformation at the UC-Denver Department of Pediatrics in the Children's Hospital Colorado.

And I am -- next week will be joining the Data Safety and Monitoring Board of Dr. Beinman-Smith's organization out in San Francisco. So I was not a prior conflict, but I -- and I have no idea what she's going to have us do, so I actually don't know if we will be involved with the use of her measure, because our orientation called to tell us what we're doing is next week. But I wanted to put that out there as a potential conflict in the future.

MEMBER MORROW-GORTON: Jill Morrow.

I'm a Senior Medical Director at Mass Health, which is the Massachusetts Medicaid Program, and I'm a developmental/behavioral pediatrician. I work for the University of Massachusetts Medical School. And I have no disclosures.

MEMBER EINZIG: I'm David Einzig from Children's Minnesota. I'm a child psychiatrist, also pediatrician by training, current president of Minnesota Society of Child and Adolescent Psychiatry. And I have no disclosures.

MEMBER THACKERAY: Good morning. My name is Jonathan Thackeray. I'm a child abuse pediatrician at Nationwide Children's in Columbus, Ohio. I'm also the assistant medical director for the Ohio Department of Medicaid. And I have no disclosures.

MEMBER DORSEY: Good morning. I'm Karen Dorsey, and I'm the director of the Division of Reevaluation and Rulemaking at the Center for Outcomes Research and Evaluation at Yale. And we got measures for the other end of the spectrum for Medicare beneficiaries. And I

have no disclosures.

MEMBER BOST: Good morning. I'm Jim Bost, Director of the Outcome Center at Children's Healthcare of Atlanta. And I have no disclosures.

MEMBER BRISTOL-ROUSE: Hi. I'm Tara Bristol-Rouse with Patient and Family Centered Care Partners. I'm a family stakeholder. And I have nothing to disclose.

MEMBER KNUDSEN: I'm Kraig Knudsen. I'm with the Ohio Department of Mental Health and Addiction Services. I'm the Chief of the Office of Research and Evaluation there. And I have no disclosures.

MEMBER HOUTROW: Hello. I'm Amy Houtrow. I'm a pediatric rehab medicine physician at the University of Pittsburgh Children's Hospital. And I have no disclosures.

MEMBER FINKELSTEIN: Good morning. I'm Jon Finkelstein. I'm a general pediatrician and the Vice Chair for Quality and Outcomes in the Department of Pediatrics at Boston Children's

Hospital.

I do have one disclosure. I'm a faculty member in Boston Children's Center of Excellence for Pediatric Quality Measurement, which has submitted measures for this call, and I am one of the co-leads in development of measure 2797, and I'll be recusing myself during discussion of that measure.

MEMBER SCHIFF: Hi. My name is Jeff Schiff. I'm the medical director at the Minnesota Medicaid Program at the Minnesota Department of Human Services, and a pediatric ER physician, and the -- I'm the immediate past chair of the, of the Medicaid Medical Director National Network.

I do disclose that I'm a co-investigator with Dr. Rita Mangione-Smith on the Seattle Children's Group, so I'll be recusing myself for those measures.

MEMBER STANLEY: Good morning. I'm Carol Stanley with Virginia Medicaid. I'm the Quality Improvement Manager for Medicaid Managed

Care and CHIP. No disclosures.

MEMBER EDIGER: Morning. My name is Maureen Ediger. And I'm here because of the volunteer role I have as an advocate on the Quality and Safety Committee of the Board for Children's Hospital of Colorado. I have four children, and you'll probably hear way more about them than you'd like to, but that's why I'm here. And, other than that, I have no disclosures.

MEMBER KONEK: I'm Susan Konek. I -- until last week -- last month was the Director of Clinical Nutrition at the Children's Hospital of Philadelphia. I'm a registered dietitian, and I retired. And I'm going to be joining Cincinnati Children's January 4 in Clinical Management again working with the program there. I have nothing to disclose.

DR. WILSON: Thank you. And I believe we have two Committee members on the phone this morning.

Lauren, are you available to do a disclosure?

MEMBER AGORATUS: Yes. Hi, Lauren Agoratus. I'm the State Coordinator for Family Voices New Jersey. I'm also a guest on the New Jersey American Academy of Pediatrics, Council for Children with Disabilities, a family stakeholder. And I have no disclosures.

DR. WILSON: Thank you, Lauren.

And, Marlene, are you with us on the phone as well this morning?

MEMBER MILLER: Yes. This is Marlene Miller. I am chief quality officer for Pediatrics at Johns Hopkins Medicine. And I have no disclosures.

DR. WILSON: Thank you very much.

And I'll just mention that we have two additional Committee members who will be joining us tomorrow, so we'll do their disclosures then.

And just a couple of remaining comments. First of all, at any time during the meeting, if you think you have a conflict, you can speak up in real time, you can approach the co-chairs, you can approach any of the NQF staff,

and we'll introduce ourselves in a minute.

Also, if you think someone else has a conflict of interest, please don't hesitate to speak up. We don't want you to sit in silence and wonder. It is better to ask the question and have us resolve the issue as we are moving through these different measures.

So, at this point, do you have any questions on the disclosures or any other information?

(No audible response.)

DR. WILSON: Okay. At this time, I think, I'll turn it over to the NQF staff to introduce themselves.

And, Helen, if you'd like to start?

DR. BURSTIN: My pleasure. Hi. I'm Helen Burstin. I'm the chief scientific officer here at NQF. Nice to see so many familiar faces around the table. Always nice when people want to come back for another round.

I think this is probably Jeff's third or fourth round at this rodeo, but really

appreciate all your help. We know this is a lot of work, but really important work, and we're delighted to have you here.

MS. MUNTHALI: Hello. My name is Elisa Munthali. I'm Vice President for Quality Measurement at NQF. Welcome.

MS. CHAVEZ: Good morning. Severa Chavez, and I'm the analyst for this project. Welcome.

MS. ALLEN: Hi. I'm Nadine Allen. I'm the project manager for this project.

DR. WILSON: Suzanne.

MS. THEBERGE: Good morning, everyone. This is Suzanne Theberge, the senior project manager on the team. I apologize that I'm not there in person with you all today, very sorry not to be there, but I had a family emergency that prevented me from attending.

DR. WILSON: Okay. Before we get started, and I turn it over to the co-chairs, Nadine has a few housekeeping things to go through.

MS. ALLEN: So the first most important thing, the restroom. The restrooms are located out here on the right. There should be someone that's sitting at the front desk. They can assist you if you need anything.

If you need to make a phone call, there's a quiet area that you can also use. You can also ask that person that's at the front desk, her name is Jennifer Green, and she will be able to assist you.

We have three dedicated breaks today. One at 10:30, lunch will be served, provided by NQF, at 12:15, we also have another break at 3:30.

Laptops and cell phones. We have Wi-Fi. The username is guest, and the password is NQFguest. Please mute your cell phones during the call. Please do not put us on hold. We tend to hear music during the Committee discussion.

Also, some additional items. We have -- if you're -- if you need to speak, please use your tent cards. Once you raise your tent cards

in the upright position, the chairs will call on you to speak at that time.

Also, we have dinner tonight at 6:30 at Mio, a contemporary Latin-American restaurant, and that's around the corner from us.

CO-CHAIR SUSMAN: All right. Well, again, welcome, everybody. We're going to do this first measure with a little bit more commentary, perhaps, than we would initially think, but only to get us into a good pattern of getting this off to a good start.

It really is a very stylized process. There's a lot of rules of the road, if you will, and the NQF staff are tremendous about making sure we sort of stay in the roadway.

If there is an opportunity to discuss an issue that becomes particularly thorny, at some point, as co-chairs, we may step in and say, "Okay, we've heard -- anybody got additional new ideas here, we want to keep the process going," because as you see, we have quite a number of measures to get through.

The good news is if I look at the agenda, I think, we've saved at least 20 minutes already, so we better stop while we're ahead. I think, there are flights out this afternoon.

At each of the measures, we're going to first allow the measure developer to provide a brief, meaning three to five-minute overview. And I'd ask for this first measure, which is the Pediatric Psychosis: Timely Inpatient Psychiatric Consultation, Number 2805, for those of you who are following your NQF bibles here.

If the -- first of all, any disclosures to make?

(No audible response.)

CO-CHAIR SUSMAN: Hearing none -- oh, yes, okay, please. Thank you.

MEMBER SCHIFF: I'll recuse myself.

CO-CHAIR SUSMAN: Thank you.

I'd ask if the developer for this measure, Rita Mangione-Smith or Naomi Bardach, are available.

CO-CHAIR BROOKEY: Should we briefly

go over the elements that we'll be voting on, so we know?

CO-CHAIR SUSMAN: Yes, that'd be great.

CO-CHAIR BROOKEY: Okay.

CO-CHAIR SUSMAN: While they're getting set up, we have some entertainment here. And this is just going to briefly review the elements that we have and how we're going to go about on this voting with your nice little clickers, which they'll describe when we get to that point.

DR. NISHIMI: So the developers will introduce their measures in two to three minutes, and then we will begin walking through the evaluation. Each measure has a lead discussant that we're asking to introduce the measure, discuss the strengths, and the weaknesses first on the evidence because that's the first thing that's up, and then we'll vote on evidence.

Evidences must pass. If it fails on evidence, then we don't discuss anything further

about the measure.

Then we move to performance gap. And again, the lead discussants will introduce the measure, and we ask the Committee to join in because all of you, of course, we're asked to look at every measure, and then we'll vote on performance gap.

Again, must pass. If it doesn't pass gap, then we don't discuss the measure any further.

And we proceed that way through both reliability and validity. Both of those are must pass.

Usability and use, feasibility are not must pass, but we still vote after we discuss those criteria. And then there's a final vote on the overall suitability for endorsement.

So you have clickers, and when the time comes, we'll do a test vote, and then we'll vote for real.

Go to the next slide. I just want to say a little bit about how the voting is

tabulated. It's automatically tabulated, but to be recommended for any of the single elements, greater than 60 percent of the Committee must approve it.

If it's between 40 and 60 percent vote on that criterion, or on the overall suitability for endorsement, NQF refers to that as consensus not reached. So we proceed with the evaluation, we proceed with the discussion, but there's a designation consensus not reached.

And what happens there is the measure goes out for public comment, we'll get comments or not, the developer may bring forward additional information, and then when you have the follow-up call after the comment period closes, we would ask you to then vote, re-vote on that, and you may still not reach consensus, but at least you will have, hopefully, received additional information.

If there is less than 40 percent, so 39.9 or lower, then the criterion fails or the overall measure fails.

Any questions about how that goes down?

(No audible response.)

DR. NISHIMI: We will alter things slightly when we get to the FECC measures, the FECC measures, because we're going to take those up in, all ten at once on evidence, all ten at once at its gap, etc., because a lot of the discussion will be the same. There was overlapping discussion. So when we get to that, we'll make it clear how you're voting there.

Okay. Great.

CO-CHAIR SUSMAN: All right. Now, I hope everybody is ready to listen to our measure developers. Thank you for attending today.

DR. BARDACH: Thank you very much. I'm Naomi Bardach. I'm at UCSF, the University of California San Francisco, and I'm part of the Seattle Center of Excellence this morning.

You'll be discussing three measures today that we're presenting on pediatric mental health in the ED or inpatient setting. The

measures are submitted by the Center of Excellence on Quality of Care for Children with Complex Needs, which is housed at Seattle Children's Research Institute.

The process for developing these measures was the same, so I'm only going to review it once. The Center's Mental Health Working Group after developing a conceptual framework for the measures determined the most common reasons for pediatric mental health presentations to the ED or inpatient setting nationally.

And then we performed literature reviews for the top conditions. The lit reviews inform the content of the measures. We then presented the measures to a multi-stakeholder Delphi panel, which included caregivers of children with mental health problems, also mental health and pediatric clinical specialists, and Medicaid health plan representatives.

Measures that met Delphi panel faced validity criteria were then operationalized and

field tested. We submitted for endorsement only those measures that performed well throughout these processes.

The measure currently under discussions, the medical record-based measure of the percentage of patients. It's called, "Pediatric Psychosis, 2805," just to orient you guys, "Timely Inpatient Psychiatric Consultation." It's a medical record-based measure of the percentage of patients age 5 through 19 who were admitted to the hospital for psychosis who had a psychiatric consult, in person or by telepsychiatry, within 24 hours of their admission.

There's just a few specific responses to questions that came up in the workgroup call that we wanted to review. Overall, we would like to acknowledge the dearth of evidence regarding the delivery of mental health services to children.

There are a couple of very important points to be made. First of all, the populations

are often hard to study since they and their families are extremely vulnerable during acute ED and inpatient episodes, and the presentation is relatively rare compared to something, such as adult acute MI.

If the Committee supports waiting for more robust evidence in order to endorse measures for accountability, it may be quite a long time before we are able to measure performance for these important groups of children.

We know from work also done in preparation of measure development that there are very few existing pediatric mental health measures despite nine percent of pediatric hospitalizations nationally for ages three and up being for mental health conditions.

Given the dearth of data and the numerous barriers to creating strong evidence to support quality measure development for the children, we believe these measures would most appropriately be considered under the evidence criteria of insufficient with exception.

Just one other thing specifically for this measure, 2805, there is one update, a study published earlier this year of pediatric patients in a tertiary care hospital who received a psychiatric consultation. The study found that a ten percent decrease in time to consultation was associated with an eight percent shorter length of stay, which was statistically significant, and a similar magnitude decrease in cost.

This study just adds to our measure validation work that found a decrease length of stay for those patients who passed the measure.

Given a limited time for this instruction, clock is ticking, I will only comment that we do have further information regarding three specific additional workgroup concerns, which were first of all the age range for the measure, secondly, a concern regarding allowing a pass for consultation performed up to 48 hours if the provider documented a justification in the chart, and lastly, there was a concern about a somehow high pass rate in our

field testing.

I'm happy to provide additional information on any of those issues when asked before going to the map.

CO-CHAIR SUSMAN: I've just been told that Virginia Moyer just arrived, so we need to go through a disclosure.

Maybe you'd like to introduce yourself first. And please use your microphone.

DR. WILSON: It's the button on the right.

MEMBER MOYER: I pushed it. It just didn't do anything.

DR. WILSON: It's on now.

MEMBER MOYER: I'm Ginny Moyer. I'm the vice president for MOC and Quality at the American Board of Pediatrics. And for today's meeting, I am going to recuse myself from four measures because of involvement with the committees that produced those measures, 2799, 2800, 2801, and 2803.

DR. WILSON: Great. Thank you very

much.

CO-CHAIR SUSMAN: Thank you for that brief interruption.

Anything to add?

(No audible response.)

CO-CHAIR SUSMAN: Okay, great. Well, that was a good example of keeping it short, sweet, and to the point. I appreciate it.

So let's go ahead and we'll first have a little bit of description, so the lead discussant will go through, again, the measure. Try to give an overview of what the analysis is, the level of analysis, what's the denominator, numerator, and then to launch off into the evidence.

So it looks like that's Dr. Keller, Dr. Bristol-Rouse, and Dr. Moyer.

Who's the lead discussant on this one?

MEMBER KELLER: Well, if that means whose name was first, I think, that was mine.

CO-CHAIR SUSMAN: Okay. Well, why don't you start us off?

MEMBER KELLER: That actually wasn't entirely clear to me, and this is my first meeting, so I was being a little hesitant.

The -- so the measure is a pretty straightforward one. As we've already heard, it was the percentage of children and adolescents, age greater than or equal five and less than or equal to 19, admitted to the hospital with psychotic symptoms who had a psychiatric consult in person or by telepsychiatry within 24 hours of admission.

I think, during our discussion, the issue of how to identify psychotic symptoms versus diagnosis of psychosis was raised, and -- but during the test phase and validation, they -- measure developers had addressed that as, in the process of being able to obtain that data out of the electronic health record in a systematic fashion from the three institutions they worked with for that.

The other issue was about the -- that was raised during our discussion was really how

to define a psychiatric consultation and whether there were times or reasons where the 24-hour limit would not be met just because of arbitrary workflow processes within the hospital that the evidence for picking 24 hours is the standard was not very clear to us in going through, in going through that measure.

The measure was based on guidelines that have been developed from what evidence there is. As we heard from the introduction, there is not a lot of evidence to go on here, but this is -- so a lot of what was done is based on consensus of experts given that, I think, everyone felt it was a reasonable thing to look at.

The -- per the algorithm for evidence, we all thought that the, that this probably did qualify, as was mentioned in the introduction, as an exception, where we didn't have a lot of evidence but that we thought this was important measure and we were hard-pressed to come up with something that was a negative that would come out

of using this as a standard that people would work towards.

The -- there was concerns raised because the three hospitals that were used are fairly advanced hospitals with well-developed QI processes and electronic health records that are -- we were wondering how well this measure would be applied in other kinds of hospitals, particularly in general hospitals that were not pediatric-specific.

And the -- and the concern about having to do -- that the chart auditing process would make it infeasible to actually bring this measure out into wider distribution. We were also concerned that at, again, hospitals that don't have large numbers of children that the N for the denominator would be relatively small, and therefore, would be subject to a lot of variability.

As far as usability and use, this is a new measure, and so is not currently in use anywhere, so we had no real way to judge how well

it's being used because it hasn't been used yet, but there were no concerns raised about unattended consequences.

Anything else I'm forgetting from that conversation? I'm looking at my fellow leads.

MEMBER MOYER: The only issue is that I'm not remembering well information from the developer about the definition of the denominator. As it reads, it's those who are discharged, with a discharge diagnosis of psychosis. It's not those who are admitted with psychotic symptoms.

MEMBER KELLER: Thank you. Sorry.

MEMBER MOYER: And so, I'm just interested from the --

MEMBER KELLER: I wasn't clear on that. There was -- there was some concern about that definition expressed.

DR. BARDACH: So to clarify, that's correct. It's based on the discharge diagnosis, and the denominator population eligible patients are identified using the administrative data, so,

yes, it would be a discharge diagnosis rather than an admission.

MEMBER MOYER: So was any work done to determine whether there were patients who were admitted with psychotic symptoms, but not discharged with a diagnosis of psychosis? And based on our group discussion, that was one of the primary concerns that, particularly younger kids with psychotic symptoms, usually are actually -- do not have psychosis as a diagnosis at the time they go home.

DR. BARDACH: So in the development work because we used the ICD-9 codes administrative of record to identify the eligible population, there was no assessment of kids who came in with the symptoms of psychosis who did not have a discharge diagnosis.

CO-CHAIR SUSMAN: Okay. So one of the first questions we're going to have in each of the measures we consider is whether this is a process or outcome measure. And I assume that this is a process measure that there is no link

directly to the outcome.

And that also brings us to the question of the evidence here as the measure developers noted. There really isn't a body of evidence, or certainly it's a very early body of evidence linking this to some patient-oriented outcome that would be of interest.

If you -- remember this wonderful table? It gives us some guidance around the ratings of evidence. And in this case, the evidence is lacking. You know, there isn't even moderate evidence as defined here, which, does the grade of evidence indicate high quality evidence, is it high grade, high quality? Answer, no, no grading of evidence, no summary, not graded or strong recommendation.

Is there empirical evidence without systematic review and grading of the evidence? That's box seven on our algorithm and going down through that.

I think, at best, does the empirical evidence that it summarize include all studies in

the body of evidence? Well, there's really not much evidence to even include here.

So we either are getting too low, or on your page here, since there wasn't any systematic review and there's not much evidence to review, we're going to get to a series of conditions that have to be met for an exception, which is, I think, what you all have been recommending as the primary reviewers.

And in box ten, it says, "Are there, or could there be, performance measures of a related health outcome, or evidence-based intermediate clinical processor outcome?" And if the answer is no, to see is there evidence of systematic assessment of expert opinion and consensus recommendation, benefits of what is being measures outright risk. And in this case, there was a formal Delphi method that came up with a consensus around this is a relatively strong measure.

And then it goes down to 12, does the Steering Committee agree that it's okay or

beneficial to hold providers accountable? So remember, this is not just for performance improvement, but it's accountability, and we don't control how these measures are being used.

So at least in my mind when we talk about accountability of providers, we want to make sure there's a certain amount of rigor, or if there isn't and we're recommending an exception that we do that knowing that people may be held accountable in some health plans or other measurement environments. And it goes on to describe then how that might be in the absence of empirical evidence.

So again, I invite comments from those of you who more closely looked at this, but at least that's how I make the evidence rating of this boiling down to probably an insufficient evidence with an exception.

Please.

MEMBER MORROW-GORTON: As part of our discussion, I think, we -- it was not entirely clear that the population of kids that you were

looking at were only those kids with a diagnosis of a psychotic disorder. And given that, there are -- is the potential for, and we have no idea what the volume of those children are, for children to have presented with psychotic symptoms in the ED and for it to have been an entirely different reason, and asking hospitals and professionals to do child psych consultations for, for all of those children when we don't know whether there's any benefit or harm to them by having a child psych evaluation at presentation.

I think, it's a little concerning in that we don't know what the volume of that group of children is. We don't know what the age distribution is. The younger children are much less likely to have psychotic disorders, and in many of these situations, child psych consult may not be helpful.

CO-CHAIR SUSMAN: Kevin.

MEMBER SLAVIN: One of my questions around this really has to do with the accountability part, and that is, especially when

you start getting into the younger children, the dearth of available psychiatric consultations, whether in-person or by telemedicine, especially as you start getting into more remote areas or areas where there are not strong tertiary centers, and so to hold accountable to a metric when the services are not accessible raises some concern in my mind.

CO-CHAIR SUSMAN: Other questions or comments?

Yes, Ricardo.

DR. NISHIMI: Just when you -- when you want to talk --

MEMBER QUINONEZ: Oh, I'm sorry. Okay. Right, I got it.

CO-CHAIR SUSMAN: Go for it.

MEMBER QUINONEZ: Yes. So I -- just to follow on that last point. I think, sometimes the, the reasons for measures is to, is to -- it's to encourage those processes being developed. I mean, I think, we would all agree that the lack of psychiatric consultation in

different areas of the country is a problem, and until somebody is held accountable for that, it's not going to change.

CO-CHAIR SUSMAN: There is a, if you will, put a measure out there, it moves out systems and the dynamics of our healthcare organizations for sure.

Other comments or questions?

MEMBER MILLER: This is Marlene.

CO-CHAIR SUSMAN: Go ahead.

MEMBER MILLER: Hello? Oh, yes, sorry. You know, I just think, on that last comment, I think, the goal of measure should be that there is evidence that may matter more so than using a measure to move policy or change systems because that accountability piece has important ramifications of where resources -- a lot of measurement resources get put, so I do think -- on many, many of these measures, we straddle what we think we should want in the ideal world versus what -- is there actually evidence that the thing being measured matters.

And I believe our focus here has to be a little bit more heavy on where the evidence is the actual measure and effects matter.

CO-CHAIR SUSMAN: Thank you.

And I can't see whose it is.

MEMBER HOUTROW: It's Amy Houtrow. Hi.

CO-CHAIR SUSMAN: Okay, thanks.

MEMBER HOUTROW: I have two points. The first point is that in many adult hospitals in which teenagers would present, they might not never see a psychiatrist either my telemedicine or otherwise. They might see a social worker and be on a medical hold until they're placed, and that would be at a completely appropriate pathway to getting treatment.

The second point is there might be many times when psychosis would show up on a discharge diagnosis that's unrelated to the, the reason the child was admitted. So for example, if you're in a car crash and you have a severe traumatic injury, you might emerge from your coma

into a state of psychosis for which you could then be discharged to a rehab facility, which would be a completely appropriate pathway for your treatment that would be completely outside of a psychiatric pathway, and so, therefore, you could end up with a psychosis discharge diagnosis, but being the scenario, a clinical scenario, which is kind of underlying not a psychiatric problem.

CO-CHAIR SUSMAN: Okay. Very good.

Why don't we just go down the line here?

MEMBER BRISTOL-ROUSE: I know the developers mentioned that part of the group that -- you put together the measures included family members and so on. I'm curious since there is this lack of evidence, you know, how many family members were part of this and what were their kind of thoughts on the importance of this since some of our discussion here today is whether it's going to matter to the children themselves.

CO-CHAIR SUSMAN: So what I'd ask we

do is if there are questions for the developer, why don't we try to put those all out there, and then we'll get a very brief comment answers if you have them. And then I'm just going to go down this side beginning with Kerri.

I'll get back to you, John.

MEMBER FEI: Thinking of this from a health plan perspective regarding the evidence, if me at a health plan is going to hold providers accountable for this measure, we have to be sure that it's based on pretty good evidence. And I know that is a struggle from developing pediatric performance measures, I know that is a struggle, and then sometimes we have to put measures out there.

However, if I'm going to either in sense or take away money from providers based on this measure, it needs to be based on strong evidence.

CO-CHAIR SUSMAN: And just to clarify the task at hand NQF staff, please keep me straight, we're going to be voting on the level

of evidence. If it turns out to be insufficient, then we'd have to separately vote whether there's an exception or not, so as you're thinking about the task at hand. That correct?

MS. MUNTHALI: Yes.

CO-CHAIR SUSMAN: Okay. So let's go down, and I can't see whose sign that is if it's -- David.

MS. MUNTHALI: You need to hit your mic.

MEMBER KELLER: Oh, sorry.

One of the things that's interesting to me is that for the three sites that, for this, which this measure was tested that the rates of positive, the rates of compliance were actually pretty high at all three sites. We didn't find a lot of, as much evidence and variation as I would have thought.

And that sort of speaks to your point, I think, a little, a little bit. What I'm wondering about is whether at smaller hospitals -- whether at smaller hospitals, you would see

more variation than you did at these large hospitals.

And certainly, where we too say going ahead here, I think, we need to look very carefully at the implications for this for general hospitals and hospitals without a pediatric focus because, I think, that you might find something different.

But I did think there was evidence that this is standard of care for at least at large children's hospitals, and that's something that we should keep in mind as we're thinking about whether or not that's generalizably possible across the nation.

CO-CHAIR SUSMAN: Okay. David.

MEMBER EINZIG: So I just want to preface this with, I really like the idea of the measure, but in terms of evidence, the question of should it be a psychiatrist who does the consult and is the evidence to say that a psychiatrist does a better job than a psychologist in terms of a psychologic

assessment.

I think, in terms of -- so the reason why a person will be held in a pediatric hospital is primarily for safety while they're waiting to be housed to get to a more appropriate setting, a psychiatric unit or whatever the next appropriate level of care is.

So this might be jumping the gun into feasibility, but just in terms of -- keep it -- focus on evidence -- is there -- is there evidence to say that a psychiatrist does provide a better, do a better job than a psychologist or other provider in terms of doing that assessment?

CO-CHAIR SUSMAN: I'm going to go ahead and take Jon and Virginia, and then we'll get some comment from our developers. And Jon had a question of the developer, so -- Jon, over here.

MEMBER FINKELSTEIN: So I too am -- I'm positively disposed to the measure. I think, most children come in with psychotic symptoms in the ED need a pathway to quick mental health

consultation and care.

Some of the discussion to me speaks to what would be exclusions where the measure in particular cases didn't quite make sense. And we'll be considering other measures based completely on claims where those exclusions will be opaque to us in the measurement process, but in this case where you're actually looking in the chart, I wonder if the developers thought about exclusions that you could also get from the chart, the rare event where psychosis, these symptoms weren't present on admission, but developed in the course of a hospitalization, Amy's case, or other, other reasons why the measure wouldn't make sense.

If it's a very large N measure, these very rare exclusions might not matter, but especially in smaller institutions where the N itself might be very small, one or two cases where there actually was a rational reason for delay or a different pathway might make a big difference.

CO-CHAIR SUSMAN: Thanks. Good points.

Virginia.

MEMBER MOYER: So I have actually another question for the, for the developers. The paper that was distributed after the pre-meeting does show an association, I'm not clear whether it's causal or not, but it does show an association between the length of hospital stay and the timing of the initial psychiatric consult.

I would just like to hear a little bit more about the -- what the expected benefit of this measure is, other than shortening hospital stay where we really don't have particularly strong evidence. We have a piece of evidence, but we don't have a lot of evidence beyond shortening of hospital stay.

What are the other benefits that we think would accrue for an earlier psych consult? And, I also just want to comment that I also have a concern that it specified that it's a

psychiatric consult when perhaps a mental health professional consultation would be what you are really looking for.

My third concern is, I think, I've already expressed, is that the kids I'm worried about -- I like the idea of the measure also because the kids I'm worried about are the ones who need early evaluation, and that's not what we're finding. The denominator is not those kids. The denominator is the kids who've had a later diagnosis.

CO-CHAIR SUSMAN: Okay.

Maureen, do you have a comment?

MEMBER MILLER: Marlene you mean?

CO-CHAIR SUSMAN: Yes.

MEMBER MILLER: Okay.

CO-CHAIR SUSMAN: I'm sorry.

MEMBER MILLER: It's okay. I get called that all the time; that's why I still reply.

CO-CHAIR SUSMAN: Yes. I need to see your name tag. That's the problem.

MEMBER MILLER: No problem.

So, yes, I wanted to go back to that comment about the three half of this was tested and had fairly high, if not standard of care already performance on this, and, I think that's a very valid point. We have a history in some of the pediatric measures, and for those of you that know -- for example -- the Joint Commission as a measures, that was exactly the conversation, the measures were very high in children's hospitals, the performance was high already, but maybe at small hospitals it wouldn't be.

And we went down this ten-year path of many, many resources being poured into measuring this, which finally -- finally this year -- is being retired by the Joint Commission because it had no evidence of any impact, and so I do -- you know, I think we have to think about that.

You know, it doesn't mean that this measure is not acceptable, but it would really need to be proven to me, I think, that there is a performance gap because the data that was given

showed very, very high performance at children's hospitals. We don't want to make measures where everyone's, you know, 95 percent already.

CO-CHAIR SUSMAN: So, I've heard from the group some concerns about the tie to the evidence that we shouldn't be holding people accountable for a performance measure that has a dearth of evidence. I've heard some questions about what outcomes are intended to be improved other than length of stay, some technical questions about specification of the measure -- for example, why not mental health provider as opposed to a psychiatrist or a psychologist being an appropriate way.

And, then I'm going to put on the parking lot for just a moment the questions about gap, which we get to vote on, I believe, separately here. So, let me turn it over to our measure developers to briefly respond to any --

DR. MANGIONE-SMITH: Thanks for all your thoughtful analysis of the measure, and I'll try to make sure that we address all of the

different concerns that have been raised.

The first one I'll take up is the issue of a young child presenting in the ED with psychotic symptoms who is not actually a child who has psychosis and inappropriately applying this measure to that sort of child.

So, when we went from the language of the draft's quality measure to operationalization of that measure, the choice was made to identify these cases using administrative data with a list of ICD-9 codes that were vetted by our mental health working group. So, we are talking about retrospectively looking at cases where there was a clear diagnosis of psychosis.

So, the measure is not -- the denominator does not include children who just present to the ED with psychotic symptoms and don't end up leaving an inpatient stay with a diagnosis of psychosis.

In terms of the concern about, is that a valid diagnosis? The first step in our medical record abstraction tool that the abstractors are

asked to do is to verify that, in fact, this child was diagnosed with psychosis in this admission, but either by their discharge summary or any other parts of the chart that the abstractor has access to.

The next thing I wanted to address that's come up multiple times is the need for this to be done by a psychiatrist. And I'd like Naomi to read to you the directions that are given to the abstractor about what counts as a psychiatric consult.

DR. BARDACH: So, the instructions are, "The consult may be in person or by telemedicine. The consult must have been done by a psychiatrist or a PhD psychologist. If the consult was done by a clinician extender, nurse practitioner, advanced practice nurse, physician assistant, licensed social worker, or licensed counselor, this is acceptable as long as the assessment is co-signed by a psychiatrist."

DR. MANGIONE-SMITH: Okay. So, it's not just a psychiatrist. It can be a PhD

psychologist that doesn't even require a co-signature, but if it's an extender, we do require that there be a signature by a psychiatrist, a co-signature.

In terms of what was our family representation and development of the measures in considering the measures, Carolyn Allshouse -- who is the lead person for Family Voices of Minnesota -- is a member of our Center. She put together a panel of ten parents who advised on the development and all stages of our measures.

They had review of our lit reviews, they reviewed all the draft measures, they gave us feedback, so even before the measures hit our Delphi panel, we had a really rich amount of parent input throughout the process. And, then on the Delphi panel, Lynn Pedraza -- of Family Voices -- was a member of our Delphi panel that assessed the final mental health measures.

An issue that was brought up with regards to this only being tested in three children's hospitals, we did have two community

hospitals in Minnesota, which we would have loved to include -- before this measure, they are included in the other -- one of the other measures that we'll be discussing today. They don't have inpatient units for children, and when they have a child present to the ED with psychosis, they send them to a tertiary care center or a center that does have an inpatient psychiatric unit for children.

I think most inpatient -- and I don't have data to support this right now -- but I would imagine most psychiatric inpatient stays for children are likely not going to be happening in general adult psychiatric units. I imagine it's possible, but I would think it would be rather rare.

So, I think these children will be clustered in tertiary care centers or in actual psych facilities. Whether we would see more variation and gap in performance if this were put in a more widely distributed swath of hospitals remains a question unanswered, and that to me

would be part of stewardship of this measure.

It would be very important for us to continue to track performance and understand whether there's no variability because if there is no variability and it's a capped out measure, I agree it's not a very useful measure for accountability purposes.

Did you have anything you wanted to add? And, I --

CO-CHAIR SUSMAN: Any -- any comments further about what the anticipated improvement in patient oriented outcomes would be?

DR. MANGIONE-SMITH: Oh, yes. Right. So, that was -- that was actually a key driver for this measure. It was not just trying to decrease length of stay or costs. Unfortunately, that's the only thing that the literature has focused on.

Our concern in thinking through this measure -- and you're going to be hearing about many more measures in this two-day period about the misuse of anti-psychotic medications in young

children --- there's concern on our part that anti-psychotics get started without a legitimate evaluation of whether a child actually needs to be on an anti-psychotic medication.

So, one of the hopeful outcomes of a measure like this would be that the appropriateness of treatment with anti-psychotics would improve. The other piece that we know is a problem is continued return to the emergency department with these types of symptoms and re-hospitalizations, so the hope is that if a psychiatrist were to become involved earlier on that appropriate follow-up care in the outpatient setting would be more likely to occur.

CO-CHAIR SUSMAN: Okay. Thank you to our developers.

And Kevin, you had a comment.

MEMBER SLAVIN: Sorry, I don't want to sort of keep on this, but there were -- I think, getting back to Dr. Moyer's comment about the younger children in particular, one of the problems with -- that I sort of foresee is that

sometimes a measure drives an action that is unintended.

So, in this case, even though it's a retrospective look-back on patients who were discharged with a diagnosis of psychosis, if a young child or even a teenager comes in with acute psychosis with no prior history of psychosis, the medical evaluation is often not completed within the first 24 hours.

But with the measure looking back and saying -- let's say this person does end up having psychosis as one of their discharge diagnoses, it might drive a psychiatric evaluation for those that were not captured in the data set when a psychiatric evaluation was not needed. And that's sort of one concern in terms of driving use of resources and use of psychiatrists.

Two other things I kind of wanted to mention. One is the supplemental study that was added or that was mentioned after the fact, when looking at that study a lot more closely, only 5

percent of the patients actually had a diagnosis of psychosis, so the shortened length of stay was actually based on about 80 percent of patients whose final diagnoses were anxiety, depressive disorders, or some out of form disorders, which I anticipate to be shorter term hospitalizations as well, and so I don't know if that actually adds to the data driving the evidence for this, and I was wondering if there are any comments about that.

And, then one question about the abstraction tool. The -- I noted the list of extenders didn't include residents, and I wasn't sure if residents actually fell into that list as well. Like a psychiatric resident might evaluate a patient, but the psychiatrist might not sign off on it until after. I didn't know if they were included in that list of the abstract-able consultations.

CO-CHAIR SUSMAN: So, why don't we have the results.

DR. BARDACH: So, I'll start, and then

Rita Mangione-Smith might have some more things to add.

Just a couple things about the age group. So, we did look at the distribution of ages in the eligible patients. There's a very small number of kids who are younger than the age of ten.

It was only five percent of our -- of the eligible population in the field testing. So, just to reassure the crowd about the younger age groups, it's just not very many were actually eligible for it.

MEMBER SLAVIN: Were those patients who were discharged with a diagnosis of psychosis?

DR. BARDACH: Discharged with it, yes.

MEMBER SLAVIN: So, I think, the concern is the ones that don't get discharged with a diagnosis of psychosis because maybe they present with encephalitis or some other medical condition.

DR. MANGIONE-SMITH: One other thing

that we have discussed as a center since the initial workgroup called because they are such a small fraction of who we even saw show up in the denominator and because of this concern that you might be pushing people to prematurely be doing psychiatric consults in kids who very well may not need them.

I don't think that we would have an issue with limiting the age range to the adolescent population for both this measure, and we'll talk about the other one later today. I think that's a reasonable suggestion, and given what we found in the field test is supported by what we found in the field test.

CO-CHAIR SUSMAN: And, just remember, we're voting on the measure as presented, and certainly in subsequent work, developers can re-tool their measure, re-target their measure, or what have you.

So, David.

MEMBER EINZIG: So, I know this is just semantics, but when I read psychiatric, I

think psychiatrist. And, I'm sensitive to that because when I got consulted as a psychiatrist by the team, sometimes they really don't want a psychiatrist; sometimes what they're looking for is a psychologist, and so I think the semantics are important there.

And, just for clarification, so if it's an LICSW who does that initial consult, that does not count as a psychiatric consult?

DR. MANGIONE-SMITH: It would count as long as it were countersigned by either a PhD psychologist or a psychiatrist. So, if there was an evaluation done by a licensed social worker and it was cosigned by either a PhD psychologist or an MD psychiatrist, it would count.

MEMBER EINZIG: Yes. And, forgive my ignorance. So when a psychiatrist or a PhD psychologist cosigns, does that imply that they also see the patient or they just review with the --

DR. MANGIONE-SMITH: My assumption --- as somebody who cosigns residents' notes all the

time and I'm expected to see the patient -- my assumption is they would have already -- they also would have evaluated the patient to some degree.

CO-CHAIR SUSMAN: Yes, I think, just my own background, it would seem that there's variability in state regulations of licensure. And, the scope of practice, certainly, there are some states where those individuals are fully licensed to independently evaluate and treat using psychotherapies at least their patients, so that is, I think, an important issue.

Okay, I'm seeing our -- I'll get your comment, but I'm seeing we're starting to wind down here, so if we can focus on evidence or critical questions that haven't yet be answered, we can go on to voting on evidence.

Ricardo.

MEMBER QUINONEZ: I just have a very technical question about the measure. The 24 hours, if a child is admitted at a -- comes to the ER to an adult facility and then subsequently

gets transferred to the either an inpatient psychiatric facility or a children's hospital -- most likely a children's hospital -- where there's a 24-hour start, and where did you come up with the 24 hours? What was the --

DR. MANGIONE-SMITH: So, the 24-hour clock starts at the timestamp of admission to the hospital where the child is treated for their psychosis. So, if you were transferred from an adult ER to a children's hospital psych unit, the 24 hours starts when you're admitted to the psych unit.

And in terms of the 24 hours, it was heavily and strongly debated at our Delphi panel whether that was the right cutoff. We started -- that measure was drafted at 48 hours and the mental health people on our Delphi panel felt that that was too lenient and wanted it to be 24 hours, but then agreed that it could be 48 hours as long as you provided a justification for why it took 48 hours.

So, that is the -- where the 24 hours

came from was from the Delphi panel recommendation that it be made more stringent.

MEMBER MOYER: Just another question for the developers. So, if the child got the psych consult in the emergency room prior to admission, how is that handled?

DR. MANGIONE-SMITH: So, if it's at the index hospital where you're measuring, that counts.

MEMBER MOYER: How would you know that? What -- is there -- is there a process by which that would have been determined?

DR. MANGIONE-SMITH: So, the chart review, you want to say what it --

DR. BARDACH: Yes. Just says, "Include in this interval any psychiatric consult that may have been done in the marker ED prior to admission if the patient was admitted by the marker ED." So, it's just part of the abstraction instructions.

CO-CHAIR SUSMAN: Okay. Any final questions?

David, is your -- no, okay.

Yes, Virginia.

MEMBER MOYER: You're probably about to do this, but if you are you can pay me later. I just need to be very clear on what we're looking for now. I'm reading the questions that are in the summary, "Is the evidence directly applicable to the process of care? Is there sufficient evidence of the relationship of this measure to patient outcomes?"

And I'm looking at those and wondering if that's what we are about to --

CO-CHAIR SUSMAN: Yes. I wonder if the staff might be able to put up the algorithm if possible. If not -- basically, we're dealing with either -- and this is what it looks like. Perhaps you have it in all your volumes of NQF-related material.

But the path I see us going down is one, at maximum, there's low evidence or perhaps insufficient evidence. We're going to be voting on the sufficiency of that evidence first.

If it votes out as insufficient, then we would be able to decide by the wisdom of this group that there should be an exception that despite having insufficient evidence, we think this is important to measure. The evidence is a must pass, and if it doesn't, then we stop our discussion at that point.

I don't know if the NQF staff want to add any additional comments or perspective.

DR. NISHIMI: I just want to clarify that the questions you see on the evaluation aren't the questions for voting. Those were to guide your evaluation.

MEMBER MOYER: I guess I need some clarity between what's low and what's insufficient. It references U.S. Preventive Services Task Force approach, and the Task Force does not distinguish between low and insufficient. Low is insufficient.

DR. NISHIMI: Low and insufficient derive from the algorithm, so if you follow the algorithm, you can conclude that sufficient

empirical evidence was provided and it will send you right to moderate or low or you can conclude that it was insufficient, and then the voting would be, you know, is it insufficient but an exception can be made because -- agree that it is okay to hold providers accountable in the absence of evidence because of the benefit to the patient, so it becomes a risk benefit that you weigh, or if you just don't feel that it is of sufficient benefit, then you vote insufficient with no exception.

CO-CHAIR SUSMAN: So, the question is, what is the level of evidence? I think, from the discussion, we're probably debating whether there's low or insufficient evidence in this, and then we'll consider --- if it's insufficient -- whether there's an exception or not.

Is that clear? Do people have --

MEMBER MOYER: I'm still unclear what the difference between low evidence and insufficient evidence. There's almost never a question for which there's no evidence. There's

something out there.

CO-CHAIR SUSMAN: Yes. So, let me read the boxes, and, hopefully, that will clarify a little bit. It is a bit confusing. If you go down the path here, is empirical evidence, or is evidence submitted without systematic review and grading?

And, you know, there really wasn't -- there isn't much evidence to systematically grade or review here as I understand the presentation. And, if that's no, it would take you to the next page here, which takes you onto a path of insufficiency.

On the other hand, if there is empirical evidence submitted without systematic review and grading, yes, that goes down to, does the empirical evidence that is summarized include all studies in the body of evidence? If the answer is yes, does the agreement that the submitted evidence indicates high certainty benefits clearly outweigh undesirable effects or risks?

And, then you get a distinction between moderate, which I don't believe -- personally at least -- that we have here, or low, or if there's little or no empirical evidence and there's no real systematic evaluation of that. That would take us to the orange boxes, and you'll see here, or are there, or could there be, performance measures of a related health outcome, or evidence-based intermediate clinical outcome or process?

Is there evidence of a systematic assessment of expert opinion? And, I think, at least in this case, one could argue that there has been here.

And, does the Steering Committee -- all of us -- believe that it's okay beneficial to hold providers accountable in the absence of empirical evidence? If we don't believe that in our second round of voting, if we come to a conclusion it's insufficient, then you'd vote against it and say, "No, you know, not only do I believe there's insufficient evidence, I don't

think it is appropriate to hold people accountable for this."

So, are there questions about the algorithm?

CO-CHAIR BROOKEY: So, I think, the question is a good one because you're asking whether or not you're just going to go to one pathway or the other. If you say it's low, it stops, right? If you say it's insufficient, then there's an opportunity to go ahead and let it pass.

So, that's a very good distinction, and I think we have to be clear about the difference between low and insufficient. But I think that left-hand bottom box is the key box to look at, and we'll have to make a decision whether we want it to move onto the second page.

Do you want to comment any more about the distinction?

DR. NISHIMI: It's really the judgment of the Committee as to whether the evidence provided was systematically reviewed, which -- or

it's not a systematic review, and so if it's not systematically reviewed, we know there's no grading, but then that will send you down to the no into the insufficient area.

But if you feel that, you know, the -- there was, you know, a lot of evidence that was there, and it was reviewed, and it just happens to be low, that sends you to the right. So, in my mind, it's what's in your mind what constitutes that without systematic review and grading.

CO-CHAIR SUSMAN: David.

MEMBER KELLER: Yes, I'm sorry; I thought I understood it, and then I was listening to John, and now I don't.

(Laughter.)

CO-CHAIR SUSMAN: Good work, John.

MEMBER KELLER: So, I just want to be, be clear. So, in order to pass this first hurdle, you need to have either moderate or high evidence?

DR. NISHIMI: No, you need to have --

the choices that will come up are high, moderate, low, insufficient.

MEMBER KELLER: Okay. And, so, insufficient doesn't move you -- so -- so --

DR. NISHIMI: And, then --

MEMBER KELLER: So, to Ginny's point, any of those three high, low -- moderate, low, move us to the next phase.

DR. NISHIMI: No. I'm confused about what you're asking.

CO-CHAIR SUSMAN: So, if it's a rating of low evidence, that does not move it. Is that correct?

DR. NISHIMI: Correct.

MEMBER KELLER: That's what I wanted -- okay.

CO-CHAIR SUSMAN: So -- so if you rate it low --

MEMBER KELLER: Moderate or high does move it.

CO-CHAIR SUSMAN: Moderate or high would move it.

MEMBER KELLER: Okay.

CO-CHAIR SUSMAN: If it gets insufficient, it doesn't automatically move unless we vote for an exception.

DR. NISHIMI: Right.

CO-CHAIR SUSMAN: Is that --

DR. NISHIMI: So, if you can vote -- if you vote insufficient, then you can then choose to vote insufficient with exception or insufficient with no exception.

MEMBER MOYER: So, if there's a systematic review that uncovers very little in the way of evidence, but there is a systematic review that's low, and we can't vote an exception in that case.

DR. NISHIMI: Correct.

CO-CHAIR SUSMAN: That is correct.

MEMBER MOYER: So, the choosing between low and insufficient is really more a matter of deciding what we want to do next than it is assessing the evidence.

CO-CHAIR SUSMAN: Could be.

DR. NISHIMI: You can decide it was not sufficiently systematic. You know, to your mind, not everything was presented, ergo it's low. I mean, there are -- there are different ways to get to low versus insufficient.

CO-CHAIR SUSMAN: I know this is a somewhat confusing decision point, but just to recap: if you vote low -- and the majority of us go there -- we'll get to no further consideration. If you vote insufficient, it will either die or we will decide that there should be an exception.

And, it should be rated on the level of evidence as presented today and in the materials provided, including the provision of systematic review of clear evidence or not. And --

MEMBER MOYER: Because the concern that I'm continuing to have is that this means that we're not -- we need to assess the evidence and its -- the totality of the evidence, the body of the evidence, independent of where that's

going to take us. First, we assess the body of the evidence.

So, we shouldn't be deciding between low and insufficient based on what we would like to be able to do. We should be deciding between low and insufficient based on whether it is low or insufficient.

CO-CHAIR SUSMAN: I think -- I think, you're right spot on. The algorithm is built in a way that it separates out those decisions, so I think we need to try to -- as much as possible -- adhere to the criteria and the processes outlined.

DR. NISHIMI: But some will consider the review that was provided to be -- some will consider the answer to box seven no, and that will send them down to insufficient. Some of you will consider it to be yes, and that will send it to the right.

That's a decision each of you make. It's not something that you all have to collectively -- it's a judgment call at that

point, so that's why I'm not being specific on which direction it takes you. That's a decision for you to make.

CO-CHAIR SUSMAN: So, are there any other questions about this, or let's then learn how to use our clickers. This is the first day of medical school. We welcome you all here, or PA school, or wherever you happen to be.

DR. NISHIMI: Severa.

CO-CHAIR SUSMAN: Professor Severa.

MS. CHAVEZ: Thank you. Okay. So, to vote, please point your clicker at my laptop here. And you can only vote once during voting per PowerPoint, but you can change your vote at any time during the ten seconds that we get to vote by pressing on the number that correlates with the answer that's shown on the PowerPoint.

Any questions?

MEMBER MILLER: Yes. This is Marlene. Can you, you know, can you tell me the choices that it's for?

MS. CHAVEZ: Marlene, I will read the

options since we're actually ready to vote.

MEMBER MILLER: Okay.

CO-CHAIR SUSMAN: Does anybody have any questions about the use of the clicker?

(No audible response.)

CO-CHAIR SUSMAN: Are we going through a practice clicker or --

DR. NISHIMI: We're going to see if we can do it on the first try.

CO-CHAIR SUSMAN: All right. Okay, this is high stakes. Okay. So, the choices -- Marlene and others on the phone -- one is high, only eligible if the QQC, which is what?

DR. NISHIMI: Quality, consistency and quantity of evidence.

CO-CHAIR SUSMAN: Okay. If it was submitted, which we do not have, I think it's safe to say.

Two is moderate, three is low, and four is insufficient. So, let's go ahead and see if we can vote.

MS. CHAVEZ: So, we're now voting for

Measure 2805, Pediatric Psychosis: Timely Inpatient Psychiatric Consultation. Okay. Ready.

CO-CHAIR SUSMAN: Okay.

MS. CHAVEZ: Go. I'm sorry.

(Pause.)

CO-CHAIR SUSMAN: I'm sorry, what did you say?

MEMBER MILLER: I was just telling Severa I sent in mine by the chat on the web link.

CO-CHAIR SUSMAN: Got it.

DR. NISHIMI: We have it.

CO-CHAIR SUSMAN: And, there was one recusal, that's correct. And, Jeff -- thank you, Jeff.

Vote early; vote often.

DR. NISHIMI: So -- so the results for evidence, 4 people voted for low and 11 for insufficient.

CO-CHAIR SUSMAN: So --

DR. NISHIMI: Yes, we were expecting

24 votes.

CO-CHAIR SUSMAN: Okay. We're going to redo this. Evidently, there was a malfunction. We only had 15 or so votes there, I think, and we have 24 expected. So, give us the word and we'll try again.

MS. CHAVEZ: Okay. So, we're re-voting. We're expecting 24 votes, 2 via -- via chat. Okay. Ready. Go.

DR. NISHIMI: And, please point your clicker directly to Severa.

CO-CHAIR SUSMAN: Okay. And, for those on the phone, please vote now through your chat function.

(Pause.)

MS. CHAVEZ: One more. Good. We got it.

CO-CHAIR SUSMAN: All right. What? Well, that resolve -- for those of you on the phone -- has one person under moderate. We've met the enemy, and it's IT.

(Laughter.)

CO-CHAIR SUSMAN: Would it be acceptable to do a hand vote for this and --

MS. MUNTHALI: Yes.

CO-CHAIR SUSMAN: Okay.

DR. NISHIMI: We'll try it one more time. We're going to do it one more time, and then --

CO-CHAIR SUSMAN: Okay. We're going to try the electronic version again. Again, for those of you over the phone, please register your votes via chat function. Let's wait just a second until our technical wizards are ready.

MS. CHAVEZ: Okay. Ready.

CO-CHAIR SUSMAN: Okay. Let's vote.

(Pause.)

DR. NISHIMI: I see 25 responses. Why does it say 29?

CO-CHAIR SUSMAN: Remember, we have two on the phone.

MS. ALLEN: Okay. We need to --

DR. NISHIMI: Until we get this sorted out, we're going to just go ahead with a hand

vote.

Lauren and Marlene, if you could say your vote over the phones? State your name first and then your vote. We'll start out with Marlene and Lauren.

MEMBER MILLER: This is Marlene; I vote low.

MEMBER AGORATUS: This is Lauren; I vote moderate, but I always vote high, so --

CO-CHAIR SUSMAN: There are always one, yes. Thank you.

DR. NISHIMI: Okay. Anyone in favor of high?

(Show of hands.)

DR. NISHIMI: Moderate?

(Show of hands.)

DR. NISHIMI: Low?

(Show of hands.)

CO-CHAIR SUSMAN: Keep your hands raised until they've got them all. It's a very technical process; it involves high-speed computing.

And, finally, insufficient. And, remember, we have one recusal.

(Show of hands.)

CO-CHAIR SUSMAN: You can't see --

MS. ALLEN: For the record, --

CO-CHAIR SUSMAN: Yes, please.

MS. ALLEN: For the record, we have 0 high, 1 moderate, 6 low, and 16 insufficient.

CO-CHAIR SUSMAN: So, the insufficients have it.

And, now we would consider a motion to consider an exception if that is the will of the group or not.

(Show of hands.)

CO-CHAIR SUSMAN: So, there is a motion on the table. A second?

(Show of hands.)

CO-CHAIR SUSMAN: Okay. To consider an exception, and remember that the exception here is: does the Steering Committee agree that it's okay or beneficial to hold providers accountable for performance in the absence of

empirical evidence of benefits to patients? Consider potential detriments to endorsing the measure -- for example, focus attention away for more impactful practices -- more cost without certainty of benefit, divert resources from developing more impactful measures.

So, we have a vote of either an exception, which will be our first or second no exception. And, we'll be taking the vote on the phone first, and then we'll take the vote of the group here present.

DR. NISHIMI: Okay. Severa.

MS. CHAVEZ: Okay. So, for the benefits of the ones on the phone, one is insufficient evidence with exception, and two is no exception. So, we'll try voting again using our clickers. Hold on. All right.

CO-CHAIR SUSMAN: And, for those of you on the phone, try to use the chat function.

MS. CHAVEZ: Okay. Go.

(Pause.)

MS. CHAVEZ: Twenty-two. Twenty-four,

okay. So, 11 voted insufficient evidence with exception; 13 voted no exception.

(Off microphone comment.)

MS. CHAVEZ: Yes. And, we have -- we have 25 Committee members right now voting with one recusal, so 24 votes.

(Off microphone comment.)

MS. CHAVEZ: Yes. Yes. Somebody --

CO-CHAIR SUSMAN: Yes. So, just to clarify. There was one recusal -- two on the phone that our staff assured were in the count -- and it's 11 voted for an exception, while 13 voted against; therefore, there's no exception. Therefore, the consideration of this measure now stops.

I will take a moment of privilege to ask if there's any further feedback to the developers. They've obviously spent a lot of hard work doing this, and clearly, this was a very divided group of folks. I think, we all had a sense that this is an important issue, and yet the body of evidence is not as well-developed

perhaps as it should be.

Any other feedback or suggestions?

Yes, please.

MEMBER MORROW-GORTON: I would just suggest that you consider going back and looking at the whole population of children who presented with psychotic symptoms, to sort of divvy out who those kids where that didn't have a psychotic diagnosis at the end, and that you look at the age groups.

CO-CHAIR SUSMAN: So, there's been some feedback about the age group, looking at presenting population of those with psychotic symptoms.

Any other feedback?

(No audible response.)

CO-CHAIR SUSMAN: Again, I want to thank you very much for your clear, concise answers. I know it's a lot of hard work, and I recognize you're probably disappointed, but thank you very much.

We were supposed to be taking a break

at 10:30, so we have a decision point of whether to go on and do our next -- we're about a half hour ahead -- or whether to take a break now, instead. So, without using the clickers, how many would you like to take a break now? Raise your hands.

(Show of hands.)

CO-CHAIR SUSMAN: And how many would like to press on?

(Show of hands.)

CO-CHAIR SUSMAN: Only a few, so we're going to take a 15-minute break; that should get us back here at 10:15. Thank you very much.

(Whereupon, the above-entitled matter went off the record at 10:00 a.m. and resumed at 10:17 a.m.)

CO-CHAIR SUSMAN: Okay, you don't get to vote if you aren't at the table. So, I know that is an incentive.

Okay, so our esteemed colleagues from the Seattle Children's Research Institute and UCSF will come back for a second round here. I

think, what, a three-round -- three rounds, okay.

This is consideration of Pediatric Psychosis: Screening for Drugs of Abuse in the Emergency Room. It is Measure 2806.

First of all, let me ask if there are any recusals for this one. Jeff or others?

Any further recusals, on the phone? Okay, so let's turn to our developers for a brief description and context. Thank you.

DR. BARDACH: Thank you. As I mentioned before, we used the same process to develop this measure as we used for 2805. So, there is a lit review, a multi-stakeholder Delphi panel, and then field testing.

This measure focuses on the pediatric patient scene in the emergency department. It is a medical records-based measure of the percent of children and adolescents age 5 to 19 years old with a discharge diagnosis from the ED of psychosis who are screened for alcohol or drugs of abuse while in the ED. The intended level of measurement is at the hospital level, which are

the results that we present.

We would like to just briefly address the most major concern brought up in the working group call regarding the appropriateness of this measure for the younger age group. In order to do this, we looked at our data by age groups. For this measure, 26 of the eligible patients or 10 percent of them were in the younger age group between 5 and 10 years old. And performance differed in a statistically significant way between the two groups with a mean score of 6 for the younger kids and a mean score of 31 for the older kids. This supports the committee members' concern that younger patients might be treated differently. And so we're amenable to responding to the concern by narrowing the age range for the measure to only include patients 12 years or older.

That's all we have to say.

CO-CHAIR SUSMAN: Okay, thank you. Short and sweet. So, we have Kevin, Karen, Martha. Kevin, you are first listed. And if you

would please just give us a quick overview and focusing on the conceptualization of the measure and the evidence.

MEMBER SLAVIN: So, as mentioned, this is a measure regarding screenings for drug of abuse for children who present to the ED with psychosis. And sort of looking through the actual measurement, it seems fairly straightforward. The evidence I think we are going to find in discussion is going to be similar to the last measure in that there really isn't a lot of empiric evidence one way or the other. However, if you look sort of beneath the surface, if this is a measure designed to identify children or youth with psychosis who have substance abuse, there certainly is a lot of information about the rates of substance abuse in youth with psychosis.

Most of the measure's recommendations come from the AACAP Guidelines but it is important to note that within the guideline for the recommendation for screening for drug abuse

or drugs that can be abused and alcohol, they do have an out which says exposure to drugs or alcohol cannot otherwise be ruled out. And that actually raises, in my mind, some questions about the actual practicality of this measure.

From a reliability perspective and validity, we will talk a lot more about that, as we sort of go on.

Actually, just getting to the gap, because I guess that is the second item on there, if you look at the measure as they tested it, there was actually -- seemed to be a decent spread of gap in the performance that was not just specifically based on the age ranges with, interestingly, some of the smaller hospitals performing better than some of the larger hospitals.

When looking at reliability, there may be some questions that come up about sort of the ranking of where psychosis appears in the list of diagnoses that somebody may present with because it does sort of focus on the first and second

diagnoses for the measure at the discharge diagnosis. And it is possible, I don't want to say likely, but it is possible or perhaps even probable that psychosis may not be one of the top two diagnoses for that particular ED visit but that doesn't mean that the screening shouldn't necessarily be done.

Validity, we will talk more about that process, I'm presuming, if we get past the evidence part.

So, I don't know if there is any other things people wanted to bring forward or discuss.

CO-CHAIR SUSMAN: Well, why don't we work with that and then as we proceed, we can get in further depth?

Other comments from the primary reviewers? Yes.

MEMBER BERGREN: Well, I had actually interpreted it differently as I thought it was to be looking at children who presented with psychotic symptoms -- symptoms of psychosis but not diagnosed psychosis and then ruling out drugs

or alcohol as a cause of those symptoms.

DR. MANGIONE-SMITH: So, actually, it, again, is looking at children who were in the ED and diagnosed with psychosis. So, again, the eligible population is identified using administrative claims data and the diagnosis of psychosis by ICD-9 codes. And what we are looking for here is comorbid substance use, which is very known to be a common phenomenon in children with psychotic disorders and is commonly missed. And if it is missed and untreated, obviously, has bad consequences.

MEMBER BERGREN: Okay.

CO-CHAIR SUSMAN: Thank you.

MEMBER SLAVIN: And actually if I -- I'm sorry.

CO-CHAIR SUSMAN: Go ahead, please.

MEMBER SLAVIN: No, that was one of the concerns I had but if you read the measure in the way it is described, it talks a lot more about the comorbidities but one of the concerns is that the measure actually measures two

different things because it does, since it is based on the ED assessment, it does also seem to be looking at potential acute triggers for psychosis, in terms of like the reliability and validity of the measure raises some concerns in my mind about how it is actually applied, rather than so much what the intent of the measure is.

CO-CHAIR SUSMAN: Okay, I see some hands up, so to speak. Maureen.

MEMBER EDIGER: Mine is just a procedural question. Is it an option for us to consider it as amended on the different age group?

CO-CHAIR SUSMAN: No. As I understand it, NQF staff, please correct me if I am wrong, we need to look at it as specified currently. If through our process this is sent back, they could make that change and then, through relatively rapid turnaround, get it back if they wanted to change the age specification.

But what we need to do is look at it as specified and with the same age. Correct?

I'm seeing a lot of nodding heads.

DR. NISHIMI: If they have the data on testing and can bring it back in a fast turnaround, then the committee can reconsider it at the post-comment phase. But today, you vote on it as you see it.

CO-CHAIR BROOKEY: Just to clarify, Robyn, if a member would like for it to come back as amended, should they just provide their input today?

DR. NISHIMI: Yes, the feedback that you give the developer will help inform their decision-making, obviously.

CO-CHAIR SUSMAN: So, as I did with our first measure, depending on our outcome, I will ask for feedback to each of the developers so that we can give them some benefit of your expert guidance.

Okay, Amy, thank you. I have got a spotter here now.

MEMBER HOUTROW: I just wanted to make sure I understood this correctly. So, if someone

presents with psychotic features in the ED and is discharged from the hospital with psychosis as their diagnosis, then we are looking for the comorbid percentage of children who have drug or alcohol abuse. But what about the children who presented to the ED who then were diagnosed with acute drug overdose and then, therefore, weren't discharged with a diagnosis of psychosis? Those children are no longer in the denominator. Is that right?

CO-CHAIR SUSMAN: What's the response?

DR. NISHIMI: That's correct.

CO-CHAIR SUSMAN: Okay, was there another comment?

MEMBER FINKELSTEIN: So, it is kind of back to the process thing and I don't want to be too obsessive about this but if that were the case, if we thought that the age range really mattered and that that was the reason to vote no on what the specified measure is we are looking at today, would that be at the phase we were at before in the evidence phase or would that be at

the validity phase, where we didn't move it forward?

DR. NISHIMI: Could be at either.

CO-CHAIR SUSMAN: Does that help?

MEMBER FINKELSTEIN: Yes. It is that the box about the exception or not that is unclear to me whether we have to be strictly about as it is written or whether, with a change, it could have an exception and it is a tweak that it needs.

CO-CHAIR SUSMAN: Well, I think you know we have to stick to what was submitted, as submitted. And anything that we would think of as tweaks, we can provide as feedback. But what we are voting on today is the documentation, materials, and the clarification provided by the developers.

Kevin.

MEMBER SLAVIN: If it helps in the AACAP Guideline that is quoted, it says specifically youth. It does not say children and youth, as the citing for the recommendation for

testing. So, that was one of the questions that I had was what prompted the age range to be extended down from youth to five years of age.

And the second thing is I think the measure, as it is reported, the diagnosis of psychosis is based on the ED diagnosis, not the inpatient diagnosis. And you can correct me if I am wrong but that is the way that I read the denominator statement.

DR. NISHIMI: That's true. It is only ED patients. So, it is patients discharged from the ED with a discharge diagnosis from the ED of psychosis. Thanks for asking for clarification.

MEMBER MOYER: If they are discharged to inpatient care, is that still a discharge or do you mean discharged to home?

DR. NISHIMI: Discharged to home.

MEMBER MOYER: So, a kid who comes into the ER psychotic and so forth and gets admitted would not fall into this group.

DR. NISHIMI: That's right. Sorry, let me just double check one second. I'm almost

sure that is right.

MEMBER MOYER: Yes, if they either got admitted or they got transferred to a psychiatric facility, one would assume they still needed to be evaluated.

CO-CHAIR SUSMAN: Okay, we will get an answer on that. I will go to Debbie in the meantime.

MEMBER FATTORI: Can the developers explain the rationale for choosing the age range that you decided on?

DR. MANGIONE-SMITH: Yes, actually we can.

MEMBER FATTORI: Thanks.

DR. MANGIONE-SMITH: This is, unfortunately, gets back again to the Delphi panel having very clear opinions about age ranges for the different sets of measures we have presented to them. We presented to them measures on psychosis, measures on danger to self and suicidality and measures on substance abuse.

The substance abuse measures, which

you are not seeing any of today, were 12 to 19 years old. This one was a psychosis measure and they said oh, the psychosis measure should be 5 to 19. And this, unfortunately, slipped under the radar of this is actually about substance abuse and should have been 12 to 19. So, it was unfortunate that that is what happened.

CO-CHAIR SUSMAN: Okay, such things happen. I wonder if the folks who reviewed this more closely could speak again to the evidence. Was there a real systematic review? How was that conducted? I understand there was a Delphi process but one of the key judgments we have to make upfront here is about the quality of evidence. And so far, I am getting a sense that we are back to this issue of low or insufficient evidence. So, Kevin, I don't know if you want to address it or one of the other.

DR. NISHIMI: Can I address the prior question?

CO-CHAIR SUSMAN: Okay, yes, address the prior question before we go on to something

else.

DR. NISHIMI: Sorry. There was just one outstanding question. So, patients who are seen in the ED, they needed to get the drug or urine or drug or alcohol testing done in the ED before getting either discharged or admitted to hospitals. So, it did include inpatients as well. Yes, they had a discharge diagnosis. The diagnosis was made in the ED before they went up.

CO-CHAIR SUSMAN: So does that mean, just for clarification, if someone was admitted and then it was on their admission orders that didn't count if it was done by say general medicine or general psychiatry?

DR. MANGIONE-SMITH: That's exactly right. So, the ED would be held accountable for having done that testing.

CO-CHAIR SUSMAN: So, you are really looking at accountability at the ED level, as opposed to a more systems-level.

DR. MANGIONE-SMITH: Simply because we thought there would be a fair number of children

seen in an outside ED and sent somewhere else. And we wanted to make sure that the index ED was being accountable for the testing.

CO-CHAIR SUSMAN: Okay. I'm sorry for that interruption but you are waiting with bated breath.

MEMBER SLAVIN: No, I'm just waiting.

(Laughter.)

MEMBER SLAVIN: So, this gets into the discussion similar to the evidence discussion from the previous measure. I can't tell you for sure that it is the same AACAP Guideline or practice recommendation that this comes from but it is part of that same process, where there is not a lot of empiric evidence stating that this screening has an impact on outcomes. There is, certainly, a lot of data suggesting that there is a higher rate of substance abuse in this population. If you look at the adult literature, most of that is tobacco use, which is not screened for in the panels that are suggested. On the other hand, that is, I think, not what

this measure is really designed to focus on.

This particular recommendation for this measure comes from the consensus panel that reviewed the literature. So, it seemed like it was an overwhelming consensus but it was not based on empiric evidence that stated that it improved short-term or long-term outcomes.

CO-CHAIR SUSMAN: Okay, thank you. Other questions? Yes, please, David.

MEMBER EINZIG: And I'm not sure if this is illustrated in the proposal here but just playing the common sense card, it does make absolute clinical sense, and again, have a direct correlation with outcome if they are chronic alcohol abusers or benzodiazepine users and that is not picked up, they get admitted to the floor in the psychiatric hospital and they get withdrawal seizures, DTs. I don't think that is illustrated in here but there is evidence it should be quality of care.

CO-CHAIR SUSMAN: Okay. Other comments or questions? Yes, Ricardo.

MEMBER QUINONEZ: I just have a quick question. When we are talking about -- and I did read the measure but I don't remember these specifics. When we are talking about drug testing, are we talking about a UDS or are we talking about a comprehensive drug testing or what are we?

DR. NISHIMI: I can read you specifically what the instructions to the abstractors was. Indicate if the patient had a urine drug screen or serum alcohol screen while in the ED. The alcohol test will be a separate test from the drug test. The drug test must be comprehensive in that it tests for multiple types of illicit drugs. Do not give credit for tests that include results of just a single drug. And then it helps the abstractor. Drug screens commonly include tests for benzodiazepines, barbiturates, methamphetamine, cocaine, methadone, opiates, tetrahydrocannabinol, et cetera.

MEMBER QUINONEZ: Okay. Well, was

there any discussion during your Delphi panel whether there was concern about how reliable urine drug screen tests are? I always remember toxicologists saying you know a urine drug test was never really developed for what it is used today and it misses a lot of things and it cross-reacts with a lot of things.

DR. MANGIONE-SMITH: It wasn't discussed.

MEMBER QUINONEZ: Okay.

CO-CHAIR SUSMAN: David Keller.

MEMBER KELLER: The other David down here. So, this kind of builds off of those last two questions and I am wondering if the measure developers gave any consideration to using non-laboratory screening for substance abuse, as opposed to using a drug test, particularly around the issue of alcohol, which is the most prevalent drug used in adolescents in general and I suspect, also, in adolescents who present with psychosis.

DR. MANGIONE-SMITH: So we actually

did have a measure that looked at alcohol screening, using a validated screener in the ED. And performance was so low across all five hospitals, that we felt that using validated screeners cannot be used as a standard of care at this point in time because adoption is so poor. They all scored under ten percent on a zero to hundred scale on that measure.

CO-CHAIR SUSMAN: And when you say so low, it is their actual use, not their performance?

DR. MANGIONE-SMITH: We couldn't find any screens in the patients we thought they were indicated for.

CO-CHAIR SUSMAN: Okay.

DR. NISHIMI: It's not that so few patients screened positive. It is that so few hospitals --

CO-CHAIR SUSMAN: Right. Or that the reliability of screening in psychotic kids was poor.

Okay, on the phone, are there any

comments?

MEMBER MILLER: No. This is Marlene. I think I am still kind of stuck in the early comments of where we are in terms of the first measure, in terms of just the overall evidence has been looking at guidelines and the extrapolations from it from youth to all these ages is relatively weak.

CO-CHAIR SUSMAN: So, there is still some of I think the same evidence concerns.

Jon, do you have a comment or question?

MEMBER FINKELSTEIN: My comment would be for me it is even more clear that the evidence is insufficient but, in my view, the credible information we have about professional consensus is stronger here. So, as I go through the orange boxes on page two, deciding whether this should be with exception or not, it seems to me both clinical practice-based validity, the Delphi process the developers went through and the guideline of the profession are all right on

target that this should be done. I think we have this issue of the age group. But absent that, I think this is one that I would rate as insufficient with exception.

CO-CHAIR SUSMAN: Okay, John.

CO-CHAIR BROOKEY: Just a quick question about specs. I'm picky about the lab results, whether they were ordered or whether they resulted and whether there is evidence that people acted on the result. Is that included in the specs?

DR. NISHIMI: Yes, so great questions and important in terms of operationalizing. It is just the instructions to the abstractor is just whether or not the patient had a drug screen or serum alcohol screen while in the ED.

CO-CHAIR SUSMAN: Ordered while in the ED or resulted while in the ED?

MEMBER SLAVIN: I believe it specifies resulted because if you read the specs very carefully, it talks about not likely to miss lab results because of the chain of what happened.

So, it is based on whether or not it is resulted, not based on whether or not it is ordered, which was one of the questions that I had because many of these patients, in terms of obtaining urine for a drug screen are not necessarily going to be as cooperative as we would hope. And so, and there are issues with elopement and leaving AMA before specimens are actually collected. So, there may be an attempt to try to collect them but not actually obtain the specimen for testing.

CO-CHAIR SUSMAN: And on the other hand, I could imagine a very efficient system, where the person is transferred to their ultimate destination, whether on the floor and results are gotten up there and where information is shared in a good manner or not.

MEMBER SLAVIN: Are any of these admissions ever efficient?

(Laughter.)

CO-CHAIR SUSMAN: Well, you know I am sure there must be the perfect hospital somewhere.

Okay, it looks like we are gaining consensus about where we want to be here. So our first question will be voting on the evidence. And Kevin?

MEMBER SLAVIN: Actually, I just had one more question. I'm not sure if it would fall under here or not but it was the numerator, the part about the numerator that allows for, for want of a better word, partial credit for one or the other, if that could be explained a little bit why this isn't an all or none.

DR. MANGIONE-SMITH: So, the decision among those of us who were operationalizing the measure was that if you had done say a serum alcohol test but you hadn't done the urine drug screen, we wanted to give you credit for at least doing part of the right thing. So, that was just a decision of us as a development team that you should get partial credit. You get 50 percent if you did one and not the other.

CO-CHAIR SUSMAN: So, how would that be reported on a system level? Was it a

facility-level process measure?

DR. MANGIONE-SMITH: So, it would roll up for every eligible patient. So, you would either have a zero --

CO-CHAIR SUSMAN: You have to have two 50 percents to equal one?

DR. MANGIONE-SMITH: So, you would score either 0.5 or 1.0. And then you roll it up among all eligible patients to get your overall score, which was still terrible.

CO-CHAIR SUSMAN: Got it.

Okay, let's go ahead, then, and get up our screen for important to measure the evidence. Remember, for those on the phone, one is high, two is moderate, three is low, and four is insufficient. And are we in clickerville today?

MS. CHAVEZ: Yes. Thank you, Dr. Susman, for doing that for Lauren and Marlene.

Okay, so we are now voting on evidence for Measure 2806, Pediatric Psychosis: Screening for Drugs of Abuse in the Emergency Department.

We have increased the time period to

15 seconds and we are, again, expecting 24 votes because we do have one recusal.

CO-CHAIR SUSMAN: So, please point your clickers. Click away. And we will get our votes from our outside, in the electrons panelists.

And it looks like --

MS. CHAVEZ: So, I see 23 responses. We have to do it again.

CO-CHAIR SUSMAN: Okay, we will do it again. We have one recusal and 24 otherwise. Is that correct?

MS. CHAVEZ: Everyone in the room. Yes, I don't see -- oh, we are not ready. Sorry.

CO-CHAIR SUSMAN: Hold on. And again, for those of you on the phone, if you could let it be known through the chat function, we are still dealing with electronics here.

MS. CHAVEZ: Okay. All right, ready, go.

CO-CHAIR SUSMAN: Okay.

MS. CHAVEZ: One more.

DR. NISHIMI: Does anyone have a neighbor missing?

CO-CHAIR SUSMAN: Now, we have all moderate. If we work hard, we will get all high.

(Laughter.)

CO-CHAIR SUSMAN: The developers will rejoice.

DR. NISHIMI: Okay, we're just going to have to do a hand vote. There is a problem here.

CO-CHAIR SUSMAN: So, let's go ahead and have our two respondents on the phone.

MS. ALLEN: So, Lauren and Marlene, staff will be voting for you in the room. We already have your votes via the chat. So you don't have to announce your vote.

DR. NISHIMI: Okay. High?

Moderate?

Low?

Insufficient?

CO-CHAIR SUSMAN: Well, clearly, the insufficients have it. The counts will be up in

just a moment.

DR. NISHIMI: Okay, we actually have to have the numbers because we record the numbers in the report.

So, insufficient?

CO-CHAIR SUSMAN: Again, hands high. We have multiple counters now. If you want a bribe for your vote, now would be the time.

DR. NISHIMI: There were 19 insufficient. How many lows? Severa, how many lows?

MS. CHAVEZ: Two lows.

DR. NISHIMI: Three.

MS. CHAVEZ: So, 19 insufficients; 3 lows; and that would make it 2 moderates.

CO-CHAIR SUSMAN: All right, so, insufficient.

We will now entertain whether there should be an exception here. As we discussed during the first measure, this is an issue of do we think it is important, despite the evidence that this be endorsed, potentially, for both

accountability and for improvement purposes.

So, are we trying the electronic version again?

DR. NISHIMI: Yes, let's try.

CO-CHAIR SUSMAN: Okay, we are going to try the electronic version. For those of you on the phone, one is insufficient evidence with exception. In other words, you're voting exception and two is no exception.

Yes?

MEMBER MOYER: The question is about the point in the discussion where we would bring up the age range. Because at this point, in terms of no exception or exception, the age range is one of the most critical pieces of information.

CO-CHAIR SUSMAN: Yes, so I think, unfortunately, the age range is as currently specified and we need to vote on whether, as specified, it makes sense to have this with exception if, for example, one believed that the specification around the younger folks in this

cohort was inappropriate, then one would vote their conscience about this.

We could, if we got to the point where this wasn't voted with exception, provide feedback to our measure developers who have obviously already thought about this issue in the deliberations.

Jon, please put on your mike.

MEMBER FINKELSTEIN: Just to say it. So I know we are in a very structured process but I am a little concerned about the process just in this way. If that is the only thing for some of us holding this back and we think there could be a rapid cycle reconsideration on a call a few weeks from now, not to have gotten to talk about the validity, the reliability and the measure testing, will really set us back. I don't know how we will --

DR. NISHIMI: You would talk about it on the call.

MEMBER FINKELSTEIN: So, will there really be time for that? If that is the process,

I understand.

DR. NISHIMI: That is the process and you would then proceed. If it then passed forward, you would then proceed to discuss it.

CO-CHAIR SUSMAN: Okay, Kevin, did you have a --

MEMBER SLAVIN: Well just in regards to that question. One of the key factors for the validity part of this is whether or not it is related to a true quality outcome. And, in my mind, the age issue falls under the validity portion not under the evidence portion, although the evidence was not cited with this age range in mind. So, it could be looked at probably either way.

CO-CHAIR SUSMAN: There is probably overlap between the two in this regard.

DR. NISHIMI: That's why I answered the original question it could be in.

CO-CHAIR SUSMAN: Okay, any other questions or points prior to vote?

MEMBER HOUTROW: I have a concern.

CO-CHAIR SUSMAN: Oh, I'm sorry, I can't see.

MEMBER HOUTROW: If we are going to use the clickers but we are going to have no verification that the clicking that we did is accurate -- because if we are really making a decision whether this goes to a next part of the discussion, I think we need to do some sort of check to make sure if we are using the clicker that the clicker accurately represented what our votes were.

DR. NISHIMI: The clicking, the problem we are having is when we have to reset it because we didn't have enough votes. So, I don't have a concern de novo. Remember we only had, whatever it was, 22 and we were supposed to have 24. So, then when we reset it, that is when the problem came in and we are going to work on that at lunch.

CO-CHAIR SUSMAN: So, I think we are ready here. Again, for benefit of those on the phone, one is insufficient evidence with

exception and two is no exception. Severa, are we ready to vote?

MS. CHAVEZ: I think. Yes. Ready, go.

DR. NISHIMI: Please remember to point your click in Severa's direction.

CO-CHAIR SUSMAN: So, we have 22 there.

MS. CHAVEZ: The software collected 22 responses. Nadine and yes, Robyn.

CO-CHAIR SUSMAN: There should be 24. So, I would suggest we do the hands again. I think the point is well taken about making sure that we get this right, since it is so close.

DR. NISHIMI: Yes, we have -- so voting for one, insufficient with exception.

CO-CHAIR SUSMAN: Voting for one, insufficient evidence with exception.

DR. NISHIMI: So, 16.

CO-CHAIR SUSMAN: And I guess just as a check --

DR. NISHIMI: No exception.

CO-CHAIR SUSMAN: Okay, 16 plus 8 equals 24. All right, good work.

So, this is actually recommended to go forward with exception. And now we will go on to the gap or opportunity for improvement. Is that correct?

DR. NISHIMI: That is correct.

CO-CHAIR SUSMAN: And so the group who reviewed this closely, is there good evidence for a gap or opportunity for improvement?

MEMBER SLAVIN: I don't want to be the only one who speaks.

CO-CHAIR SUSMAN: Yes, your teammates are letting you down there.

MEMBER SLAVIN: You know in their testing, this was tested at three children's hospitals and two community hospitals. The volumes in a couple of the hospitals were fairly low. One had only 15 patients over the two-year period; one had only 18 patients over the two-year period. The range that they identified for performance was basically 18 percent to 83

percent. Interestingly, the hospital with the second smallest eligible patients, they had one of the best performances. So, in looking at this, it seems like there seems to be consensus that this should be done. And so there is a performance gap in terms of its not being implemented across the board and there is also a wide performance gap in the different hospitals in their emergency departments.

CO-CHAIR SUSMAN: Thank you. Any additional comments from our panelists? Questions?

If not -- oh, yes, please.

MEMBER MOYER: Very briefly, from the developers, the data on the younger kids in terms of a gap.

DR. BARDACH: Sorry. You want us to just give you the data on the younger kids again, on performance?

MEMBER MOYER: Yes, are there data in terms of the gap on the younger kids, the under 12s?

DR. BARDACH: Oh, where we just looked at the under 12 and seeing the range.

CO-CHAIR SUSMAN: Did you stratify by age?

DR. BARDACH: I don't think we looked at the performance range.

CO-CHAIR SUSMAN: They are arguing. Hold on.

DR. BARDACH: Hold on one second. Let me just look it up. We may or may not actually have it. We have it actually for the older age group but not for the younger age group, the range in performance variation, if that makes sense.

But you are specifically interested in the range of performance variation in younger children rather than older children. It was only five percent of our sample. So, I anticipate that it is going to be very difficult to say anything without just looking at noise.

MEMBER MOYER: That may be enough information, just to know that we actually don't

have a clue in that age range.

DR. BARDACH: Yes, I think that is a fine conclusion.

CO-CHAIR SUSMAN: Anything else you want to say, Kevin, on this?

MEMBER SLAVIN: I think the numbers were submitted with the supplemental submission and the numbers were very small.

DR. BARDACH: Yes, we didn't submit it by hospital. We didn't do individual hospital performance measurement. It was just for the group of kids in the younger age group. But performance was quite low, yes.

CO-CHAIR SUSMAN: Okay, any other questions about -- yes.

MEMBER BOST: I'm having trouble. What were the denominators associated with the rates that you are providing, the count of the denominator? Because usually if it is less than 21, I tend to not even think it is worth looking at.

DR. BARDACH: Yes, that makes sense.

So, it is provided in materials but I will read it out loud.

Hospitals, overall, had 257 eligible patients and then we just have them Hospitals A, B, C, D, and E. Hospital A had 36 and a performance of 25. Hospital B had 166 eligible patients, performance level of 18. Hospital C had 18 patients and 83 percent -- sorry -- a hospital performance score of 83. And then Hospital D had 22 eligible patients and a mean performance of 66. Hospital E had eligible patients of 15 and a mean of 40.

If you would like, I can read you the confidence intervals.

MEMBER BOST: No, but it just sounds like the first two are high enough, I think, to consider. So, that does narrow the gap a little bit.

CO-CHAIR SUSMAN: Any other questions or comments? Okay, let's go ahead and consider the vote on this one.

MS. CHAVEZ: So, we have increased the

time to vote to 30 seconds.

(Laughter.)

MS. CHAVEZ: Hopefully, there won't be any more problems.

CO-CHAIR SUSMAN: When you get to three days, let us know.

MS. CHAVEZ: And for Lauren and Marlene on the phone, one is high; two, moderate; three low; four, insufficient. And we are voting on performance gap.

CO-CHAIR BROOKEY: So, is there any harm in having everybody try to click it twice during the 30 seconds?

MS. CHAVEZ: There shouldn't be.

CO-CHAIR BROOKEY: Can we try that, just to see if we can get everybody in?

DR. NISHIMI: What we are going to do is announce we only have 22, keep clicking. Because we know that all the clickers are working because we have had a vote where they all showed up. So, it is clearly someone's is not quite hitting it.

CO-CHAIR SUSMAN: Okay, Severa, are you ready?

MS. CHAVEZ: Okay, ready, go.

CO-CHAIR SUSMAN: Go.

MS. CHAVEZ: Fifteen, twenty-three. One more.

DR. NISHIMI: One more.

MS. CHAVEZ: Twenty-four. Ten seconds left.

(Laughter.)

CO-CHAIR SUSMAN: First adherence.

MS. CHAVEZ: Okay, so two voted high; eighteen voted moderate; three voted low; and one voted insufficient.

CO-CHAIR SUSMAN: So, moderate carries it and we will move forward.

Next is the quality construct, let's see, and the issues of reliability and validity. So, reliability, any key comments on this portion of the measure? Are they consistent results?

Yes?

MEMBER BERGREN: Well, I just thought

it would be nice to work off of, is it Kevin, I can't see your card --

CO-CHAIR SUSMAN: Thank you. Kevin appreciates that.

MEMBER BERGREN: So, the reliability was 100 percent but there were only four patients sampled in the reliability but there weren't any concerns with it either.

The validity testing was done via consensus, face validity, through the Delphi process. And the consensus was that if this is performed, then that should result in high quality.

There is believed to be meaningful difference, based on whether or not the measure is performed and the likelihood of missing data is not likely. It is not likely that this would be missing data because it is data that is already captured in the chart.

CO-CHAIR SUSMAN: Okay, so we are going to be considering the reliability first, things like the statement of the numerator,

denominator, consistency of those results, what sort of reliability testing was done.

Are there other comments or questions? Let's do Kerri and then I will get over across the way.

MEMBER FEI: Okay, thank you. My question is about the denominator. And I just wanted to make sure that I am clear and thinking about this the right way.

The way it is stated here is that it is patients 5 to 19 seen in the ED with psychotic symptoms but it is really patients 5 to 19 discharged from the ED, could be to home, could be to another setting of care. I find that to be very confusing. And I get it. I think the denominator could be reworked for public-facing folks, so that they know every day people aren't going to get that part.

So, if the denominator could be restated to be reflective of what they are actually measuring, the population you are actually measuring for clarity purposes.

CO-CHAIR SUSMAN: Okay, across the way.

MEMBER MOYER: The discomfort that I am having, the measure of reliability seems to be pretty good for the older kids. I'm very uncomfortable that we have a whole bunch of -- we have a chunk of patients for accountability and we have people we actually haven't got the vaguest idea whether this is reliable in that younger age group.

And I'm trying to figure out where, and Jeff, I think we need your guidance as to where this fits in terms of our vote. I am not going to be comfortable approving this measure as it is currently stated.

CO-CHAIR SUSMAN: Well, remember at the end we vote on the measure altogether. And it is conceivable that it could squeak by and in the end we would say it wasn't something that we felt comfortable with going forward.

When there is a part of the population that reliability testing has been performed

either very scantily or not at all, it is going to question gee, we just can't say whether it is reliable or not. And there are also, I think with that, concerns about the evidence, as we just discussed and concerns about the validity because at least the face validity is much less. And as we described from the measure perspective, developer perspective themselves, this may not have been the intended outcome with the age range.

So, all, I think, important points that might limit our enthusiasm for the measure.

MEMBER MOYER: I guess I just have a concern if we vote positively -- I am sort of voting on what I understand about the older age group because it has a lot of positives. But in the end, there is a chunk that we can't address.

DR. NISHIMI: I think you need to weigh how you are going to vote. If it goes down, it can be reconsidered. If it goes forward, it can be reconsidered. I mean, so, but it will be reconsidered if it goes down.

CO-CHAIR SUSMAN: I mean -- oh. You are shaking your head. Obviously, there is something there that bothers you.

DR. NISHIMI: During the comment period, if you get all negative comments, let's say, even though it has gone forward, then you can reconsider it. If you voted down, the developer has the opportunity to bring new information forward and it gets reconsidered.

MEMBER MOYER: So, in any case, it can be reconsidered.

DR. NISHIMI: Right. It is unlikely to be reconsidered if all of the comments come in positive and you put it forward.

CO-CHAIR SUSMAN: So, I mean this is an iterative process. Once we go from here to get the comments, there is consideration, and then it goes to the bigger committee to be considered. But we really need to do our jobs well and with the best of our ability, given the specification we have, which is the larger age range, to decide where we go.

Let's go down to the end and then we will come back up.

MEMBER BOST: I just wanted to clarify. I don't think I heard it but besides the percent agreement, there was also the reported interclass correlation coefficient of 0.42 across the five hospitals, which is considered high by reliability standards, again, with the caveat that three of the hospitals had pretty small denominators.

CO-CHAIR SUSMAN: Thank you. Okay.

MEMBER STANLEY: I have some concerns about the denominator as well and concur with what Kerri had said about public face of this particular measure. But I am wondering if you can tell us, is there any -- do you have any data that shows, for example, the visits to the ED where there are psychosis symptoms but yet at discharge, there is not a psychotic diagnosis? Because if you are carving out those who have perceived psychotic symptoms but don't end up with a diagnosis, is there a missed opportunity

there for testing?

CO-CHAIR SUSMAN: Go ahead.

DR. MANGIONE-SMITH: Yes, I would say absolutely that is a missed opportunity for testing. We were, unfortunately, limited to the data sources we have to identify the population and, given over two years in one of these community hospitals, we only saw 15 cases and the number of charts you would have to review to try to find kids who had psychotic symptoms without a diagnosis wouldn't be logistically feasible. I wish there was some way to do that.

If we had EHRs in place where we could troll for something that picks up psychotic symptoms, rather than psychosis as a diagnosis, absolutely, we would want to, I think, look at those cases in the teenage age group. But, given the data sources we have for specifying this measure, it is not possible at this point in time.

CO-CHAIR SUSMAN: David, did you have another comment or --

MEMBER KELLER: Yes, I was thinking. I was processing what you said and I had almost forgotten what my question was. Just a process question going forward.

When a measure is being reconsidered at the follow-up phone call and the measure developer is allowed -- I heard the language you used Robyn was bring additional information to the discussion -- is the measure developer allowed to modify the measure being considered at that time as well? Because that is different than bringing additional information.

DR. NISHIMI: Yes, that would be encompassed by that, if they modified the measure and brought forth testing data to support that measure. They can't modify the measure and not bring you data to support it.

CO-CHAIR SUSMAN: I mean as it is, from a practical standpoint, most of their data is on the older age range anyway. So, the five percent or what have you probably aren't going to meaningfully change the testing data. Is that --

DR. MANGIONE-SMITH: So, that is just about what I was about to offer to the group. Given the workgroup call, we have actually already redone all of the analyses just including 12- to 19-year-olds and the results are extremely similar to what are in front of you today. And you know, we would have no issue with resubmitting it with that information.

CO-CHAIR SUSMAN: And just to chide that we need to consider the broader age range because that is what was submitted.

And Kevin.

MEMBER SLAVIN: So, my question just has to do with the reliability related to the coding part. And it kind of gets to Carol's question about not having psychosis as the final diagnosis. Since this only looks at the first two diagnoses of somebody coming out of the ED and it is possible, I guess the questions sort of are is the intent if the patient has psychosis and presents even with something else, they should still be screened because they have a

known diagnosis of psychosis or is it really only if they present with acute psychosis? And if there is something that sort of takes precedence over the psychosis reason for coming to the ED, then psychosis very likely could fall to the third or fourth diagnosis.

And this, again, all gets into coding is really designed to maximize billing opportunities as opposed to identify patients with real problems.

So, is it worth thinking about extending that diagnosis code further down or in terms of missing some of those opportunities?

DR. BARDACH: So, just a point of clarification for all these administrative measures. So, it says in the specifications, it uses lingo which I apologize for, it says a primary or secondary diagnosis. But the secondary diagnoses actually mean anything after the primary one. So, it includes every single, all diagnostic slots. And I apologize.

CO-CHAIR SUSMAN: Thank you for that

clarification.

David, did you have another question? Okay, no problem.

Okay, anybody else have any questions related to reliability?

And please vote on reliability alone here. And are we trying to do the electronic version?

PART 2 Section B

MS. CHAVEZ: Yes. So, we are now voting on reliability. Again, for those on the phone, one for high; two, moderate; three, low; four, insufficient. We are expecting 24 votes. Ready, go.

Nineteen so far; twenty-three; twenty-four.

Okay, so one voted high; thirteen voted moderate; nine voted low; one insufficient.

CO-CHAIR SUSMAN: Okay. So, we go on now to validity. And I think this is where many people have had some questions about the validity, particularly in our younger aged

patients.

So, Kevin, our go-to man on this measure. We might as well go to you again. Anything further about validity you would like to call to our attention?

MEMBER SLAVIN: No, I think the process for determining validity is the same as the previous measure, which we talked about a little bit, although, we didn't get that far in the actual discussion. And it really comes down to the validity for the age range, in my mind, at least, the validity for the age range as specified.

CO-CHAIR SUSMAN: And so as I understand the validity of this measure is based on the Delphi method, the recommended procedures through some august bodies, which if we follow that through, means that it is going to be either a low or moderate degree of validity, by definition.

DR. NISHIMI: Right. So, the highest -- the eligible ratings are moderate, low, or

insufficient because it was only tested based on face validity. So, it is not eligible even for high under the NQF rubric.

CO-CHAIR SUSMAN: Questions further about the validity? Yes, Kevin, you have a comment.

MEMBER SLAVIN: Actually yes, it is just procedural. Is this like the other measures that if it is voted low, it would stop at this point?

CO-CHAIR SUSMAN: That is my understanding that this is a must-pass. Is that correct?

DR. NISHIMI: Yes, it is must pass.

CO-CHAIR SUSMAN: So, low will get you out; two, which is moderate; and then insufficient, which is insufficient.

Any further comments? If not, let's move on to voting on this. We are getting our electronics in gear here.

MS. CHAVEZ: We are voting on validity. Again, it is the same one for high;

two, moderate; three, low; four, insufficient.

Ready, go.

Fourteen, twenty-two, twenty-four.

Okay, so nobody voted high; nine moderate; fifteen voted low; and zero insufficient.

CO-CHAIR SUSMAN: So, just a procedural question with the majority being low, does that preclude further discussion?

DR. BURSTIN: In some ways, particularly if there is an opportunity for change post-comment, I just suggest the committee finish up the remaining criteria, just so you don't have to repeat it on a phone call, which is always harder.

DR. NISHIMI: And I just want to announce for the record that the vote on reliability actually was in the gray zone, as well but when it is in the gray zone, you continue.

You have to be outside of 60 and outside of 40.

CO-CHAIR SUSMAN: All right. So, we are going to continue on at Helen's suggestion and look at feasibility. Are there any concerns about the feasibility, how this would actually happen in practice? Any concerns from those who closely reviewed? I didn't see any but I don't want to give it short drift.

MS. CHAVEZ: No.

CO-CHAIR SUSMAN: Thank you. Well, if we are ready, let's go ahead and vote on feasibility.

MS. CHAVEZ: So, we are voting on feasibility. One, high; two, moderate; three, low; four, insufficient. Ready, go.

Twenty-two, twenty-three, twenty-four.

CO-CHAIR SUSMAN: All right. So, it looks like this one is moderate or high, on average.

MS. CHAVEZ: Eleven voted high, twelve voted moderate, one voted low, zero voted insufficient.

CO-CHAIR SUSMAN: Okay, usability and

use. So, is this currently in use? Is it publicly reported? Has there been any information? What are the unintended consequences?

Any comments on usability?

MEMBER BERGREN: It is not currently in use. And it is to be used for benchmarking and quality improvement. And there were no unintended consequences reported.

And I don't recall -- I did look at the transcript of our phone discussion and didn't find concerns with the usability.

CO-CHAIR SUSMAN: Would usability be affected by the broad age range of this population that is currently specified? Again, it is a relatively small number of patients but those kiddos were younger.

Ricardo.

MEMBER QUINONEZ: So, I have a question for anyone if they can think of it because I couldn't think of that looking at this measure. What could be possible unintended

consequences from this measure? Because especially looking at administrative data, right? I mean the one I can think of is someone gets labeled as a drug user and is not because, again, it is not a very reliable test but you can't really see that through administrative data.

CO-CHAIR SUSMAN: I will go over here.

MEMBER MORROW-GORTON: And actually, you would probably look at it anyway because the young children have a risk of getting into somebody else's stuff like alcohol and whatnot.

So, this would be part of -- at least in my mind, this would be part of the workup of psychosis, even in a younger child presenting to the ER.

CO-CHAIR SUSMAN: I guess the unintended consequences, occasionally, could be things like diverting resources from a high-return, high-impact measurement an improvement efforts versus ones that are much lower; the excessive testing or use of resources.

Please.

DR. MANGIONE-SMITH: I just want to be completely clear. Our intent was not to work up causes of psychotic symptoms in the ED. It was really to look for comorbid substance use among people with psychosis.

CO-CHAIR SUSMAN: Kevin.

MEMBER SLAVIN: I mean, in my mind, the unintended consequences really are kind of a downstream effect. If you test more people and you have a test that is unreliable, you may be either incorrectly are labeling people early on and if somebody has psychosis, it may affect family dynamics. It may affect further treatment where they are seeking care. If they are labeled inappropriately, it also affects the false negatives, in terms of missing people who may definitely have an issue or problem.

So, the unintended consequences I don't think are so much, in my mind, for testing. In my region, you cannot admit somebody with any mental health issue to anywhere without having a drug screen done before they leave the ED to go

somewhere else. So, we already see this sort of put into play. It is just the consequences of having a test that has some unreliable results and what it does at the local level to that individual, as opposed to on a broader scale.

CO-CHAIR SUSMAN: Ricardo.

MEMBER QUINONEZ: And I agree with you. Again, just the point that that probably cannot be obtained from administrative data, that level of consequences.

CO-CHAIR SUSMAN: Okay, thanks for everybody's comments. Let's talk about usability and use and go ahead and take our vote.

MS. CHAVEZ: Okay, for those on the phone one, high; two, moderate; three, low; four, insufficient. Ready, go.

Twenty-one votes; twenty-four.

So, three voted high; fifteen voted moderate; five voted low; one voted insufficient information.

CO-CHAIR SUSMAN: Okay. Now, given our voted previously, do we need to vote on the

measure or not? Okay, so we are going to move to overall suitability for endorsement. One is yes, and two is no.

Any final comments on this? My only observation is that truing up the age range on this one would be relatively easy. And I would feel, personally, more comfortable with it trued up, rather than voting for suitability now.

Any other overall comments? Yes, please.

MEMBER KELLER: Just, again, to make sure I understand the procedure. So, at this point, if we were to vote no at this time, then it goes back to the measure developer and they have the opportunity to bring back both additional data and a revised measure for us to consider at the follow-up phone call, at which point we would change our vote.

DR. NISHIMI: Right. You are already in the gray zone for reliability. So, it is going to have to be addressed somehow.

MEMBER KELLER: And so we would have

to think about the reliability question and then think about our total question, if this turns out to be a no.

DR. NISHIMI: Yes.

CO-CHAIR BROOKEY: Well, just to add to the confusion, whether you vote yes or no, it still could be amended. If we have recommendations for the age range to be changed, it still could happen irrespective of the vote on this particular questions.

DR. NISHIMI: Right.

MEMBER KELLER: Got you.

CO-CHAIR SUSMAN: Let's go ahead and vote. We may want to give some feedback, say through a straw vote thereafter, depending on how this turns out. So, we have a one and a two. Yes, is one; two is no. Severa.

MS. CHAVEZ: Okay, everyone ready? Go.

Twenty-three. Okay, so six voted yes, eighteen voted no.

CO-CHAIR SUSMAN: Then maybe just as

feedback by a show of hands, if the age range was amended, as we have discussed to the older kids, the adolescents, would that change your belief around suitability for endorsement?

So, let's just have a show of hands yes and no. If the age range were changed, would you be more likely to vote in favor of this? And those otherwise, no? Okay.

So, it looks like the age is one of the specific and probably relatively easily correctable.

Any other work? Do we need to ask for comments or anything along that line? Okay.

Thank you very much. That was a, again, very helpful discussion and feedback.

So, unfortunately, we have fallen a bit behind but I'm sure this group will catch us up, as we talk to 2807: Pediatric Danger to Self: Discharge Communication with Outpatient Provider.

And any recusals on this one? Okay, then, let's move forward to our measure developer.

DR. BARDACH: Thanks.

CO-CHAIR SUSMAN: Guess who?

DR. BARDACH: Back again.

This measure is also a medical records-based measure. Eligible patients were children 5 to 19 years old who were admitted to the hospital with dangerous self-harm or suicidality.

The measure is the percentage of eligible patients with documentation in the hospital record of a phone or email discussion between the inpatient and outpatient providers, regarding the plan for follow-up. Communication can occur anywhere between 24 hours prior to discharge up to 48 hours after discharge.

Regarding a couple concerns from the workgroup call, there was a concern about the lack of evidence for this measure. Again, we mentioned the difficulty in generating this evidence, as well as pointing out that there is strong clinical consensus on this measure in the national guidelines from the UK that we cite, as

well as endorsement from our multi-stakeholder Delphi panel.

In addition, there is randomized control trial evidence from the literature on inpatient to outpatient transitions that supports this process measure for the larger population of patients with complex conditions. We suggested this literature and the studied processes of care from this literature are generalizable to the proposed subset of patients with complex chronic mental health condition who are as much as, if not more so, in need of care coordination and successful transitions of care.

There was also a concern from one member that communication would be unlikely to be documented in the chart, if it had occurred, with an implied concern the providers would not be getting credit for doing this handoff.

A couple things in response to this concern. There are certain aspects of care for which lack of documentation is, in itself, an indicator of poor quality. For instance, it is

standard of care to document a history and physical and to document vital signs that are taken during the course of care. The Delphi panel felt strongly that a core aspect of discharge care for a suicidal pediatric patient was documentation of communication with a follow-up provider. This type of warm handoff communication will just not be as effective if it is not documented was the thinking from the Delphi panel.

And furthermore, from an operational standpoint, we suggest that it will be relatively straightforward to create a structured field for this in an EMR to facilitate that documentation.

Lastly, there was a concern about the exclusion from the measure of patients whose outpatient psychiatrist works in the hospital's outpatient clinic. So, this exclusion was strongly supported and acknowledges the real possibility that it was strongly supported by the Delphi panel and it also acknowledges the real possibility that a patient might be admitted

inpatient into a well-integrated inpatient and outpatient system. For instance, in Kaiser or in an inpatient facility with an outpatient follow-up program, in which providers within the same subspecialty group work closely together and it would be much easier for an outpatient provider to contact the inpatient provider and have access to patient medical records from the hospitalization.

Under these scenarios, the delivery system design is facilitating the care coordination that this measure otherwise would be needed to instigate and, therefore, the exclusion was decided to be included.

Again, we are happy to provide any more details, as asked but those seem to be the highest priority issues from the call.

CO-CHAIR SUSMAN: Thank you very much. Let's turn to the primary reviewers, Jill, Tara, and Craig.

MEMBER MORROW-GORTON: So, I get to be on the hot seat, right?

CO-CHAIR SUSMAN: Yes, thank you.

MEMBER MORROW-GORTON: All right. As we looked at this one, I think we all felt like the warm handoff, the general studies that are not specific to this population, so not kids with harm, not always kids, but did show some evidence that there was improved care. Most of this was based on evidence -- or most of the evidence was based on expert consensus and opinion through the American Academy of Child and Adolescent Psychiatry and the NIH, the National Institute for Health and Care Excellence.

I think that when we looked at this, we had the concerns about some of the exclusions and the assumption that communication is better within an institution than without. And there also was a significant performance gap in terms of very low rate of communication, which could have been that it wasn't documented in the chart but also could have been that it didn't happen.

Do you all want to add anything?

CO-CHAIR SUSMAN: Other comments?

MEMBER BRISTOL-ROUSE: I would just add that this is, even though there isn't a lot of evidence, as Jill said, for this specific population that this is an extremely patient-centered measure. So, while it may not be translating into some clinical or financial outcomes, that it is, across the board, important to patients and families.

CO-CHAIR SUSMAN: Ricardo and then David.

MEMBER QUINONEZ: So our hospital sees 75 percent Medicaid. That is our population. And a large percent of those patients cannot identify a primary care provider. And so that is one problem that I see.

The other one is that communication varies a lot. There has been a lot of publications in the hospital medicine literature about trying to standardize how we communicate with outpatient providers. But one of the things that we have learned is that it is incredibly varied. And not only varied in the way the

hospital is communicating or the hospital provider is communicating but the way the primary care physicians want to have communication given to them.

And so for example, we use various methods and one of them is a patient portal that some of our primary care clinics have access to, in which we put out a discharge summary of the patient and then they can access those records. So that, while technically would not be documented communication in the record, there was very good handoff of patients.

And so those would be two of my main concerns.

CO-CHAIR SUSMAN: Thank you. David.

MEMBER KELLER: So, you mentioned the concern about the within the institution communication problem, that exclusion, and I just have to reecho that.

I, personally, worked in an institution where the psych department decided to blind all the rest of us to any encounter in the

psych department without telling us. So, patients I knew were on psych meds all of a sudden had no record of it. And I screamed and they fixed it.

But I think there is an incredibly wide misunderstanding of the rules regarding confidentiality and psychiatric matters. And that really interferes with this kind of communication in a major way, both within institutions and outside of institutions. And I'm sorry to hear the Delphi panel thought it wasn't a problem within institutions because I absolutely think it is and am interested in their evidence that that is not a problem, besides their own experience.

CO-CHAIR SUSMAN: Well, I think by the nature of the process, it is a Delphi panel. It is, you know I think part of the question, again for this, is this isn't high-quality evidence. There aren't randomized control trials and it is important to provide the feedback.

MEMBER KELLER: I mean I have worked

in three different systems in the last five years and it has been a major problem in all three of them in three different states. So, I just have a hard time imagining that it wasn't perceived by a group of experts.

CO-CHAIR SUSMAN: I'm going to go over to Debbie.

MEMBER MILLER: Can I also get in queue?

CO-CHAIR SUSMAN: Sure, go ahead.

MEMBER MILLER: Oh, sure, this is Marlene. Yes, I was going to agree on that with David about the inter-institutional issues. We have some of those same issues of blinding the data for confidentiality reasons. It makes it difficult.

So, again, I would echo I'm surprised that the committee, the Delphi group didn't discuss it.

But I think the bigger issue with this to me is the documentation and what that is. Making sure that it is effective documentation

actually makes a difference. It is very hard in these types of measures where you just say, you can just put a sentence that I had this communication. That doesn't mean it was heard or you actually talked to the person or was it the quality that actually impacted.

Again, I find myself, I agree with the concept that there should be warm handoff but anytime we break that down to a simple documentation in a chart, usually all of that richness that we are actually looking for gets lost because of just the nuances of documenting and auditing and that kind of stuff.

CO-CHAIR SUSMAN: Thank you very much. Debbie.

MEMBER FATTORI: So, my comment was very similar. Certainly handoffs and transitions of care are critical junctures in the care of any patients and this is a vulnerable group.

But I am wondering from the developers how did you look at the patient record. How did you find this? Because I know in my institution

it sometimes isn't documented or if it is, it is buried in a progress note. So, how did your team and your data collection deal with that issue?

CO-CHAIR SUSMAN: Why don't you respond to that?

DR. MANGIONE-SMITH: Sure. So, at each of the hospitals where this was field tested we had nurse research abstractors who were very familiar with that particular hospital's charting system and went through a training about what we would count as a documentation of a warm handoff. And to Marlene's point, we are pretty lenient. If they say Dr. So-and-So was called, follow-up plan discussed and that is all they wrote, that was given credit.

I mean so it was, in my view, a relatively low bar of documentation we were asking them to look at but they were quite familiar with their own hospital's charting system and I would hope would have known to look for this sort of information.

CO-CHAIR SUSMAN: Thank you. Kevin.

MEMBER AGORATUS: This is Lauren. I also have a question when you get a chance.

CO-CHAIR SUSMAN: Go ahead, Lauren.

MEMBER AGORATUS: Okay, thank you. Sorry, it is hard because I am not there to raise my hand or anything.

CO-CHAIR SUSMAN: Yes.

MEMBER AGORATUS: I agree with the conversation that this is an extremely important measure. I also agree with the comments regarding the confidentiality issues. I am wondering how this dovetails with minor consent to mental health treatment because, again, in some states, that could be as young as 16, even 14. So, I don't know if that is a question for the developers.

I actually heard from families who have begged to have the child stay and the child has convinced everyone they are fine and then they go home and commit suicide.

So, I guess that is my concern is the minor consent issue.

CO-CHAIR SUSMAN: Do you want to respond as developers?

DR. MANGIONE-SMITH: So for this particular measure, to me it would seem even more important that that handoff occur if the scenario occurred that you just outlined. If a child has convinced everybody they are ready to go and they really aren't, I would hope that outpatient follow-up had been a warm handoff. But this wouldn't get into confidentiality of their parents seeing anything because the measure really doesn't address the parents being shared this information, necessarily.

CO-CHAIR SUSMAN: Kevin.

MEMBER SLAVIN: I just had a couple of questions about the abstraction process. I guess the questions really are who would be responsible for documenting. Does it have to be a physician or if it is, I hate the term, but a physician extender, a licensed clinical social worker, what the ongoing treatment plan is, would they be acceptable as far as the documentation? And I

think it has already been mentioned that just because somebody says I spoke with so-and-so doesn't necessarily guarantee that the quality of the discussion and that the information has been received on the other side.

And then I had a question about the exceptions. And this is going to sound really bad but in the area where I practice and this is, again, dealing with mostly medical stuff, we have large group practices where one physician rounds in the hospital, the patient is followed by technically the same group and the same set of providers in a different office, and a lot of times the ongoing care plan is not discussed between people in the same group.

So, I had some concerns about the exception for, and wanted some clarification on the exception for same practice, same physician, same group. Because when you are talking about that kind of communication, if somebody is going to a different office that is ten miles away and nobody has spoken with each other, they may still

be considered to be excluded from that when in actuality, the discussion hasn't really occurred with the person who is going to do the follow-up.

DR. BARDACH: So, the abstraction instructions, just in terms of who that person is, they are instructed to say the hospital, indicate one of the hospital providers communicated by telephone or email with a follow-up provider, which is either a PCP or a psychiatrist.

So, it is relatively nonspecific but focusing on the idea that it is somebody who is part of the care team who is communicating with the follow-up provider.

DR. MANGIONE-SMITH: So, I know there is a lot of angst around sort of letting people off the hook if it was what was considered an integrated system by the people on the Delphi panel but the problem was is that we had a person on the panel who said so you mean to tell me that if I saw this kid and took care of them as their psychiatrist in the inpatient setting, and I am

the one they are following up with, I have to get in contact with myself to get credit. So, that was the genesis of this exclusion.

To be completely honest with you, we were going to say it didn't matter if you were in the same system, that some kind of handoff had to occur with whoever was following you outpatient.

And I imagine that we could have written it more strictly saying you were only excluded if it was the same provider and maybe that is what we should have done, rather than saying provider or provider in the same institution. But it is what it is.

CO-CHAIR SUSMAN: Okay, so let's go down this side. Carol.

MEMBER STANLEY: So, can you talk in a little more detail about what you could as a numerator hit? So, if the hospital were to just fax over some notes or evidence of what has happened and it lands in a physician office fax and someone picks it up, I mean with the algorithm, where did you accept -- what did you

specifically accept as numerator hits?

DR. BARDACH: Thank you for the question for clarification. We actually had a big discussion about whether fax was acceptable or not and it was decided it was not acceptable. So, it is not included in the abstraction instructions. It had to be telephone or email.

MEMBER STANLEY: So, if anybody from the hospital made the phone call and talked to anybody at the physician office, that would count?

DR. BARDACH: Yes, the PCP or psychiatrist follow-up provider was the -- anybody in the hospital, the hospital provider on the care team for the patient making a phone call or an email to the outpatient PCP or psychiatrist who would be following up.

MEMBER STANLEY: So, does there have to be evidence that an actual clinician spoke with an actual clinician on each end?

DR. BARDACH: So the abstraction tool says indicate one if the hospital provider

communicated by telephone or email with the follow-up provider, PCP, or psychiatrist during the time window of 24 hours prior to discharge or 48 hours after discharge.

So, I think you want it to have been more -- you want to be able to say exactly what the definition was the abstractors were instructed to pay attention, mostly clinician, care team providers.

MEMBER STANLEY: Okay.

DR. BARDACH: Does that help?

DR. MANGIONE-SMITH: Just to be clear, we were not specific about who on the care team in the inpatient setting had to make the handoff but it did have to go to either a PCP or psychiatrist. So, they couldn't simply call say a nurse at the follow-up office and do the handoff to the nurse. It did have to be either a PCP or a psychiatrist.

And one other quick thing I wanted to address with Ricardo, the other thing that we require is that if the child did not have an

identified follow-up provider, we hold them culpable for having identified one prior to discharge in communicating with that person.

CO-CHAIR SUSMAN: Jon.

MEMBER FINKELSTEIN: So, one quick thing. It is funny to me that we allow email but not fax, if we don't say that the criterion is someone had to respond to the email. Just sending an email isn't very warm.

(Laughter.)

MEMBER FINKELSTEIN: But almost more important is we are coming to this on several measures, and I knew we would, the idea that the measure doesn't cast a wide enough net. That in one case, it should be internal systems that have had communication as well.

I think for many of these measures what is important, especially given the state of evidence in pediatrics, the diversity of health systems, especially for these mental health conditions for kids, we are going to have to be narrow enough so that for accountability when we

are measuring something, we know exactly what we are measuring, even if there are things outside those borders that would also be really important. No one is saying they are not important. No one is saying within system communication isn't a big problem. It is. But we might not be able to hold every system accountable in the same way for things outside those narrow borders. And I think the performance is low enough that we have room to move even in the narrow scope.

CO-CHAIR SUSMAN: So, I don't want to cut off discussion but I also would like to remind us we have a full group of measures. So, for those of you who have your cards up, we will get to you. Please make your comments quickly and let's not repeat.

So, with that, Virginia.

MEMBER MOYER: I think that I have a similar to Jon's except I have the opposite conclusion from the same data, which is that I think this is clearly one of the most important

things we can do for patients is to make sure that they get appropriate follow-up and that the information be there. But our ability to measure whether that has happened, it seems to me, is extremely poor.

And so we exclude email when we send all our faxes by email. So, our fax would count because our fax goes by email. And so the measure -- the problem isn't that that was a bad decision, it is that it is almost impossible to make a good decision about how to measure the issue of a warm handoff.

CO-CHAIR SUSMAN: Okay, down the line.

MEMBER BOST: The discussion that I have heard is not necessarily about the appropriateness of the numerator and denominator but about the information not being documented appropriately to actually calculate this rate.

But the developers have said that they believe that is also an important contribution to bringing this measure forward and I would just tell folks to think about that when making --

whether that is appropriate or not when making the decisions about evidence.

CO-CHAIR SUSMAN: Okay. And just remember, we are voting on evidence here.

Debbie, did you have another comment or are you okay? We'll give you that one.

CO-CHAIR SUSMAN: Any other comments? Hearing none, let's move on to the vote about evidence.

MS. CHAVEZ: Okay, we are now getting ready to vote on evidence for Measure 2807: Pediatric Danger to Self: Discharge Communication with Outpatient Provider.

And for those on the phone, the choices are one for high; two, moderate; three, low; four, insufficient.

We are expecting 25 votes for this measure and we have 30 seconds. Okay, ready? Go.

Twenty-one, twenty-two, twenty-four. One more.

CO-CHAIR SUSMAN: Jeff, are you

recused on this one or not? You are recused. Thank you.

MS. CHAVEZ: Okay.

CO-CHAIR SUSMAN: Do you want to go over to the park for a while, take in the Washington Monument?

MS. CHAVEZ: Okay, so seven voted moderate; eight voted low; nine voted insufficient.

CO-CHAIR SUSMAN: So, great developers in the cloud, what does this mean to us? Gray zone?

I mean three and four -- three would not move us forward. I mean seven --

DR. NISHIMI: It doesn't move forward but with nine, they could consider an exception.

CO-CHAIR SUSMAN: I guess I am having a hard time interpreting how you determine gray zone and whether it goes forward or whether because it is low and insufficient, we need to consider an insufficient with exception vote. It is more of a process.

DR. BURSTIN: The measure did not pass on evidence. It is like 70 percent against. But your next decision would be determine whether this is a measure you would like to potentially consider for exception.

CO-CHAIR SUSMAN: Okay. So, clarified now, we are going to have the vote about insufficient with exception. Remember, that is based on a preponderance of evidence being or the thought being that this would be a positive thing, there wouldn't be unintended consequences. One being insufficient evidence with exception and two, no exception.

MS. CHAVEZ: Okay, read.

Fifteen, twenty-two. One more.

CO-CHAIR SUSMAN: Keep voting.

MS. CHAVEZ: Twenty-four.

So, 14 voted for insufficient evidence with exception and 10 voted for no exception.

CO-CHAIR SUSMAN: Okay, so I guess that is another gray zone. We continue on forward. Let's see if we can pick up the pace

and go for opportunity for improvement or gap.

So, is there evidence for gap in this measure? Okay, yes.

MEMBER KNUDSEN: Yes, there is. They provided some performance results for this measure using some data N of 177 over two years, three hospitals and only 20.5 percent of the hospitals actually recorded this happening. So, that is a significant gap.

MEMBER KNUDSEN: There were no statistical differences in terms of disparities when they looked at this as well.

CO-CHAIR SUSMAN: Thank you. Virginia, did you have a comment? No. Okay.

Any other discussion about gap? If not let's turn to voting on gap.

MS. CHAVEZ: Again, one for high; two, moderate; three, low; four, insufficient. Ready, go.

Nineteen. We have twenty-one, twenty-two.

CO-CHAIR SUSMAN: Someone has probably

got no battery.

MS. CHAVEZ: Twenty-three. One more. Three seconds. Okay.

So, nine voted high, thirteen moderate, and one voted low.

CO-CHAIR SUSMAN: Reliability testing. Comments from our primary reviewers about reliability of this measure?

MEMBER MORROW-GORTON: So, I think the reliability testing they did in rate of reliability had a high kappa and an ICC testing at the hospital level that was also relatively high. And there were 117 records, so it was a fair number of records that were looked at.

So, I think we thought it was fairly reliable.

CO-CHAIR SUSMAN: Okay, any more comments about reliability? Kerri.

MEMBER FEI: I have another denominator question. Earlier, the denominator said it was patients discharged for self-harm. Is it actually they are admitted for a self-harm

or suicide diagnosis and then subsequently discharged after treatment?

DR. MANGIONE-SMITH: Right, so the denominator population is identified using E codes and V codes for suicidality.

MEMBER FEI: Okay. All right, thank you.

CO-CHAIR SUSMAN: Okay, unless there are other comments about reliability, let's move to vote.

MS. CHAVEZ: Okay, voting on reliability for Measure 2807. Ready, go.

Twenty, twenty-two, twenty-three. Thank you.

Okay, six voted high, fifteen voted moderate, three voted low, and zero for insufficient.

CO-CHAIR SUSMAN: So, moving on to validity. I think we have talked a lot about validity overall. Anything new to say or bring up specifically?

MEMBER KNUDSEN: I think that it is

important to note that there were no statistical significant differences between those meeting or those failing the measure in readmissions or ED visits. So, I think that is really important.

CO-CHAIR SUSMAN: Okay, any other comments? Kevin did you have one?

MEMBER SLAVIN: And I'm not sure how other institutions work but one of the things I think that may affect the validity is at what point after discharge does the hospital close the EHR for documentation purposes. Because if you are allowed up to 48 hours after somebody is discharged to send off and to document but if the hospital closes the EHR before then so that they can get the coding and billing done as quickly as possible, documenting that is going to be difficult.

CO-CHAIR SUSMAN: Please.

MEMBER MORROW-GORTON: I think in our small group discussion, we also had a discussion about the kids that left AMA. And given that they were in the hospital for danger to

themselves or suicide, that excluding them from this might be problematic as well.

CO-CHAIR SUSMAN: Thank you. Any further validity questions?

Let's move forward, then, to vote.

DR. BARDACH: Can I?

CO-CHAIR SUSMAN: Oh, yes, please.

DR. BARDACH: I was just going to offer one piece of information, which is that in our testing there were zero people who met that exclusion of leaving AMA.

CO-CHAIR SUSMAN: Thank you. Okay, voting on validity.

MS. CHAVEZ: You are voting on validity, one, two, three, four options.

Eighteen, twenty-two, twenty-four.

So, 12 voted on moderate and 12 voted on low.

CO-CHAIR SUSMAN: So, I guess that is gray zone again. And we will proceed on.

So, let's go ahead to feasibility. Is this easily collectable data? Is it feasible to

get the measurement? Any measurement concerns?

Yes, so Virginia.

MEMBER MOYER: This is for the developers. You used pretty well-trained nurse extractors. I looked at this and I looked at all the questionnaire and everything and I wondered how challenging this would be in the setting of not well-trained nurse abstractors, but people who perhaps a little less well-trained. How usable is this outside of that? Did you do anything with that, with evaluating how hard it was to train them?

DR. MANGIONE-SMITH: So, the feedback we got, with the caveat that they are all experienced abstractors was that the tool we designed, the electronic data collection tool was actually quite easy to use.

The person who designs the tool to collect the information specifically designs it so that abstractors don't have to think very hard and they never have to make a subjective judgment. So, and many times they don't even

know what measure the data they are putting in is feeding into. So, and I think the average abstraction time, I don't know if you have that recorded down, but for the entire tool, which collected many more measure than just this one, was well under an hour. It was more like 30 minutes to collect the data for several different measures.

So, our feedback was that it was quite user friendly and easy to use. But again, in the hands of somebody who is less experienced, I don't know the answer to that, obviously.

MEMBER BOST: -- what percent were required reviewing the notes by the trained nurse versus actually being able to collect?

DR. MANGIONE-SMITH: So, I know for a fact in two of the hospitals there was a designated field. They could look for whether there was documentation of a call to the follow-up provider. But in three of the hospitals I don't know whether they had to actually look into notes or not. It is possible that they did.

CO-CHAIR SUSMAN: Okay, any other questions about feasibility? If not, let's vote.

MS. CHAVEZ: Voting on feasibility. One, high; two, moderate; three, low; four, insufficient. Go.

Nineteen, twenty-two, twenty-three, twenty-four.

Zero voted high, 12 voted moderate, 12 voted low, zero voted insufficient.

CO-CHAIR SUSMAN: This is a tale of gray today.

All right, let's move on to usability. Any questions about usability? This is not currently in use?

Kevin.

MEMBER SLAVIN: Just in terms of the types of communication that are allowed and HIPAA compliance. I'm not sure -- you know email is allowed. Not all email is secure. So, there is some concerns about accessibility to information, depending on how it is communicated.

CO-CHAIR SUSMAN: John.

CO-CHAIR BROOKEY: I just want to make a comment about so we have a lot of electronic processes, one of which is putting in default items and smart sets and smart phrases. And a measure like this could potentially allow somebody to put in the right phrase, pulling it into a smart set without actually having not done anything related to what they said.

So, I just wanted to say that pulling the data out maybe problematic, even in an EHR.

CO-CHAIR SUSMAN: Carol.

MEMBER STANLEY: Can you talk a little bit about -- when I look at potentially using this measure, to me it is more of a notification, a measure of notification and not really a warm handoff. Can you explain how you decipher between a warm handoff and just a notification?

DR. MANGIONE-SMITH: So, to me notification would be more like the fax being sent or a discharge summary being sent to an outpatient provider. A warm handoff, at least the way we operationalized it was there need to

be a conversation either by email, which certainly there could be concerns about security where that is concerned for confidentiality reasons, or a documented telephone call.

CO-CHAIR SUSMAN: Virginia, did you have a question? No, okay.

Any others on usability? Let's go ahead and vote.

MS. CHAVEZ: Voting on usability. One, high; two, moderate; three, low; four, insufficient. Go.

Twenty-one, twenty-four.

One voted high, ten voted moderate, twelve voted low, one voted insufficient.

CO-CHAIR SUSMAN: Okay, still a split here.

And let's go ahead and vote on the measure. If there is other comments, final comments, anything additional to add.

Okay, so this is an up/down. One, yes; two, no.

MS. CHAVEZ: We are now voting on

Measure 2807, acceptability for endorsement.

Ready, go.

Twenty, twenty-two, twenty-three, twenty-four.

Ten voted yes and fourteen voted no.

CO-CHAIR SUSMAN: Okay, again, very close.

Any comments or suggestions to the developer here? Yes, please, Dave.

MEMBER EINZIG: So, it just kind of feels like sort of a clunky measure. It reminds me of driving a car with square wheels, in a way. I mean it kind of gets you there but it is not going to work right. You know just kind of a bigger picture of until we get a universal healthcare record, I mean it is just going to be kind of clunky along the way until we get there.

CO-CHAIR SUSMAN: Other feedback? Is this something you would want to see in some revised form? And if so, what would it look like?

MEMBER MILLER: This is Marlene. I

would just encourage the developers to really think about what documentation means, both what is acceptable and what has to be in it so that it is actually more likely a meaningful conversation, a meaningful handoff and not just the smart phrase comments. What really struck home with me is how easy this would be to almost game with a smart phrase.

CO-CHAIR SUSMAN: I think the handoffs, internally, at least for me, would be important to incorporate. I think the exception that you were trying to get to wasn't handled as artfully as it might be.

Any other comments to the developer?

Okay, we now have time to ask for public comment. Robyn, do you want to?

DR. NISHIMI: Right. So, if there is anyone here on-site or, operator, if there is anyone on the line who wants to give public comment.

OPERATOR: Okay, at this time, if you had wanted to make a comment, please press * then

the number 1 on your telephone keypad.

There are no public comments from the phone line.

CO-CHAIR SUSMAN: Okay. Well, this has been a lot of work. I appreciate very much the efforts of our measure developers. This was a yeoman's work, so to speak. And I hope we see revised measures coming to us.

Certainly, the ideas that are being incorporated are certainly the direction. And for a variety of reasons, it is very difficult in this area to come up with really a well-tested, valid, evidence-based measure. So, don't get discouraged.

All right, this brings us to lunch. And I'm not sure. Do we intend to have a working lunch or how do you all suggest we go forward? Okay, it is up to us.

We are behind by one measure. So, why don't we -- it's about ten after. Why don't we take about a 15-minute break to get started and then about 25, 30 after, I am going to ask that

we reconvene and get ourselves going, try to get back caught up.

So, thank you.

(Whereupon, the above-entitled matter went off the record at 12:08 p.m. and resumed at 12:31 p.m.)

CO-CHAIR SUSMAN: The first of the ADHD measures is 2817, accurate, ADHD diagnosis from AHRQ, CMS, and I guess God or someone else. And we'll start out with any recusals. Okay, hearing none, let's turn it over for a brief explanation from our developers.

DR. WOODS: So statistics provided by the CDC, 5 million children between the ages of 4 and 17 have been diagnosed with ADHD. This is increasing and the rates of ADHD diagnosis increased 5.5 percent per year from 2003 to 2007. Validated tools based on DSM criteria have demonstrated effectiveness for diagnosing ADHD -- for distinguishing ADHD from the diagnosis of other conditions. When less rigorous methods are applied to the diagnosis of ADHD, the positive

existence of the condition may be missed, leading to potential social and academic struggle.

In November of 2011, the American Academy of Pediatrics published a new evidence-based guideline for ADHD diagnosis follow-up and treatment based on extensive review of the existing evidence. One recommendation with a high level of evidence indicated that when diagnosing ADHD in children 4 to 18 years of age, primary care clinicians should determine the DSM criteria have been met, including documentation of impairment in more than one major setting, with information obtained from reports of parents or guardians, teachers, other school mental health clinicians involved in the child's care.

To make a diagnosis of ADHD -- so there is a -- this new guideline and there are several recommendations within this guideline, this guideline recommendation is, and I'm reading verbatim, "to make a diagnosis of ADHD, the primary care clinician should determine that Diagnostic and Statistical Manual of Mental

Disorders criteria have been met, including documentation of impairment in more than one major setting. The information should be obtained primarily from reports, from parents or guardians and teachers and other school and mental health clinicians involved in the child's care. The primary care clinician should also rule out any alternative cause."

The evidence is grade B with strong recommendations which is defined as RCTs or diagnostic studies with minor limitations, overwhelmingly consistent evidence from observational studies. This level of evidence is based on high to moderate quality scientific evidence and preponderance of the benefit over the harm.

The gaps in care include that -- a survey done by the APA regarding the guideline, 91.5 percent of physicians were familiar with the guideline recommendations. However, an additional study found that approximately 50 percent of children with ADHD seen in practice

settings obtained care that matches the guidelines of the American Academy of Pediatrics.

The pathway between the process measure on the outcome, ADHD diagnosis increases appropriate treatment, decreases inappropriate treatment, improves quality of life and improves care. In order to work on a measure in this space, we engaged a technical expert panel, 25 experts and stakeholders that included psychiatrists, psychologists, nurses, school nurses, school psychologists, pediatricians, developmental pediatricians, social workers -- did I say developmental pediatricians -- also parent and patient stakeholders, other school stakeholders. And this particular recommendation stuck out to this group as something that would be very impactful in the lives of children to make sure that an accurate diagnosis is made according to the appropriate criteria.

There are validated tools that use the DSM criteria, so they recommended that the measure look at validated tools or actual direct

assessment of the criteria through reports given to the physician.

The measure itself -- the denominator criteria are all patients aged 4 through 18 with a diagnosis of ADHD. And the numerator criteria are patients whose diagnosis of ADHD was based on a clinical exam with a physician or other health professional as appropriate, which includes confirmation of functional impairment in two or more settings and assessment of core symptoms of ADHD including inattention, hyperactivity, and impulsivity either through use of validated diagnostic tool, based on DSM-IV criteria or for ADHD through direct assessment of the patient.

We tested this measure as a chart review measure in the primary care networks of four hospitals in the Chicago area, a teaching hospital, two safety net hospitals, and a suburban hospital. And we also from that testing worked with the American Board of Pediatrics and they have incorporated the specification of this measure into their maintenance of certification

Part 4 payment program. And since then, 313 physicians have used this measure as a measure for improvement which they generally have to do 100 pre- and 100 post- actual patients.

Should I respond to the questions here, too? There were a couple of questions unresolved at the end of the conversation.

CO-CHAIR SUSMAN: Why don't you try to hurry it up so --

DR. WOODS: Okay, I just wanted to know if I should give you everything right now?

CO-CHAIR SUSMAN: No.

DR. WOODS: Okay.

CO-CHAIR SUSMAN: I really appreciate it, but given the pressures of time, I'm going to turn to our group of -- Martha, were you on --

DR. WOODS: Can I say one more thing, just one more thing?

CO-CHAIR SUSMAN: One more thing, okay. We'll let you -- one more.

DR. WOODS: That these measures were also vetted through a public comment period where

we did active recruitment of comment.

CO-CHAIR SUSMAN: Thank you. Sounds like it was very thorough.

Martha?

MEMBER BERGREN: So the call that we had did think that the people on the phone did believe that this was a very important measure, but there was a lot of discussion about the measure and almost every component of the measure. So the reason that it's so important, just to echo what you said is the implications of false negatives and false positives. Both have significant implications for the children that are either diagnosed correctly -- incorrectly, or not diagnosed when the diagnosis is present.

So one of the --- there were a ton of issues with the numerator with many of the people on the call believing that having actual disagreement with the DSM criteria that perhaps not all three symptoms are needed to have ADHD and the implications of that on what the results would be.

I want to make it clear that there was a lot of discussion about what constituted meeting the numerator. It can be either using a validated tool for the symptoms or using direct clinical assessment where the DSM criteria are used to basically evaluate those same criteria.

And then the denominator is all children between 4 and 18 years old with a diagnosis of ADHD. And the diagnosis has to have been within the previous year from the visit. So the evidence is a recommendation and as you said, it's grade B evidence, based on RCTs and diagnostic studies and graded as strong.

The performance gap is present and there is a performance gap within ethnicities with 55 percent of African-American and Hispanic patients meeting the criteria compared to 81 percent of white patients.

For the reliability, I don't see my notes on the reliability.

DR. NISHIMI: We can just discuss that when we get to the reliability section.

MEMBER BERGREN: Sure. Okay.

CO-CHAIR SUSMAN: Any other major points for the committee?

MEMBER BERGREN: Oh, I'm sorry, do you mean the other people?

CO-CHAIR SUSMAN: No.

MEMBER KELLER: The concern about reliability was that reliability testing had only been done on the numerator portion of the measure and we were wondering why there had been no reliability -- we didn't see reliability testing of the denominator.

DR. WOODS: We did do reliability testing of the denominator -- that diagnosis. In order -- we pulled charts and then assured that the diagnosis was present in the chart. And we excluded then any -- actually, I guess you're right. We didn't present how many we excluded. We didn't exclude many, but we did exclude some.

MEMBER KELLER: It was a concern. We were wondering why we hadn't -- we didn't see that data.

DR. WOODS: We can get that for you, but we didn't know you were interested in it.

CO-CHAIR SUSMAN: Okay, any other major comments overall from those who reviewed this closely?

Well, then let's go ahead and concentrate on evidence. This is a process measure. It has gone through systematic reviews, some RCTs linking process with the outcomes of note. Any questions, comments?

I know you've indicated, Martha, some of the concerns about specification. We probably are not going to rewrite DSM IV or V at this committee meeting, so take it for what it's worth. It must be the post-prandial slump.

MEMBER KELLER: I'll jump in. I think the biggest concern in the numerator statement was the clinical exam with a physician or other healthcare professional because it just seemed relatively straight forward to document and to count what measure -- what standardized tools were used. If you're using a Vanderbilt, you can

find a Vanderbilt in there. But it wasn't clear to us exactly what counted as a clinical exam with a physician that would be adequate for the diagnosis of ADHD having done a number of -- having seen a lot of external records sent to me from kids who have been diagnosed by other physicians. Because I manage ADHD a lot in my practice, there's a fair degree of variability about what goes into those exams and so we were wondering how standardized, how reproducible that particular piece would be.

DR. WOODS: So in the chart abstraction tool itself, really the way that they were instructed to abstract it was to first identify the date of the ADHD diagnosis and then evidence of ADHD diagnostic, clinical exam by physician in the chart. So any evidence, evidence in the chart of assessment for symptoms of ADHD including inattention, hyperactivity, impulsivity, to evaluate a diagnostic tool. Evidence in the child consists of core symptoms of ADHD including inattention, hyperactivity,

impulsivity, certain things. They were instructed to look broadly on effectiveness on a visit-based assessment.

MEMBER KELLER: One other issue that arose was the question that is done in a number of parts of the country where different parts of the ADHD assessment are done not on a single encounter, where people do an assessment --

CO-CHAIR SUSMAN: I think we have cross calls here.

DR. WOODS: Operator, you need to close your line.

In the conversation on the phone, it had slipped my mind that that was the way we instructed the chart abstractors to account for that very thing. So I apologize.

CO-CHAIR SUSMAN: That's okay.

MEMBER KELLER: That's why we get to ask the question again. Thanks.

CO-CHAIR SUSMAN: Ricardo.

MEMBER QUINONEZ: So I just have a question about the evidence, graded as B. Being

a little bit familiar with the ADHD evidence simply because I'm very interested in over diagnosis, I don't remember and maybe you can educate me as to actual RCTs that linked inaccurate diagnosis of ADHD with bad outcomes. I remember most RCTs are whether treatment works or not. So I mean is there direct evidence to cite such a high level of evidence that complying with this measure would improve outcomes?

DR. WOODS: Okay, so I can provide you with more information, but one of the things that I think is particularly concerning, there are kind of two things that are particularly concerning. One is when another type of mental health diagnosis is described as ADHD. They will get potentially stimulant medication that could exacerbate the symptoms of this other condition, if not properly diagnosed. So that's a fairly bad outcome.

The other is in the age of stimulant medications. One of the things that happens is kids will show up thinking that they can get more

focused or get some drugs and they'll show up and try to get ADHD medications and if you don't do a very systematic diagnostic process, they can pass. So those were -- that's the nature of the literature. I can try to get you specific citations if that would be helpful.

CO-CHAIR SUSMAN: So it sounds like more anecdotal than from RCTs?

DR. WOODS: No, I'm just saying that off the top of my head I can't list off the 12 studies. But let me -- while you're discussing, let me look in the guideline.

CO-CHAIR SUSMAN: All right, are there -- yes, Virginia.

MEMBER MOYER: To add a little bit to what Ricardo is saying, there's clearly adequate evidence. In fact, high quality evidence for treatment, appropriate treatment. And evidence that if you don't treat, it's not good for the kid.

What we don't have, and I know the literature well enough to know that it's actually

not there, is evidence that the -- there's not trial which you wouldn't expect, but not even observational evidence about the misdiagnosis. And that, I think, is what we're looking for. We're not looking for randomized trial of accurate diagnosis. We're looking for a study that would tell us how frequently inaccurate diagnoses occur.

CO-CHAIR SUSMAN: Okay, well, we'll wait for a response from our developers while we're further discussing.

Other points around the evidence here, that this is a measure that's going to -- by having a structured approach to ADHD diagnosis, it's going to lead to improved outcome in the patients? That's a process measure here.

I don't see a lot of more questions or input, so I'll give you just another second or two here before we decide to vote.

Yes?

MEMBER FATTORI: Just a question for the group, particularly those who -- I appreciate

the comments that were just made, but doesn't treatment, effective treatment rest on the fact that you have an accurate diagnosis?

CO-CHAIR SUSMAN: So there's some shaking of head, yes and no. Let's take pro and con here. Virginia?

MEMBER MOYER: Absolutely. I mean that's sort of a first principle. But what we don't have is documentation that inaccurate diagnoses are leading to bad outcomes. We think they would, but we need documentation that that's happening and we also need documentation that using this approach provides you, makes the patient have a better outcome. That's what we don't have. It's not that there's not a theoretical reason --

DR. WOODS: So there actually is an interesting new study in Nature and Neuroscience last month that was looking at the ability to focus and doing functional MRI scans for the ability to focus. And they were able to find a particular pattern map, a signature for focus.

And then they prospectively got scans of children who had had the standardized tools, both positive and negative, and prospectively predicted very well the ability of the testing. More to come, but that's an interesting little piece.

CO-CHAIR SUSMAN: David, do you want to add to the conversation?

MEMBER EINZIG: Yes, just a clinical perspective. So people can have, obviously, trouble with concentrating for a lot of reasons, fetal alcohol, autism. There's a lot of other variables. So I'd be less worried about using these forms to diagnose ADHD accurately for the purpose of appropriate treatment with a medication because medications may be appropriate even if you don't have ADHD. But more for the worried, well, the people who are trying to get into their Ivy League colleges and get artificial advantages where this might provide more useful information.

CO-CHAIR SUSMAN: Okay, any further comments? I'll take one further comment from our

developer, so make it good.

DR. WOODS: A question to you?

CO-CHAIR SUSMAN: I thought you had something further, your colleague was indicated, but if not, we can move on.

DR. WOODS: I had wanted to comment about the issue of finding all of the elements in one visit. That was the last final thing I think I was going to tell you.

And as has been already mentioned, it is highly problematic to begin a treatment for ADHD on a child who does not have ADHD. And so therefore, it's very problematic and an important measure. And there's a considerable gap in this at this time.

CO-CHAIR SUSMAN: Okay. I'm not seeing or hearing a lot of further discussion on evidence. Yes, David?

MEMBER KELLER: So the only other comment was about the performance gap that was identified that that was based on a practice survey of four clinical sites in the greater

Chicago area and we were -- the committee wondered whether there might be more. I expect there would be more variation if you did a broader sample, but that's a fairly rarified sample on which to identify a performance gap. I didn't think that was a huge sample.

DR. WOODS: So the performance gap is really identified through the literature and what we found was consistent with the performance gap that exists across the country. And this is a national standard, so wherever you find it, you find it. I mean you should find it. But the performance gap is in the literature. Does that make sense?

CO-CHAIR SUSMAN: Let's go to vote on the evidence. We've had a nice discussion of what this measure does and doesn't do and I would turn it over.

MS. CHAVEZ: Okay, we're now getting ready to vote on evidence for measure 2817, accurate ADHD diagnosis. And for committee members on the phone, the enter options are 1 for

high; 2, moderate; 3, low; 4, insufficient. Okay? Ready, go.

CO-CHAIR SUSMAN: Remember to click toward our esteemed NQF colleague and click more than once and think good thoughts.

MS. CHAVEZ: Twenty-one, 23.

CO-CHAIR SUSMAN: Are we trying to get 25?

MS. CHAVEZ: Twenty-four. I think 25. Okay, so 2 voted high; 16 voted moderate; 5 voted low; and 2 voted insufficient.

CO-CHAIR SUSMAN: Okay, and we've had a lot of discussion about gap and documentation or not of that. Any further comments on gap? If not, perhaps we can go on to vote on gap.

MS. CHAVEZ: Okay, moving on to voting for gap, same enter options, 1, high; 2, moderate; 3, low; 4, insufficient. Ready, go. Eighteen, 23, 24, 24, 24, 25. Okay, 3 voted high; 21 voted moderate; 1 voted low; and none for insufficient.

CO-CHAIR SUSMAN: Okay, reliability.

Any questions or further comments about reliability?

MR. FINKELSTEIN: I just wonder if the developers want to say anything about the variation in the Kappas across the elements which is pretty striking.

DR. WOODS: So on reflection, I've been actually in the measure development business for about 20 years. In 1995, I was part of the initial measure development. When I thought about it, generally things that are not monitored and tracked do not have good standard methods for documentation. There's under documentation. There's documentation in a lot of places. When something is monitored and tracked, people streamline that pretty readily, sometimes over a couple of years or more.

When I looked at the results, I thought to myself, well, this abstractor found this, but that abstractor didn't find it. They didn't look maybe as hard. I don't know. Clearly, from what we -- so we went back. We

looked at it. And that was kind of what we found. People were looking in different places and not finding things or finding things based on where they were looking because we had them -- the abstraction tool asked exactly -- one of the elements was where did you find it?

So that's what I would say about it, the natural life course of a measure. And also, probably it's probably appropriate to say for mental health measures, so the mental health documentation in two of the practices that we looked at were not on electronic medical records yet. So there's a diffusion curve that exists.

CO-CHAIR SUSMAN: So what are the differences in reliability based on the construct used to define ADHD? So using the Vanderbilt versus clinical assessment, for example?

DR. WOODS: We did not see much -- we did not see that one reviewer was finding a lot of direct assessment and the other was not, if that's what you're saying.

CO-CHAIR SUSMAN: Well, I'm thinking

about, for example, the operationalization of a clinical assessment of ADHD according to DSM criteria might be subject to variation based on the ability to pick up documentation, the interpretation of the elements which would go into assessing things that are in the three domains of ADHD.

DR. WOODS: So we only saw two cases of the direct assessment where people found the practice meeting the measure. They found it either with a diagnostic tool that was either validated or not validated.

CO-CHAIR SUSMAN: Yes?

MEMBER MORROW-GORTON: One of the things that we discussed in the small group was that typically when you make a diagnosis of ADHD, you're not using either clinical history, what you're calling observation --

DR. WOODS: No, assessment.

MEMBER MORROW-GORTON: -- or checklists. You're generally doing both and using them from multiple places so that you get

them more than one location.

DR. WOODS: So we're getting the diagnosis, the date of the diagnosis, and we're looking for evidence in the chart of those elements, a validated tool, a physical exam, and the symptoms and impairment in more than one setting. And we're looking for any evidence.

It is not necessarily -- I should have clarified that on the call, but all of our instructions and the abstraction tool indicate those things and they're given -- each of the elements that would be included in the validated tool are included as individual items for the direct assessment.

CO-CHAIR SUSMAN: Okay, Kevin.

MEMBER SLAVIN: One of the questions that came up on the call and that it would be -- I think part of the reason why the denominator reliability information would be useful is there were questions about patients who were diagnosed elsewhere who come into a new practice and it's the first diagnosis of those tools and that

assessment is not going to be available within that patient's chart, at least where it's being looked for. So I think in terms of the reliability, the denominator information would be kind of useful for those types of issues.

DR. WOODS: So we also went back to the -- and met with the chairs of our expert technical panel, Mark Wolraich and Karen Pierce, and presented that question to them. And their assessment -- they work in different kinds of practices, so they had similar opinions, but about different kinds of practices. And their perspective is that generally when a child is moved from one clinician to another, there should be passing forward of this information because the pediatrician is then responsible for the school accommodations, for specific treatment, and one other thing which I can get for you.

So they thought that it is standard of care that the new physician should be receiving that information or if they don't get it, they should be looking at doing another assessment.

That's their clinical opinion about how this should play out.

MEMBER SLAVIN: But I guess the question would be is would that information include the full diagnostic assessment of the child who has been diagnosed or just this is what the child's current needs and the recommendations for the on-going management of the ADHD are?

DR. WOODS: Well, their opinion was that the information should be gotten from the sending or the leaving clinician, that there should be -- that appropriate care is really that the pediatrician who is managing the patient should have a criteria-based understanding of what their condition is. And they should get it, however, they can get it, but they most often will get it from the practice that the child is coming from.

CO-CHAIR SUSMAN: So I've got David and we'll just go up the aisle here. David E.

MEMBER EINZIG: I just wanted to make sure I'm understanding correctly. So we've got a

child who is diagnosed with ADHD at the age of five. Moves to Washington, D.C. Is 16 now. Has been stable on meds. So the expectation is that the new provider tries to obtain that document from when they were five? And if they don't get it, they do it again?

DR. WOODS: It's unlikely that that would have been their only assessment because -- and in fact, another recommendation which we'll talk about in a minute is that there should be regular follow-up and reassessment of children with ADHD, chronic care follow-up.

CO-CHAIR SUSMAN: Okay, I'm going to try to wrap this up.

David, please?

MEMBER KELLER: I keep -- I hate to keep hammering on that point, but I think the reality is that you sometimes can get those records and sometimes can't and as a practitioner, you then have to decide how you're going to handle it. Most of us probably wouldn't do what we would consider a full intake

evaluation on that child. We would probably do some sort of grading scales with multiple observations, but not the rest of the assessment that goes into deciding if there are any other diagnoses going on. We would just try to get an assessment of functional status basically to try to make sure that we can keep treating the child because the family is going to be interested in not changing a lot, and that's what they're going to be looking for.

I mean it's measuring a different problem because then the problem isn't one of a physician not doing an adequate assessment. It's a physician not being able to get records. And I'm wondering if that was the intent, to kind of conflate those two issues within this measure.

DR. WOODS: No. I'll go back to -- as I'm hearing you, you're going to be interested in how this child is actually functioning. You're going to be interested in understanding whether the medication, treatment, or behavior therapy treatment is actually managing the ADHD symptoms

and what symptoms are falling out of that.

So I mean the idea is there ought to be criteria-based understanding of the diagnosis of the child. And there are a lot of ways to get that. But that's the standard of care and that it should be based on the DSM criteria.

CO-CHAIR SUSMAN: Okay. Let's move on. Any new thoughts about reliability? If not, let's vote on reliability. And we have the data concerning Kappa. We have some information that's been described about the potential reliability of the denominators. One, high; 2, moderate; 3, low; and 4, insufficient.

MS. CHAVEZ: Voting on reliability for 2817. Ready, go. There's 23, 24, 25. Okay, 0 voted high; 12 voted moderate; 10 voted low; and 3 voted insufficient.

CO-CHAIR SUSMAN: Okay, let's move on to validity. Was there any testing of validity in this measure?

Martha?

MEMBER BERGREN: So there is a 25

person expert panel which was described and it was considered case validity and there was not any data associated with that assessment.

CO-CHAIR SUSMAN: So by definition, this would be low or moderate?

MEMBER BERGREN: Yes.

CO-CHAIR SUSMAN: Or insufficient. I'm sorry.

DR. WOODS: We did provide data about the face validity in qualitative form and also our public comments on --

CO-CHAIR SUSMAN: But this wasn't tested out in the field, correct?

DR. WOODS: Correct.

CO-CHAIR SUSMAN: Other thoughts from the group at large about validity?

David, did you have a thought? Failure to turn. You don't have to keep it up like that. You can just talk.

MEMBER MORROW-GORTON: I think one of the conversations that we had during the small group was the sort of diagnostic dilemma around

DSM-IV, DSM-IV-TR, DSM-V. And sort of taking a process and sort of moving it to where people are supposed to be moving which is DSM-V. And likely DSM-V will be like DSM-IV and be around for -- what is it, 20 years?

DR. WOODS: I can respond to that. The changes in the diagnostic criteria for ADHD in DSM-V do not affect the measure. The changes include that you can also assess for autism spectrum as well as ADHD which had not been previously described. And that if there are symptoms of ADHD, initially you had to have symptoms by age 7 and they raised that ceiling to age 12. Those are the only differences which don't affect our criteria.

CO-CHAIR SUSMAN: Virginia.

MEMBER MOYER: Just one brief comment which is that having been on the call, the committee pre-evaluation comments are quite complete and are a good reflection of what happened at that net conversation.

CO-CHAIR SUSMAN: Well, let's go ahead

and then consider validity voting here. So remember, this is moderate, low, or insufficient by definition.

MS. CHAVEZ: Voting is open. Eighteen, 24, 25. Zero voted high; 9 voted moderate; 11 voted low; 5 voted insufficient for validity.

CO-CHAIR SUSMAN: So again, we have this issue of the low and insufficient being the majority. How would you like to proceed, NQF staff? They're doing some gyrations, calculations, but it's going to be 16 versus 9.

DR. NISHIMI: Doesn't pass.

CO-CHAIR SUSMAN: This is a must pass, so this will stop here.

DR. NISHIMI: So just for the record, reliability was in the gray zone, but validity is a not pass.

CO-CHAIR SUSMAN: Okay, thank you very much. Any feedback to the measure developers here before we move on to the next one?

MEMBER KELLER: So I echo what Jennie

said earlier which is that we all think that accurate diagnosis of ADHD is important. I think where we're stuck is exactly what we're measuring and we're concerned. I think some of what you heard and the concern here is that we want people to focus on what's important is going to improve outcomes and not find themselves doing a lot of things just to make the chart look pretty. I think that's underlying a lot of what you heard here, that we want the work that we do to get accurate ADHD diagnosis to matter. And that's what we -- I think that's where we're all struggling in this.

MEMBER MOYER: I would also suggest that more empirical testing would probably have made us feel more comfortable. The face validity and the description which was pretty broad didn't leave us feeling as comfortable as we would like to have been.

CO-CHAIR SUSMAN: I would say the reliability was a question when you start to look at the individual components.

David?

MEMBER EINZIG: Just one more quick comment. And also just to emphasize that sometimes ADHD medications are appropriately used, even without an ADHD diagnosis, so I think that's --

CO-CHAIR SUSMAN: Okay. We have another ADHD measure which is on chronic care follow-up, 2818. I think we've talked a lot about the issues, so if you could confine your comments to specification and why that specification is valid and important.

DR. WOODS: So this is a claims-based measure, also deriving from the 2011 AAP ADHD Guideline that recommends that ADHD be considered a chronic condition and that patients with a diagnosis of ADHD be treated as children and youth with special healthcare needs and that it is very important to treatment appearance to have follow-up visits. This measure is specified to begin a year after ADHD diagnosis. There should be -- the treatment should be managed fairly

frequently in the first year where there should be several visits and phone calls to titrate medication or to assess behavior therapy. But following the year after diagnosis, as a chronic condition, in a medical home, the patient should be seen by a clinician at least yearly.

In the call, we discussed that there may be many other times where there's a phone call or other kinds of communication more frequently and there could be visits, more frequently, but it's at least one visit every year to manage ADHD. That is currently not happening.

In terms of the gaps, data from community-based samples indicate average time to discontinuation of medicine is four months and that families are fully compliant with treatment regimens for an average of only two months.

GPA has been shown to be significantly higher during the treatment adhered marking periods than non-adhered marking periods for Medicaid-eligible children diagnosed with ADHD.

So it is this chronic care follow-up is guideline-based recommendation which should give it some evidence, strong evidence and it is actually graded B and strong in the guideline itself. And patients with ADHD who receive follow-up visits are more likely to receive treatment which, in turn, improves function, quality of life, and reduces symptoms.

There are no unintended consequences from this measure. However, without this measure, negative consequences may occur including poor treatment adherence, ultimately resulting in decreased function and quality of life.

CO-CHAIR SUSMAN: Okay. Very good. Thank you.

And Virginia, Jill, Kevin, comments, high points in our consideration here?

MEMBER MORROW-GORTON: I think we actually were pretty interested in this measure thinking about ADHD as a chronic condition, thinking about follow-up. We did have some

conversation about whether or not one visit was adequate, although if you think about it, if you have somebody that is followed by a sub-specialist or who doesn't tolerate medication, there was some question about whether that extra visit would be a burden. I think that's probably not the case. So I think we were pretty positive about this in general.

CO-CHAIR SUSMAN: Other comments, Kevin, Virginia?

MEMBER MOYER: So one of the concerns that we had is that the evidence is inferential. It's not actually specific to this disorder and so that was a concern. It's basically using the chronic care model and making an assumption that because this is a chronic disease that the chronic care model and the data that have arisen from that would also apply to this disorder. So I think there was -- there is evidence, but the evidence is inferential. It's not direct.

DR. WOODS: What I just described to you is actual evidence from studies that

demonstrate that follow-up visits lead to better adherence and poorer adherence when there aren't follow-up visits.

MEMBER MOYER: Right.

DR. WOODS: So that's the evidence for ADHD about the type of follow-up. And I know the discussion about inference was really whether we could believe that ADHD exists, someone brought that up on the phone.

CO-CHAIR SUSMAN: Let's not go down that path. I think by a vote how many want ADHD, right?

Okay, John.

DR. BURSTIN: Folks need to turn their mics off.

CO-CHAIR BROOKEY: So just to clarify, if this visit, especially for these older children, is in the context of a well-child visit and they code both a well child and an ADHD, it's going to count, is that correct?

DR. WOODS: Correct.

CO-CHAIR BROOKEY: Thank you.

MEMBER MILLER: I'm sorry, this is Marlene. Did you say that would count or would not count?

CO-CHAIR SUSMAN: It would count.

MEMBER MILLER: So if a child has five diagnoses, the order doesn't really matter? I was sort of stuck on those last four words of this thing that says "as the primary diagnosis."

DR. WOODS: It's specified as primary or secondary diagnosis.

CO-CHAIR SUSMAN: So any order, it will count.

Kevin?

MEMBER MOYER: So the actual specifications are that the follow-up visit also is primary? That isn't what's stated.

DR. WOODS: Primary or secondary.

MEMBER MOYER: For both. Okay.

MEMBER SLAVIN: The studies that you quoted suggest that follow-up leads to improved adherence with medications, but that doesn't necessarily show that there's improved outcomes

based on just the follow-up appointment and adherence to the medication.

I know there's studies that show that treatment plans do improve certain functional outcomes, but the ones that you're specifically citing, at least if I understand them correctly, just specify that because patients follow up they take their meds, but it doesn't necessarily mean -- it doesn't necessarily get us that final step that the outcome is improved.

DR. WOODS: As part of this activity, we did a systematic review which was published in -- what was it, Journal of General Internal Medicine or Annals of General Medicine, something like that. And I can get you the article that actually links. Actually, we were going -- we were supporting an activity of the NCQA who were trying to develop an outcome measure for ADHD because it was possible to demonstrate through the literature that medication and behavior therapy adherence did improve outcomes of the ADHD condition. But if treatment is stopped,

there's return to the condition specific symptoms impairment.

CO-CHAIR SUSMAN: David.

MEMBER EINZIG: So I was just curious as to why just one visit for follow-up for a year? How do you come with one?

DR. WOODS: At least one, that's the way it's stated, at least one. It can be more, but it doesn't need to be more for all children.

CO-CHAIR SUSMAN: It's a low bar.

DR. WOODS: Well, we understand that some of the situations that people were discussing in the other measure, the child may be managed. And you just have to check in and make sure.

CO-CHAIR SUSMAN: Carol and then John.

MEMBER STANLEY: Yes, can you explain a little bit about does the prescribing provider have to be the one to do the follow-up visit? Because with Medicaid population, there's frequent changes in PCP and in health plans sometimes, especially if foster care. So when

operationalizing this measure, was it taking into account, did it have to be the same provider that was prescribing that did follow-up?

DR. WOODS: We did not go with the prescribing provider because that was one of the key issues with ADHD measures. Previously, states could not really use the measure because federally qualified health centers don't provide the DEA numbers of the physicians.

MR. FINKELSTEIN: So these are very quick questions and neither is disqualifying, but I just am trying to understand. What if a patient is seen in my practice and then moves to Nebraska? So there's no evidence of any visit at all in that following year, but there's also no documentation in the chart that they've left. So I just wonder, if we just have to not worry about that, so that's number one.

Number two, can you clarify for me are these calendar years? So if I'm diagnosed in December, December 2015, is it that there's a visit in 2016, which is kind of month 2 through

14 and if I'm diagnosed in January of 2015, it's month 14 through 26?

DR. WOODS: Yes. It's based on the time between diagnosis and the next visit. So it's a year and a day from your diagnostic.

MR. FINKELSTEIN: Not calendar year, it's not calendar year, it's a year and a day.

DR. WOODS: Okay, what you described I thought --

MR. FINKELSTEIN: It's not calendar year, right.

DR. WOODS: Right, because the idea is it is recommended practice to see a child many more times in the first year and that's not the chronic care management part of things. It's the titration of medication. It's the determination of effective treatment, whereas beyond a year, that's where the chronic management takes up.

CO-CHAIR SUSMAN: Okay, Ricardo.

MEMBER QUINONEZ: So I just wanted to comment on the one-year issue. I actually think the opposite. I don't think it's a low bar. If

a clinician who is managing this condition has very good communication with the family, including phone calls, emails, etcetera, that's good care. And so I think a year is not bad.

CO-CHAIR SUSMAN: A year seems to me to be arbitrary.

MEMBER QUINONEZ: Arbitrary, so I'm saying I don't think it's a low bar. It's actually -- you could argue if you have good communication with the family and the child is actually doing well, then a year may be too much.

CO-CHAIR SUSMAN: Today, in most advanced systems, certainly there are other means of communication than a visit to the physician. I think you make a good point there.

David.

MEMBER KELLER: So one of the things that happens in the management of this chronic disease is that families decide to opt out, that there are families who decide that their children should not be on medication. They don't want their child on medication and for those families,

I've always found it challenging to bring them in for anything other than their well visit. And we'll get into my coding issue with the well visits because I think that is an issue later. But for kids who are not actively and they opt out of behavioral health treatment, behavioral therapy as well, and just say they'll manage it by themselves. I had a substantial number of those folks. I don't know how prevalent that is nationally. But I'm wondering if that came up in any of your discussions because with families, the decision to medicate is actually a pretty major one, that families make for a variety of reasons, some of which is evidence based and some of this is just based on what they hear and from a variety of people or their previous experience with medications and their family members and things like that.

DR. WOODS: There was some discussion of that issue. Our expert panel believed that it was a part -- like if someone doesn't take their diabetes medicines or they don't take their blood

pressure medicines or cholesterol medicines, that it is a part of this chronic care management to bring a patient in and have those difficult conversations. Difficult conversations happen.

And what are the concerns? Well, maybe there are some side effects. Maybe there are things that the physician doesn't know about. So we did discuss that and this -- our expert technical panel that included parents felt that this is part of good chronic care.

CO-CHAIR SUSMAN: So let's try to wrap this up.

Virginia?

MEMBER MOYER: Remind me whether this is a provider level or a health plan level measure.

DR. WOODS: So we were unable in the testing context, we were working with the Truven MarketScan database and they cannot give us provider-level information, not that they don't have it. It is with their contract, the way they get their data, they're not allowed to share that

information with us, but they could share with us whether it was a Medicaid or the types of insurance.

CO-CHAIR SUSMAN: So this is a population-based measure?

DR. WOODS: Yes.

CO-CHAIR SUSMAN: Thank you.

MEMBER MOYER: It's intended to be a health plan level measure because that is actually relevant to John's question about somebody moving to Nebraska. They're out of the health plan, so they're no longer in your denominator.

CO-CHAIR SUSMAN: If they have continuous enrollment for a specified period.

MEMBER MOYER: Right.

CO-CHAIR SUSMAN: Okay, are there any other new perspectives? Not seeing any, let us move on to consideration of evidence.

MS. CHAVEZ: Okay, voting for evidence for measure 2818, ADHD chronic care follow-up is now open. We are expecting 25 votes. I guess 24

votes. I see 22, 23. It would be 23 votes. There are two people out. Twenty-four.

CO-CHAIR SUSMAN: Are we set?

MS. CHAVEZ: Okay, so 2 voted high; 17 voted moderate; 3 voted low; 2 voted insufficient.

CO-CHAIR SUSMAN: Okay, gap. Any comments on gap? I think we've at least touched a little bit on this.

MEMBER MORROW-GORTON: I think from the conversation we had that their performance on the measure is fairly low, 50 percent. We don't know individual provider. It's probably really variable and there were some disparities in terms of minorities and the documentation of follow-up visits for them compared to the general population.

CO-CHAIR SUSMAN: Do we have actual plan level data on gap or is it all amalgamated?

MEMBER KELLER: We have Medicaid.

CO-CHAIR SUSMAN: Okay, but not one plan versus another plan.

MEMBER KELLER: Although there was a disparity between Medicaid and commercially-insured patients.

CO-CHAIR SUSMAN: Any other comments about gap? Let's vote.

MS. CHAVEZ: Okay, voting on gap. Eighteen, 24, 25. So 5 voted high; 19 voted moderate; zero for low; and 1 insufficient.

CO-CHAIR SUSMAN: Moving on to reliability. Was there any empiric reliability testing of this measure? And if so, what were the results?

DR. WOODS: What we did was compare a sample with the remainder and found strong reliability.

DR. NISHIMI: So in this case, the developer appears to be relying on validity testing at the data element level. So however the committee judges that approach would then carry into the reliability field. So if you vote moderate validity at the data element level, then it would be moderate validity at the -- we don't

require separate reliability testing if they've conducted validity testing that you judge as adequate at the data element level.

DR. WOODS: And I can read the results if that would be helpful.

DR. NISHIMI: I think the committee has it in front of them.

MEMBER KELLER: So this is where though I had some concern only because of what I've been told by various coders over the generations which is the well visit, how to code a well visit. And I've been told at different times to code many diagnoses and to code only a V code, depending on what they believe payers are paying at that given time.

PART 3 Section B 1:35:54

There was a time where a number of payers would only pay -- basically, if you did a code 25 and merged a prevention code and an illness code at the same time, the payer would pick the one that cost the least and pay that one and deny the other claim.

And particularly in Medicaid which at least in two of the states I've worked in pays significantly better for preventive care than it does for acute care, we were encouraged to not code for diagnoses like asthma and ADHD during visits that involved well child care.

I don't know how prevalent, again, that practice is, but I'm concerned that a number of visits where ADHD -- it's pretty typical to address ADHD during a well visit when you're seeing a child who has ADHD. So whether you code for it or not, I'm concerned that we would be losing a lot of that information if we relied solely on coding.

CO-CHAIR SUSMAN: Kerri.

MEMBER FEI: In reviewing here, I'm noticing that the exclusions use the medical and patient reasons for exclusion, those buckets. Are those specified -- are all of those possible reasons specified out and available via coding? Because usually that's a provider level method of exclusion for measures and aren't able to be

collected administratively.

DR. WOODS: You mean like -- the exclusions are codes for autism, substance abuse, anorexia, mood disorders, and anxiety.

MEMBER FEI: So only those and they can't do anything else. Okay. Usually those aren't all coded out. Secondly, since it is a health plan measure, I see continuously enrolled during the measurement year, not continuously enrolled during the measurement year, excluded. There's no allowable gap.

CO-CHAIR SUSMAN: So one of the questions I think remains around this issue of can the validity testing stand in for reliability testing here, and I guess I'm not clear that you've done anything more than face validity testing or maybe I'm not getting it, seeing it.

DR. WOODS: Reliability testing.

CO-CHAIR SUSMAN: Or maybe one of the folks who really took a better look at this could help out.

MEMBER KELLER: We had that same

concern that the validity testing seemed to be face validity testing.

DR. WOODS: "For critical data element testing, each measure component, numerator and denominator exclusions were tested through implementation. Results were reviewed and reliability was assessed based on comparison with the total ADHD population and Medicaid, CHIP, and commercial insurance respectively. Results of the analysis of the measure led to substantial changes in the initial proposed specifications. The components were iteratively tested until results indicated the measure specifications were capturing the correct population. For performance measure score, the measure was implemented in a Truven MarketScan database and performance was compared to performance of the initial core ADHD follow-up measure."

"Administrative claims" -- so there's also a different measure that we compared it to. "In the critical data element testing of the Medicaid population, 22.52 percent of the

denominator population had a valid specific psychiatric E&M visit with an ADHD diagnosis code in the measurement year. Similarly, 13.43 percent of the denominator population had a valid other psychiatric E&M visit with ADHD" -- I can continue. Shall I? I'm trying to give you a sense that you have --

CO-CHAIR SUSMAN: I'm not sure. Virginia, do you want to comment?

MEMBER MOYER: Yes. I actually read all of that several times and I still don't understand what was done for validity testing to know whether what you say you were measuring is what you are actually measuring. Testing it against the rest of the sample tells you that you got a good random sample.

DR. WOODS: Okay, so initially we were asked about reliability testing and that's what I was just reading for you is what we did for reliability testing. I can read you what we did for validity testing.

MEMBER MOYER: I've read it. I don't

need you to read it to me, I need to understand it better. That's where I'm struggling.

DR. WOODS: Help me understand what was problematic.

CO-CHAIR SUSMAN: Use your microphone, if you would. Thanks.

MEMBER KELLER: So what we don't -- what I'm not understanding is what was the gold standard and --

DR. WOODS: Complementary analyses.

MEMBER KELLER: Say that again?

DR. WOODS: Complementary analyses.

MEMBER KELLER: What's that?

CO-CHAIR SUSMAN: What does that mean?

DR. WOODS: We implemented the existing CHIPRA initial core measure of ADHD and compared it to our proposed version.

MEMBER KELLER: So you took the initial -- the current standard, the current CHIPRA standard and compared the results against this?

DR. WOODS: Yes. We also examined the

likelihood that children met the follow-up requirement with an E&M visit versus a non-psychiatric visit.

All individuals in the denominator -- so we also assessed the denominator eligibility, inclusion, and exclusion.

MEMBER KELLER: So essentially, you used the CHIPRA standard to create a gold standard and then you compared and said this works the same as the CHIPRA standard does.

DR. WOODS: Better, works better.

MEMBER KELLER: In the Truven database. Okay.

CO-CHAIR SUSMAN: Was there any then reporting of the statistical analysis between the level of agreement between those two databases?

DR. WOODS: "Results of testing of the new specification of the enhanced ADHD follow-up measure to assess chronic care follow-up were strong. High-level results include that 63 percent of Medicaid enrollees and 49 percent of commercial enrollees who had sufficient coverage

and were diagnosed with ADHD in 2010 had any valid E&M visit for ADHD diagnosis code in the measurement year."

CO-CHAIR SUSMAN: Okay.

MEMBER MOYER: That was results, but did you have comparative results?

DR. WOODS: I'm looking for them. It appears we may not have reported on that, but we have it, so I can get it for you. Just to tell you what we were thinking, we were concerned -- we wanted to be sure that we were not losing a lot of children, but our look-back periods and you know that year, and we would consider it non-valid to have a lot of children falling out and so we found that they were not falling out and we should have reported.

CO-CHAIR SUSMAN: I guess I'm still personally unclear what you did with regard to validity testing beyond face validity and reliability testing and the description that it was done doesn't feel sufficiently detailed for this process. But let me go on and get Carol.

MEMBER STANLEY: I think maybe it would be helpful to hear how you know that conducting this measure using administrative data will give you the same results as using medical record abstraction.

DR. WOODS: We did not -- and generally, it's not done to do reliability testing of administrative claims versus chart review. And our goal was -- our goal was to have this measure used by Medicaid which was part of the program that we were involved with. And Medicaid, generally, won't use chart review measures. So we did the best that we could with an electronic administrative claims measure. Administrative claims measures have their challenges. And this will equally have those challenges, but they won't be used.

CO-CHAIR SUSMAN: Okay, clearly, there's limitations in any data, some stronger than others. We might go on to Jon.

MR. FINKELSTEIN: So if there's an opportunity to come back to this measure, I would

ask you to go back and really look carefully at the specifications for the numerator and the denominator because I don't think they make clear what you were saying before. There's language about calendar year. There's language about a measurement year and a prior year. These are these 12-month periods you're talking about and it needs to be much more clearly specified so there's no confusion.

DR. WOODS: Right. So it ends up being a bit of a lag, right? So they have to be continuously enrolled and then it has to be a year.

MR. FINKELSTEIN: So I understand --

DR. WOODS: -- from the diagnosis.

MR. FINKELSTEIN: I understand what you're aiming for. I'm saying it's not clear as written. I've now read it several times.

CO-CHAIR SUSMAN: So unless there's anything new, why don't we go ahead and vote on reliability and then validity. I think there's been a lot of useful comments. One through four

on reliability.

MS. CHAVEZ: Okay, voting for correct is now open. Eleven, 21, 23, 25. Zero voted high; 5 voted moderate; 13 voted low; 7 voted insufficient. And this does not pass reliability.

CO-CHAIR SUSMAN: So this doesn't pass. I guess we do not need to go on, but I guess my feedback would be to really try to work with the NQF around the validity/reliability testing and making it clear. You may well have done everything that's necessary, but it was difficult to tease that out and albeit we're not as facile with the data as you are. Any other feedback?

CO-CHAIR BROOKEY: I just have one comment about moving towards more virtual medicine, especially in this field. It may be appropriate to have video visits. It may be appropriate to have telephone visits, especially for the stable children. And developers I think should consider the fact that those are really

legitimate ways to provide care. And I think they should be included in some of these measures.

CO-CHAIR SUSMAN: That goes along with what Ricardo was saying earlier.

Okay, you've had enough of me, so we're going to make a switch and John has the unenviable task of going through a host of related measures. And hopefully, he'll find some magical way to get us back on time.

DR. WOODS: I had one question about the accurate diagnosis measure. Is there any follow-up that I can do or we can do regarding that?

DR. NISHIMI: We'll follow up with you after the meeting.

DR. WOODS: Okay.

CO-CHAIR SUSMAN: Thank you very much. Should we take a five-minute stretch before we go in --

CO-CHAIR BROOKEY: Why don't we take a five-minute stretch and then we'll go into the

ten measures.

DR. NISHIMI: Literally five. This will be a little bit of a slog.

(Whereupon, the above-entitled matter went off the record at 1:49 p.m. and resumed at 1:55 p.m.)

CO-CHAIR BROOKEY: It turns out to be ten measures. We're on 2770 family experience as for coordination of care, FECC measure set. And I understood that from the discussion with the member group that it was decided that we would vote on all of these ten individually. Is that correct? Okay.

And so the good news is that for some of these, the evidence is based on the same studies. And so we may be able to sort of lump some of these together although we will still vote on them individually. So we're going to have to be -- we're going to ask both the developers and the members who are on point for this, I think it's Tim and Marlene, to try to clarify if a particular measure, if the

discussion doesn't need to be as long because we just talked about something where the evidence may have been the same for this particular question on the survey.

So having said that, let's turn to the developers to give us an overview of these measures.

DR. LION: My name is Casey Lion. I'm with Seattle Children's Research Institute and I am going to be introducing to the family experiences with coordination of care, FECC measure set which is 2770 from the Center of Excellence on Quality of Care for Children with Complex Needs. And there are measure development processes similar to what you heard about this morning for the mental health measures. I will review it very briefly now.

So our Care Coordination Working Group began by developing a conceptual framework. We then used the conceptual framework to guide six separate literature reviews in domains that seemed to be related to care coordination related

processes that might have impacts on short and long term health outcomes. We then used the evidence from these reviews to develop each of the proposed measures. Then presented the measures to a multi-stakeholder Delphi panel which included caregivers of children with medical complexity.

Measures that met these validity criteria were then operationalized and underwent cognitive interviews with families in both English and Spanish.

We then field tested the measures in a sample of over 1200 caregivers of children with medical complexities in two state Medicaid programs. Of the 21 original FECC measures that we field tested, we've submitted for endorsement the 10 measures with the strongest evidence from the literature, demonstrated performance caps, and the most compelling testing results for reliability and validity.

The FECC measures that we have submitted includes ten separate measures which

can be used either independently or in any combination to assess to quality of care coordination processes provided to children ages 0 to 17 with medical complexity. These are all survey based caregiver reported measures as caregivers are presently the most reliable source for this information which addressed the family perspective and are not reliably documented in the medical record.

Examples include whether the child's care coordinator assisted with completing specialty refers, whether the child has a shared care plan. Measures do use billing data to identify the overall denominator population of children with medical complexity using the pediatric medical complexity algorithm or PMCA. The PMCA has also been separately tested and demonstrated excellent sensitivity and specificity for identifying children with medical complexity in both Medicaid claims and hospital discharge data.

The ten measures are all supported by

some empiric evidence with the exception of FECC-14 which is supported by strong expert consensus. Additional evidence for FECC-14 and 17 were recently circulated to the committee following the work group call. The majority of measures demonstrated good reliability although two were limited by small sample size and all of the measures demonstrated excellent face facility and convergent validity with at least one other care experience outcome measure.

CO-CHAIR BROOKEY: That was a great summary. Thank you. Very concise. Before we move on, are there any recusals from voting? Okay.

Okay, I understand that for the evidence discussion that the first few measures, 1, 3, 5, 7, 8, or 9 rely heavily on one RCT and so I wonder if maybe we should just begin our discussion about those measures first and we can vote on those because I think it might be more helpful to kind of break these apart just a little bit.

I don't know if that makes sense, Tim or Marlene, but just a suggestion. And is it going to be you or Marlene that leads off the discussion?

MEMBER MILLER: We really haven't talked about it, so I'm fine or Jim can do it, too.

Do you want me to open with my comments then?

CO-CHAIR BROOKEY: Tim has volunteered to go first, okay?

MEMBER MILLER: Okay.

DR. NISHIMI: You need to use the mic.

MEMBER BOST: For FECC-1, there was one randomized control study, one cohort study, and five case series, case control or historically controlled studies that demonstrated that outcomes improve when caregivers of children with medical complex report that their child has a designated care coordinator. For FECC-1, besides what you already pointed out, about one RCT, there was also the committee was concerned

about a lack of clarity about who the care coordinator is. Many insurers assign care coordinators for high utilization patients. That coordinator would have different responsibilities than a clinically assigned care coordinator.

MEMBER MILLER: Yes, this is Marlene. I guess I would add in the one RCT was really limited is how I best say it because it only involved 100 children and it only followed them for 6 months which seems insufficient to really comment on improvements in chronic conditions when you're following them for 6 months.

I think more importantly other than the fact that there is that one very, very short RCT in there was that the RCT was not about involving a care coordinator. It was the small type factorial intervention of which it's impossible to say that the breakdowns that have happened into these six or seven questions that stem from them are the logical pieces at all. Particularly, you know, having a care coordinator as Jim just said that may have other names in

other settings of care, so trying to extrapolate from this very, very short study involving 100 children for 6 months and then using that one name of a care coordinator, I didn't see the evidence that was there to warrant the measure.

CO-CHAIR BROOKEY: Does the developer want to respond to that concern?

DR. LION: Sure. So to begin with in the survey, we way we operationalized the survey, we actually set up the questions to try to figure out exactly who might be coordinating care for these children. So we allowed for the fact that it might be somebody, it might be the main provider. It might be someone within the provider's office. It might also be someone outside of the provider's office, for instance, a care coordinator assigned by an insurance plan for high utilizers, for example.

So the questions were actually framed as did anyone in the main provider's office help you to manage your child's care or treatment from different doctors or care providers? And then

did anyone else outside of the main provider's office help you to manage your child's care or treatment from other doctors or care providers?

And we developed that language through our cognitive interview process with families because care coordinator was not actually universally understood by families as meaning precisely what we thought it would mean.

And then with regard to the concern about the studies that we used as evidence, mostly relying on multi-factorial interventions, that is essentially true across the board of just about all of the evidence that we have. And at the end of the day, those are the studies that had been conducted. They've all been sort of bundled interventions.

We also know from other research that bundled interventions are more likely to be successful than single component interventions. So it may not be possible to actually or even advisable to try to extricate individual components of these bundled interventions. It's

something that we recognize and we own, but we did the best that we could with the evidence that's available.

CO-CHAIR BROOKEY: Okay, are there questions or comments from the members? Go ahead.

MEMBER KELLER: So yes, I would echo what you just said and just wanted to point out that I think one of the strengths of this measure is that unlike the studies where someone was designated by an external force to be the care coordinator, what this measure is looking at is the parent's perception of whether or not there is a care coordinator. And I would submit that that's actually much more important than where the care coordinator is located. If the parent perceives that they have one, I would bet that that's an important measure. So I like that. It really builds on the evidence that's out there.

CO-CHAIR BROOKEY: Other comments? Go ahead, Amy.

MEMBER HOUTROW: So I really

appreciate how in the survey that you guys were able to distinguish what type of person was providing the care, but the use of one RTC that uses an inside force for six months, I think when we talk about the evidence that's where we're talking about being concerned.

And the expectation of an in-practice case manager in all of these different studies is really kind of different than if families were identifying that their care coordinator was somebody from their insurance company, for example. And maybe that's not such a big deal with whether they have it or not, but the activities that fall below. So maybe my comment goes better with the rest of the different items that we're going to be talking about that follows. But I think you were very wise to work on perception and identification in the survey.

CO-CHAIR BROOKEY: Other questions? So just to clarify, if I could ask, are you saying that the evidence is stronger for the entire set of measures as opposed to any

individual component of the set?

DR. MANGIONE-SMITH: I would say that that's a fair interpretation. When we saw how -- the intervention in any evidence we found were bundled, it was by choice that we felt measures should look at those individual components because there was no way to tell which of them drove the better outcomes that were arrived at in those studies. So that's exactly right. We really feel the evidence for several of these measures come from those bundled interventions.

CO-CHAIR BROOKEY: I know there's a question down there, but I just wanted to bring up that the question for the group then, the decision previous to this meeting, was to go ahead and vote on these as individual components as opposed to a bundle. So I want to be clear if that's still the direction here. Are there any comments about that before we move on?

CO-CHAIR SUSMAN: I mean it seems to me from what you're saying and reading through this this really makes more sense as a whole, as

a bundle rather than breaking it down into all these little component parts. I didn't study it as long as some of you did, but that's just my general sense.

MEMBER MILLER: This is Marlene. I guess I would say that I would take it even to a different level. Instead of ten questions that parse out tiny aspects of the bundle, was there -- is there a possibility we could instead have two or three questions that get at more of a larger construct that don't confine us to, for example, exact wording, did your care coordinator ask about a concern. I know it's very hard to think at that minutiae level when we know that the whole intervention was much more than that. So when you say bundled or individual, I guess I'm saying is that I think -- I kept wondering myself is could we have not done ten questions and maybe done two or three at a bit higher level and really gotten something a little easier to wrap our heads around.

CO-CHAIR BROOKEY: Go head.

MEMBER DORSEY: My question for the developers is related, which is just trying to grapple with what's the rationale, given the conversation that we've had to break these out into their component parts. And since it doesn't seem to be directly supported by the evidence that you all produce in the application, is this more an issue of how you intended to be used and that you're trying to make discrete information about specific components so that individual providers can evaluate where their care coordination may be breaking down? I mean it's not explicitly stated, but I'm trying to figure out sort of what's the balance or rationale here.

DR. MANGIONE-SMITH: That's exactly why we didn't want to go from our global constructs. Our hope is that people could track these measures over time and understand where are they falling down in terms of their care coordination services, so what are families telling them they're not meeting in terms of helping us get community services, helping us get

sub-specialty appointments when we need them. Several of the more detailed things that have been broken out here.

It's also partly why we wanted these to stand as individual measures is these are not measuring a single domain. There are several different aspects of care coordination being captured by the different measures and depending on what it is you're trying to accomplish with your care coordination project or program, not all of these are going to apply. And that's why we wanted to suggest to people you don't have to ask all of them to understand whether you're giving high quality care coordination or not. You can ask specific measures that make sense for the program that you're trying to implement and improve on care coordination with. So in that way, it's very different than say, for instance, the CAHPS measures where you really are supposed to ask the whole survey, right, in order to get at whether experience is good. That's a much more global sort of thing, but this was really

trying to help people understand where they're doing well in care coordination from the family perspective and where they're not doing so well and they may want to put some of their improvement efforts.

CO-CHAIR BROOKEY: So what I would say is that we put on the table the question of bundling two to ten measures and vote on them individually with the understanding that in the future these could be brought back together as a bundle. So I think that that would be cleaner today if there is no disagreement. We'll go ahead and talk about each measure with the understanding that if some of them do not pass, it doesn't mean that in the future they could be reconsidered or even brought back together as a bundle. I'm not quite sure any other way to do it, especially since the subgroup had decided that it would be better to vote on them individually. Any objections to that?

MEMBER MILLER: This is Marlene, but I guess I'd bring up the question, you know, it's

one thing to use it to improve your practice, but again when you get to someone being accountable and if we are approving them as individuals, that leaves the door open for some entity to pick or choose one or two of these things where we know the evidence doesn't make sense at this granular level of these ten questions, and then put resources and drive things, holding people accountable to things where there's not direct evidence at that minutiae level.

DR. MANGIONE-SMITH: So the level of analysis for these measures is considered the health plan or health system level. They can be used for quality improvement intervention evaluation, but the intent is for them to be used as accountability measures for the complex, medically-complex child population to look at quality of care coordination. So I want to be clear about that. The intention here was to have these be measures of accountability and that the health plan and/or health system would be responsible for improvement on these measures.

There were on our Delphi panel, Medicaid health plan representatives. There were parent representatives. There were providers who care for these children who are content experts. These were individual aspects of care coordination that they endorsed as having very high face validity in terms of indicating you did better on these, you were getting higher quality care.

CO-CHAIR SUSMAN: So I was trying to look through each of them and see which wouldn't necessarily be applicable for a child with complex disease. I mean could we say that care coordinator helped to obtain community services wouldn't be really germane or appropriate written visit summary content wouldn't be germane. It seemed to me the elements would be applicable to almost every child with complex healthcare needs and therefore why wouldn't we want to measure all those in a single bundle. So I guess I'm going back to that set of concerns that by parsing these out, I mean let's say has care coordinator

is the one that some health plan takes for its accountability measure. That could mean anything or not much at all.

CO-CHAIR BROOKEY: The real question I have is whether these are actually tested as a bundle versus individual measures.

DR. MANGIONE-SMITH: So the survey is actually not a very long survey. It takes about 20 minutes to complete and you actually can get 20 measures out of it. As we were saying we've only put ten forward for endorsement for the reasons that Casey stated up front. So my assumption and I can tell you from the people who have already asked us for the survey, and are using it for different purposes, currently are all doing the complete survey. They have not been picking and choosing different measures to do. And it was field tested as a whole.

CO-CHAIR BROOKEY: Amy.

MEMBER HOUTROW: So I think it might help us, maybe to look at the concept map that was provided in this packet because it helps you

look at which of these different measures are intended to address different aspects of care. And I also think that when you think about the answers to the questions that parents provided as a survey, let's say they said that their care coordinator was knowledgeable, supportive, and advocated for their child's heath. My guess is that they also said yes, that their care coordinator asked about concerns in health changes. And I bet that those two are highly correlated. But perhaps there are other measures here that are left correlated to the other one. So for example, saying yes to question 8, knowledgeable, supportive, and advocated for child's needs, might be highly correlated with concerns in health changes, but not so highly correlated with getting an appropriate after visit summary because those are content and conceptually different. They're not as close to each other on the map.

And so for me, the map helps us think about how these things are interrelated to each

other and I think that you're making a good point, that these things are important for kids who have complex needs. I don't think anyone is really doubting that those are important, but I think for me, the way I am looking at it is it is about how conceptually close are these different measures and could we potentially lump some of them together as very similar? I think the reason that they're kept apart is because there are unique aspects to the measure that the developers and the Delphi panel thought were important.

There were other measures that didn't make it to the table for us, right? So that's why we see not number four, for example.

CO-CHAIR BROOKEY: So I'm going to go Tim and then I'm going to start going around the room this way, so Tim, go ahead.

MEMBER BOST: So the reason, one of the reasons we've split them was that you will see later on some items basically fail criteria. FECC-15 has no validity assessment. So if based

on that, no validity for one of the questions you would have to give validity low. You're going to throw them all out because of one. So it was later on in the process that we decided we needed to split these if you wanted to fairly assess each item because of differences associated with those.

But I would also say listening to the two folks, if this is a health plan level, and a health plan can pick and choose the subset they want, you can't get accurate benchmarking because you're basically picking and choosing the ones that you're best at and especially if it then also is used somewhere down the line to evaluate health plan performance. So, you know, I get both sides of this coin.

MEMBER FEI: And actually Jim just mentioned what I was going to bring up from the health plan side and to add on to that, he's absolutely right in that respect. If it were to be something that was eventually publicly reported for health plans and consumers were able

to make choices, but you're allowed to pick and choose the measures, there's no benchmark -- there's actually then no real benchmark or you're being compared to a benchmark that contains everything, when you plan only to choose maybe the things you scored really well on. So if it's going to be endorsed for accountability and someday gets out there for consumer choice or provider incentive or something else, it needs to be this is how you use it as a packaged deal and everyone uses it the same way.

MEMBER MILLER: This is Marlene. I would go along with that. If there are some questions, if someone said FECC-15 where there's no validity and maybe that should not be on this panel of questions, comparable to FECC-14 which is really an extrapolation of a very tangential extrapolation from the evidence and maybe those should be removed and we should get to a smaller set of two or three or maybe four questions that we all agree should be asked and always asked as a bundle.

DR. MANGIONE-SMITH: So just to clarify, there is validity data for FECC-15. There is not reliability data.

CO-CHAIR BROOKEY: Ricardo.

MEMBER QUINONEZ: A lot of the things I was going to say had been said, but just to again iterate part of why I think in the phone calls we thought it was important to separate these is there's probably a good reason why NQF puts evidence bases as the number one thing you have to pass first to go on to consider measures. And although a lot of these indicators rely on the same evidence, there are some where the evidence is a lot weaker and so I think that's one of the reasons to vote for these individually because the evidence bases for some of these is very different and much stronger for some and not as much for others.

CO-CHAIR BROOKEY: So can I ask you to turn your name tags this way so we can read them? I still can't read your name though. What is your first name again?

MEMBER MORROW-GORTON: Jill.

CO-CHAIR BROOKEY: Okay, we can't read them because of your microphone.

MEMBER MORROW-GORTON: Oh, because it's upside down.

CO-CHAIR BROOKEY: There you go. Thank you, Jill.

MEMBER MORROW-GORTON: There you go. I'd like to come back to what Amy was talking about in terms of potential correlations between questions. You've got ten questions. Are there questions that track every time? You know if you get an excellent on one, you get an excellent on three, five, and nine, and it's not very helpful to ask all of those questions because it doesn't differentiate. Or is there one or two or some subset that are reflective of people that follow the process so that you didn't need to have all ten of them?

DR. MANGIONE-SMITH: Yes, so our great hope going into this was that we would do a factor analysis and we would find like, you know,

a beautiful set of five things we could just say this is it. They all measure something different and it's going to get us exactly what we need. We didn't find that. So we just -- we tried and I think to Amy's point, these are really getting at some very separate constructs around care coordination and even though it seems like some of them should really run together, we just didn't find that. They were not correlated with each other and we were hopeful that they would be.

CO-CHAIR BROOKEY: Jon.

MR. FINKELSTEIN: I agree with keeping them separate because we're in a very difficult zone with evidence, right? There's a huge consensus in the field that kids with chronic conditions, especially complex ones should have a medical home. Care coordination should be part of that. If you try to parse it too finely on which aspects of most evidence, you'll never get there and we'll never have any quality measures on anything to do with this. So I'm really

worried about that.

I think if we leave them separate, we may come again, through this pathway of insufficient evidence on that micro thing, but meeting all of the criteria for the exception and then seeing some of these ten measures come to the top as kind of more overarching important things that could -- that plans could be held accountable for now with other measures not being endorsed for accountability in that way, but being part of an instrument that could help health systems improve. And I think that would be a fine outcome of this process.

CO-CHAIR BROOKEY: So going back to my earlier recommendation, voting separately for all ten measures, which does not preclude in the future requiring the survey to be whole for the health plans to have consistent measures across all health plans. So not to confuse everyone. That means that we'll vote up and down for all ten but there may be further discussion about bundles and I think that might be the best way to

get through this. Any disagreement with that? Okay.

So any other discussion about the evidence for -- and I guess one question that we have to entertain when we look at the first measure is whether we consider the review to be sufficient to be able to rate it a high or not. So you have to consider that when we vote. Otherwise, if you don't consider the review to be sufficient, it may limit our choices of voting.

Are there any other questions or comments about this particular measure and the evidence to support it?

DR. NISHIMI: Just to be clear, the developer conducted its own review, but they did report on the quality, consistency and quantity of the evidence. So that's why you can march down that path. If you don't consider that sufficient, then it's not eligible for a high.

CO-CHAIR BROOKEY: Questions, comments, confusion? Go ahead.

MEMBER FATTORI: Are there any

measures that are connected? For example, if we don't move past with the care coordinator question, can we move forward with the other ones that have to do with the care coordinator?

CO-CHAIR BROOKEY: Yes. And we're going to do them independently, even though many of them are sort of linked. For this purpose, we're going to go through them one at a time.

MEMBER MILLER: This is Marlene. I was confused about the statement you were just making about -- are you trying to suggest we should be ranking based on the thoroughness of literature review or based on what the evidence is showing? I was confused about your comments about --

DR. NISHIMI: It's what it's eligible for. It's eligible for high, moderate, low, insufficient. I'm not making a comment on what it should be. It's what you then consider it to be eligible for, the quality of it.

MEMBER MILLER: But our lens is still on what does the evidence say about the measure

being asked?

DR. NISHIMI: Yes.

CO-CHAIR BROOKEY: Correct. Jeff.

CO-CHAIR SUSMAN: So I have a question, maybe best to the developers. We have these multiple measures. Did we do a separate literature review about each specific element down to the level of, for example, care coordinators, this was specialist service referrals and all the -- or was there just this one larger -- not larger, but one RCT?

DR. LION: So we conducted six separate literature reviews that were informed by domains that were identified based on the conceptual model that we've included in the packet, so the domains, the literature reviews were organized around things such as shared care plans, goal setting, information exchange, care coordination.

And in all of those separate literature reviews, most of which were conducted by separate people, we were looking for evidence-

based links between process measures that were related to that particular item and short and long term outcomes. In most cases we allowed, we preferred pediatric studies, but because of the dearth of pediatric studies, we also included some adult studies, particularly adults with chronic disease.

And so on the basis of the literature reviews, we developed the draft indicators that went before the Delphi panel. Did that answer your question?

DR. NISHIMI: And just to let you know, the developer did provide the information measure by measure, and so then they just aggregated their reviews and supply the evidence for each individual measure. So they did do that.

DR. LION: Yes, so started with -- sorry.

CO-CHAIR BROOKEY: Marlene.

MEMBER MILLER: I was just going to say but even if you did it though with six

different reviews, with the vast majority of these, the evidence still comes back to same one RTC.

CO-CHAIR BROOKEY: That's correct.

MEMBER MILLER: So you're saying that you did six different literature reviews, but they all kept pointing back to one RCT?

DR. LION: That is generally true. So for the first set of measures, the one through eight, those all came from the same literature review. Those all related to care coordination within the medical home and care coordinator functions specifically. And so those did come from the same literature review which we then disaggregated in order to present the evidence measure by measure.

CO-CHAIR BROOKEY: So is it fair to say that for those measures, we probably will be looking at the same outcome in terms of whether we support them or not? Okay.

So Amy, you have a question?

MEMBER HOUTROW: I have a procedural

question about the issue between high and moderate for number one which is how is the -- care coordinator. When we're looking at the evidence, the evidence is about a bundled set of activities that a care coordinator does. But if you have hired a care coordinator, they exist. You must then assume they are doing things. I mean I'm having a little trouble with the high to moderate based on just the presence versus the activities that exist that they do.

DR. NISHIMI: That's a decision you need to make. Yes.

CO-CHAIR BROOKEY: If you have issues with that, then you can't rank it higher than moderate, so that's an individual decision you'll have to make. Are there other questions or comments before we -- are you ready for a vote? Okay. Let's move forward.

MS. CHAVEZ: Okay, we're ready to vote on measure 2770-1, family experiences with coordination of care, FECC-1. We are expecting 24 votes on this. Ready, go. For the folks on

the phone, it's 1, high; 2, moderate; 3, low; 4, insufficient. Nineteen, 21, 22, 24. Thank you.

Okay, so 5 voted high, 15 voted moderate, 2 voted low, and 2 insufficient.

CO-CHAIR BROOKEY: Great. That's a great first start. And we have nine more to go. We're going to go through evidence on each one. So we're going to go to the next one which is FECC-3, I believe, and I think either Jim or Kerri are going to lead on this discussion.

MEMBER BOST: Basically, the evidence is identical.

CO-CHAIR BROOKEY: Any comments about this measure?

MEMBER MILLER: Could you use the microphone?

MEMBER FEI: I thought I was. I'm sorry. It builds on the first question.

CO-CHAIR BROOKEY: So are we comfortable voting on it based on the evidence being the same as for the first? Okay. So let's move forward.

MS. CHAVEZ: Okay, voting on FECC-3 evidence. Eighteen, 22, 23. Okay, 2 voted high; 17 for moderate; 3 for low; 1 insufficient.

CO-CHAIR BROOKEY: The next evidence is FECC-5, care coordinator asked about concerns and health changes and again, Kerri and Karen are the leads.

MEMBER FEI: Again, it's the same, it's the same one, right? There's not much else to say. I do think it's -- I think from the patient's side it's important, but that's outside of the evidence.

MEMBER DORSEY: I agree. Nothing to add for the evidence decision.

CO-CHAIR BROOKEY: You guys are going way too easy on me here. Wow. Are you ready for a vote? Go ahead.

MEMBER KELLER: I'm actually looking at the nice QCC table on page 36 as I'm going through. I'm noticing that this measure only has one article referenced, whereas the earlier measures had multiple articles referenced. I'm

presuming that's because this was a relatively new concept, the article that's referenced is the most recent one.

DR. LION: The other differences that the various articles included a variety of detail, level of detail in what was actually included in their intervention and so some authors provided far more detail which allowed us to understand the individual components of the intervention in great detail.

MEMBER KELLER: Got you.

DR. LION: So in some cases, they just didn't give us enough information in the study to know whether a particular, whether this particular thing was a part of their intervention, so we didn't cite it.

MEMBER KELLER: I was going to say because I'm looking at -- conceptually, this seems to be something that's pretty much standard is everything I've ever read, certainly in every guideline I've ever read around care coordination. Interesting. Thanks.

CO-CHAIR BROOKEY: Good point. Jon.

MR. FINKELSTEIN: I just wondered if the developers want to comment on the denominator issues that are different here. So in this one, it's only if you had contact in the last three months, so someone could have a care coordinator who never calls and then they're not in the denominator of this and the two patients they call they ask about these things and you're 100 percent, but you're still doing a terrible job in your system overall. And I understand pros and cons because I can think through why you did this, but I think it's important for you just to tell us about that.

DR. LION: One of the indicators that did not make it to you all was actually about having been contacted in the past three months. I forget why we dropped it. But we have that as a separate indicator and it didn't make it through the various hurdles although I forget on which criteria it failed.

DR. MANGIONE-SMITH: But you could

measure it from the survey. We're just not putting it forward for all of you.

MEMBER EDIGER: How is that different than the one that we have?

DR. MANGIONE-SMITH: So it's the feed-in measure to this one. It would have been FECC-4, you know, they go from 3 to 5.

MEMBER EDIGER: Just whether or not you've been contacted?

DR. MANGIONE-SMITH: Exactly.

MEMBER EDIGER: And then our question is asking about --

DR. MANGIONE-SMITH: Now we're looking at the subpopulation of people who were contacted to see whether these things happen when they were contacted.

MEMBER EDIGER: Okay. My comment is I'm a mother of one of these kids and three months seems incredibly short to me. When things are going well, and we're not in the hospital, I am more than happy to not talk to our primary care physician for at least three months. When

we go through intense periods where we're being hospitalized all the time and surgery, surgery, surgery, but even with that, three month seems awfully short to me.

DR. MANGIONE-SMITH: So interesting that the main evidence we cite said at least once each month and we got the same feedback from our Delphi panel, Family Voices person. She said that would be so irritating. And so it got extended out to three months based on her commentary and that's what, in fact, the working group pointed out to us. Why does the evidence say once a month and you're allowing three months. I explained the same thing to them. But that's what we settled on in the Delphi panel as being a compromise.

MEMBER MILLER: This is Marlene. I guess this still gives me pause. I take to bring it back, but you know, the logic of separating out these questions because -- and I even have more concern knowing the missing questions that are in there that would make these other ones

make sense is to whether it is really logical to consider these individually.

CO-CHAIR BROOKEY: Well, again, I think we decided as a group we're going to move forward with the understanding that this isn't the final, final, that things can be renegotiated later, but I don't think we're going to be able to get through this as just conversation about ten measures all at one time. So I would suggest that we just move forward and then we can always provide input and Maureen provide input about the frequency and things like that. Those can be considered for the future, definitely.

Any other comments before we vote on FECC-5? Okay.

MS. CHAVEZ: Okay, now voting for evidence for FECC-5 is open on the phone. One, high; 2, moderate, 3, low; 4, insufficient.

Seventeen. I'm reading 30, 35. And to those on the phone, the slide is showing 16, all 16 moderate.

CO-CHAIR BROOKEY: It's busted, okay.

So are we not going to be able to do this electronically? Should we do it manually? Right, now that we've kind of messed this one up, don't we have to go manual?

MS. CHAVEZ: Should we just go ahead and do this manually until you guys work it out?

DR. MANGIONE-SMITH: For dash 5.

DR. NISHIMI: So this is for FECC-5 and high, moderate. So that's 18 moderate. Low, five. Insufficient, one.

Okay, so it's evidence for FECC-5, 0, high; 18, moderate, 5, low; 1, insufficient.

CO-CHAIR BROOKEY: Okay, we can move forward to FECC-7, care coordinator assisted with specialist service referrals. And the leads are Karen and Lauren And Lauren, I believe, is still on the line.

MEMBER DORSEY: Yes. So I think the go-to study that we've been talking about, there's not really anything new to say about the evidence supporting this piece. I would say that the developers added an explanation of a few

other pre-post design studies that talk about utilization which I think is the logical outcome associated with this particular component of the measure which is being connected with appropriate specialty referral. So they do add that to the sort of go-to study we've been talking about which I think is important.

PART 3 Section C 2:40:33

CO-CHAIR BROOKEY: Lauren, anything to add?

MEMBER AGORATUS: Yes. I just wanted to add we had the same discussion on the three-month interval which was clarified that it was changed because of the Delphi review panel, and also there was a comment on clarification of assistance with appointments also includes complex care scheduling, which also showed up in the comments under validity, and just a general comment that the body of evidence was weak.

CO-CHAIR BROOKEY: Any other comments or questions?

MEMBER MOYER: This one strikes me as

being -- this actually has two things in it. It has did they help you with getting appointments and were they successful at getting you appointments? So if your child with complex healthcare needs has autism, and they live in the city that I used to live in, there's no way to get an appointment in three months. You're lucky to get one in 12 months. And that's if you know somebody. So I'm wondering whether conflating somebody helping you get appointments and getting appointments means that a parent who is trying to answer this wouldn't know what the right answer was -- wouldn't be able to reflect their experience.

CO-CHAIR BROOKEY: Can you read the question?

DR. LION: So the question itself asks -- so first it goes through and describes what specialists are. And it says, during the last 12 months, did the main provider tell you that your child needed to see a specialist? If you said yes, you moved on to the next question which was,

did the person who helped with managing your child's care contact you to make sure your child got an appointment to see a specialist? And the possible answers were never, sometimes, usually, and always.

DR. MANGIONE-SMITH: This was operationalized both as a survey measure and a medical records based measure. I think the three-month time window was retained for the medical records based measure, but it was not retained in the question on the survey for parents.

CO-CHAIR BROOKEY: Is everyone clear on that? Questions? Okay, if not, let's move forward on FECC-7.

MS. CHAVEZ: Okay, voting for FECC-7 is now open. Evidence. 1, high; 2, moderate; 3, low; 4, insufficient. Twenty, 23, 24. Zero voted high; 14 voted moderate; 7 voted low; 3, insufficient.

CO-CHAIR BROOKEY: Okay, so we're moving on to FECC-8: care coordinator was

knowledgeable, supportive, and advocated for child's needs. The leads are Lauren and Amy.

MEMBER AGORATUS: Okay, this is Lauren. I had two sets of comments. The first under evidence was the same as under the previous measure, which is body of evidence was weak. And then there were several comments under validity specifications. It's unclear why no ICC and Spearman-Brown were reported for this measure. Is it also because of sample size?

Another comment, not comfortable with saying this clearly demonstrates reliability because of internal consistency alpha for the sub-items. And also these questions are ideal for test/re-test reliability, but this was not done and would have been a better assessment of reliability. And then the last comment whether this should be evaluated separately just on reliability grounds. That's all I have.

CO-CHAIR BROOKEY: Great. Amy?

MEMBER HOUTROW: All right, so as you guys are aware, it followed that same RCT, but

unspecific to that is direct advocacy for needed care which is a component of this question which is whether they're knowledgeable, supportive, and advocate for the child's needs. And so there is potentially a little more specificity from the RCT, related to that one of the three aspects of this question.

FECC-8 does hit on a number of different areas and I think that as we think about it kind of throughout, we need to be considering that it hits on a number of different concepts, and it also requires that they are thinking yes to each one of these to get a positive response. But basically the evidence, similar to the ones before, is that main RCT and some additional studies.

CO-CHAIR BROOKEY: If you look at 8, you might presume they would have answered yes to 8 as well as 7, depending on the way they interpret the question. Eight may be a better measure than 7, looking at the two side by side. That's my question. Yes, Jon? Use your mic.

MR. FINKELSTEIN: It's warming up. There it goes. So 8 is the one where I'm really concerned about the denominator because I like it, so that these kids should have a care coordinator who is knowledgeable, supportive, and advocates, seems like the denominator should be the kids. And you shouldn't be able to score 100 percent on this because you don't provide care coordination services, except to a subset. It's kind of like FECC-1 with a little more specificity than I like, if you had that denominator. So I personally don't see -- I think it has to be interpreted with a denominator of children. I wonder if the developers have a special reason for doing it this way.

DR. LION: We considered both approaches and we were concerned that essentially going through, because so many of the subsequent measures build on that first measure, we were concerned about essentially repeatedly penalizing a provider group, a health plan, etcetera, for the same initial fault of not providing a care

coordinator. And so we thought this would provide us with, sort of, additional detail and specificity of again exactly where the problem is arising in order to be able to better hold people accountable and facilitate improvement activities.

CO-CHAIR BROOKEY: Amy.

MEMBER HOUTROW: So to Jon's point, I think that's an important one. So what are the activities of the care coordinator to do? In their mapping, the developers have provided that this question relates to collecting information, synthesizing information, sharing plans, executing plans, and determining where failures occur, which is basically almost everything that they set out to be important in their concept mapping. And so I think, Jon, you're making a very strong and valid point, that this is kind of the essence of what we meant when we mean care coordination.

CO-CHAIR SUSMAN: I'm getting really uneasy, though, that we're mixing up all these

data points with discussion around evidence. I see that there's very little evidence in this. It all ties to one RCT where they've amalgamated all these elements together and yet, we're disaggregating this by element. So to be able to say that this one has high level of evidence or any of the others or even moderate level of evidence, I just feel very uncomfortable with it. But obviously, we do need to vote with each of these.

I just have a hard time parsing out each one and then saying well, this element is important. This element -- and it seems we've sort of taken the whole gemish and we have a fruit cocktail here.

CO-CHAIR BROOKEY: I just want to make sure I understand Jon's point. Are you suggesting that we mush 1 and 8 together or -- I'm just trying to understand exactly --

MR. FINKELSTEIN: You can't talk until it turns red, which is ironic. It should be green. So what I'm -- so as we get to the end

and I agree with Jeff, this business of the evidence for each one, and in response to Marlene's comments, I could see us as a group saying there are a few of these that people really should today in 2015 be held accountable for and others that aren't ready for that level of accountability.

And if you ask me, and I know this isn't the way the developers have framed it, I would today make people accountable for these kids having a care coordinator who is not knowledgeable, supportive, and advocates. And that, to me, would it be an umbrella metric I'd be comfortable with. But I understand we're parsing right now, but we may at the end get to that.

And just to the point that Casey made, I think you do penalize if someone isn't providing a service, you penalize them for not providing a service. You then don't give them credit for doing it well on the small number of times that they do it. You keep penalizing them.

CO-CHAIR BROOKEY: So to be clear again, 1 and 8 should be really 1 in a way because it's not adequate to say you have a care coordinator. They should be doing all these things. I think that's what you're saying. And I think no one would disagree with that. Any other comments about this particular measure? Carol?

MEMBER STANLEY: Do you mind reading the exact question?

DR. LION: Sure. So it's a series of questions. The first is: in the last 12 months, did the person who helped you with managing your child's care know the important information about your child's health and care needs? Would you say yes, definitely; yes, somewhat; no. And then there's also don't know and refused options.

In the last 12 months, did the person who helped you with managing your child's care seem informed and up to date about the care your child got from other providers? And again options were yes, definitely; yes, somewhat; or

no.

And then In the last 12 months, did the person who helped you with managing your child's care support your decisions about what is best for your child's health and treatment? Yes, definitely; yes, somewhat; or no.

And I believe this one also: In the last 12 months, did the person who helped you with managing your child's care help you to get appointments to visit other providers?

CO-CHAIR BROOKEY: Which was the previous question, in a way.

DR. LION: Sorry, and there's another. So for a few of these, we had -- for the A, B, and C subsections, we had a couple of component questions that also --- in the operationalization of these particular questions and going through the cognitive interviews, we couldn't just ask families did the care coordinator -- did the person who helped you advocate for the child's needs, we actually needed to sort of parse out what that would look like.

And so -- sorry, another component of it was, in the last 12 months did the person who helped you with managing your child's care help you to get special medical equipment your child needed like a special bed, wheelchair, or feeding tube supplies? So getting help to get appointments and getting the special equipment was part of advocating for the child's needs.

CO-CHAIR BROOKEY: And there's an N/A response in there?

DR. LION: Yes, for all of those.

CO-CHAIR BROOKEY: Okay.

CO-CHAIR SUSMAN: How did you end up operationalizing the scoring of all that?

DR. MANGIONE-SMITH: So it is. Like she was suggesting, we took a construct like advocates for the needs of your child. And when you go to cognitively interview, you find out people have no idea what you're talking about when you say that. So then we say well, we mean like -- so do you understand when I say if I ask you did that person help you to get appointments,

and help you to get services you needed, and equipment you need? Oh, yes, yes. That all makes sense, right?

So then in building the survey questions to try to get at that construct, we end up with these two different sub-questions, and if they answered -- it was top box scoring, so if they answered yes, definitely, okay, then they passed that subpart of that construct. Okay, advocates for child was help me get appointments and help me get equipment and other services that I needed right? So they would have gotten, they would have had to say yes, definitely to both of those to get 1.0 on that subpart of the measure. And then there's two other subparts, is knowledgeable. That was captured in one question or two?

DR. LION: That was also two.

DR. MANGIONE-SMITH: So that was knowledgeable was captured in two of the questions, 5a and 5b. Supports the caregiver was captured in one question, 5c. So you get partial

credit in the ones that have two subparts. You get 100 percent in the one that only has one subpart. And then we roll it all up.

If we can go to the specifications of exactly how it scored, but it's essentially -- there's a lot of room for partial credit in there, right? So you can for any question or any construct that had two subparts, you can get 50 percent, right? For the one that only has one question, it's 0 or 1. And then you roll up what your individual scores were on all those subparts to get your score on the measure.

CO-CHAIR BROOKEY: Which is also why it makes more sense to be a composite instead of having a stand-alone measure.

DR. MANGIONE-SMITH: It is a composite measure.

CO-CHAIR BROOKEY: It is a composite within one measure.

DR. MANGIONE-SMITH: Exactly.

CO-CHAIR BROOKEY: Kevin.

MEMBER SLAVIN: How does a don't know

get scored? What happens with those answers because if there's a question and then there's a composite underneath of two questions that leads to that, somebody could answer yes on the first question and then don't know on the next two. So how does that all get kind of rolled together?

DR. LION: So we only scored measures for which we had complete information, so if somebody skipped, refused, or said don't know, essentially we dropped them from that measure, because we did not feel that it was fair to hold the practice or providers accountable for something that a parent may actually legitimately not know whether someone was working behind the scenes to help make something happen.

DR. MANGIONE-SMITH: Just to be clear, right now we're still talking about the evidence.

CO-CHAIR BROOKEY: Right. So no one else has asked, so let me just ask in terms of literacy levels, in terms of health literacy, was that diligently addressed during the creation of the tool, understanding that 50 percent of our

population, 60 percent in L.A. are health illiterate?

DR. MANGIONE-SMITH: Right, and I think if you look at our mode results, you'll see that we've got the best response rate in our mixed mode which was mailed, followed by phone, and the people who were in the phone part of that mix mode tended to be people who were either low English proficiency or minority. So I think that tells us pretty clearly -- and also of lower education levels -- so that tells us pretty clearly that those are people who received the mailed survey and were not comfortable answering it and returning it. So I think even though we went for a sixth grade reading level, and we did cognitive interviewing and all of that, there were people who were not able to complete it as a written survey.

CO-CHAIR BROOKEY: Any other questions or comments? I know we're struggling with this going through one measure after the other. Again, I just don't see an alternative, so if we

can just continue to go through these. We're just asking about evidence right now. Are we ready to vote? Do you remember which measure we're on?

MS. CHAVEZ: Okay, voting for evidence for measure FECC-8 is now open. One for high; 2, moderate; 3, low; 4, insufficient. Thirteen, 21, 22, 23, 24. Zero voted high; 19 voted moderate; 3 for low; and 2 for insufficient.

DR. NISHIMI: So for 8 that -- I just want to note for the record that on 7 that was in the gray zone, so we'll continue to discuss it, but did need to note that for the record.

CO-CHAIR BROOKEY: All right, we're moving on FECC-9, appropriate written visit summary content. I believe this is the last one that was primarily based on the one RCT, but there are other papers as well. I believe Amy and Jim are going to be commenting on this particular measure.

MEMBER HOUTROW: Yes, so this actually, 9 is whether or not there was an after-

visit summary which included a problem list, a current medication list, drug allergies, specialist involved in care, planned follow-up, and what to do related to problems from the outpatient visit. And this one used evidence from not the RCT, but the AAP consensus statement and Palfrey study in 2004 evaluating a medical home model with an N of 117, which was about a written care plan which may or may not be the same thing as an after-visit summary. And so that came up in our discussion of what this actually was.

We also discussed how this relates to the expectations from Medicaid Meaningful Use, which this is more encompassing than an after-visit summary from Medicaid Meaningful Use. And in our topic of discussion as a group, we talked about how this was different than a care plan and how that -- there was a general lack of evidence related to this question because it's actually not the same thing as a care plan.

CO-CHAIR BROOKEY: Go ahead.

MEMBER BOST: The only thing I would add is that it's scored never, sometimes, or always on each of the six components. And as Amy said, most of the concern was around: are these the right six?

CO-CHAIR BROOKEY: Any other comments or questions? None?

MEMBER KELLER: You mentioned this, but I wasn't clear on the outcome. How do these things line up with meaningful use? How do these line up with the requirements for meaningful use after visit summary?

MEMBER HOUTROW: Meaningful use includes these things, but this includes more than what meaningful use says.

MEMBER KELLER: It's more than meaningful use. Thank you.

CO-CHAIR BROOKEY: Questions, comments? Ready to vote? Okay.

MS. CHAVEZ: Okay, voting on evidence for FECC-9. One, high; 2, moderate; 3, low; 4, insufficient. Open. Sixteen, 22, 24. Zero

voted high; 11, moderate; 11, low; 2, insufficient. This puts this in the gray zone.

CO-CHAIR BROOKEY: So we keep going, we keep going forward. Okay. So are we up to 14 now? Is that right? Okay, FECC-14 is a little bit different: healthcare provider communicated with school staff about child's condition. And we have Marlene, Ricardo, and Sue are going to comment.

MEMBER QUINONEZ: I can go first. So this one is different than the others because it's supported by just a couple of studies and -- sorry, one study, that looked at kids with traumatic brain injury. And better outcomes if -- or perceived better outcomes by the authors if they had a communication with the school and if they transitioned back to school with good communication from the providers. So the evidence basis for this is fairly weak, almost consensus basis, based on one paper.

MEMBER MILLER: This is Marlene. I would completely agree with that. It's very

problematic to me that this is one study with traumatic brain injury where you could see -- the language of the measure goes on to talk about exact educational impact of the child's condition and it makes sense in traumatic brain injury, but it may not make as much sense for the whole myriad of other chronic conditions. And so it seemed this is a very, very large extrapolation with a TBI study involving 66 kids, so all kids with all chronic disease to warrant this level of communication with the details as specified in this measure.

MEMBER KONEK: There was additional information asked for on the call, although I wasn't actually participating on the call, I was listening. And they did come through with some additional information, an article by Weil about school reentry after cardiac transplantation; one by Hart which was a 2015 systematic review of 10 qualitative studies emphasizing the importance of communication, and also they cited the AAP Medical Home Policy of 2002 which basically says

that this is communication between medical home and the school is very important. So there was additional. Some of it, there was a systematic review, so that was something that wasn't there for the initial call, but that was sent to us on the 24th which was right before Thanksgiving.

CO-CHAIR BROOKEY: Right, that's very helpful. Maureen?

MEMBER EDIGER: To me, this just sounds like something else that parents are going to have to coordinate, because ideally your medical professionals are talking to the school, but in reality it's going to be the caregivers and the parents that are having to facilitate that so it can get checked off. And I think requiring it just might add another layer of burden or complication to families.

CO-CHAIR BROOKEY: Any other? Go ahead.

MEMBER BRISTOL-ROUSE: I would just add from the parent perspective too, having a child with special healthcare needs that as we

were kind of transitioning out of services I wouldn't have -- I was trying to downplay with the school system like to help normalize and move into that space. And so that's something as a parent I would have wanted to solicit from my provider to connect with the school. And I know that's not appropriate for every family, but just from my experience.

CO-CHAIR BROOKEY: And speaking as a pediatrician, I think there's a few in the room, we wouldn't do that without the parent's permission and probably without the parent asking us to do that. So it's just to say from a logical perspective, although it's very helpful what you stated in terms of the support for this practice.

MEMBER BRISTOL-ROUSE: Right.

CO-CHAIR BROOKEY: It isn't just the provider, it's the parents and the provider, right?

MEMBER KONEK: Right. I do have another comment. The thing that I -- when I

looked at this more closely it emphasized that what the coordinator can bring to the school is information perhaps and of course with the family, perhaps, but to -- about whatever their condition is or chronic condition, complex condition is, it results in their ability to learn. It's not just how they're doing physically, it's a lot more than that. And that, I think, goes into the individual education plans and things like that. I learn more by reading those things.

CO-CHAIR BROOKEY: So again, we're talking -- go ahead.

MEMBER AGORATUS: Hi. I'm also a parent of a child with five life-threatening conditions and autism, just to keep things interesting. And one of the tools that we utilize at the Parent Training and Information Centers is an individual health plan, which is an addendum to the individual education plan. So yes, while it would be helpful for a pediatrician to do this, it may not be necessary if everything

is in writing and it's being followed.

CO-CHAIR BROOKEY: Great. Great point. From the developer's standpoint, anything you want to add? We are talking just about the evidence right now?

DR. MANGIONE-SMITH: The only concern that I would put out there for people is I think we have incredibly savvy parents around the table, and on the phone and I would worry a little bit about parents who cannot advocate for their child in the same way as what we're hearing.

CO-CHAIR BROOKEY: All good points. Go ahead, David.

MEMBER KELLER: I would say similarly that part of our routine care coordination process, we've been piloting care coordinators in our office since 1997, so we've been doing this for a while. When we get a family hooked up with our care coordinator, one of the first things we do is get permission to talk to school as -- we never do it without letting the family know that

we're doing it, but we always have an up-to-date consent form so that we can have that communication. So I do think this is really important.

I also think it hasn't been studied well. I just took a look through the studies that they sent and of course, they were in very specific instances of brain trauma and cardiac disease. There's some work on communicating around asthma that I think probably has gone a little bit further than most, but the area of communication with the school has been poorly studied, even though having the school as part of the team when you're dealing with a child with special -- with medical complexity is acknowledged by just about everyone. So I agree that there is very little evidence, but there is certainly a lot of clinical juice behind this.

CO-CHAIR BROOKEY: Martha?

MEMBER BERGREN: So one of the things I didn't mention before is that prior to becoming a faculty member, I was a school nurse in four

states. And the importance of the school understanding the child's health conditions and the program of care is essential. And if you don't have it, you are flying in the dark.

I agree that those of you around the table and on the phone are very savvy parents who can probably case manage your own kids without the assistance of the school, but often it's the school that's the case manager and is the only person who's talking to every specialist and the primary care providers.

So I'll take it upon myself as having previously worked for the National Association of School Nurses that we need to up the ante on the evidence in this area. I will commit to that, but I really do think it's essential. And I agree with David that it really is on the primary care provider in the school to initiate that consent, the authorization to exchange information, so that all the care that's going on in the school is communicated with the primary care provider, and the primary care provider's

goals for that child, but both health and educationally, are known to the school.

CO-CHAIR BROOKEY: Karen.

MEMBER DORSEY: I was just going to say that in a case like this where we imagine the entire spectrum from this kind of requirement being intrusive for some families and seen as not helpful, and being essential for other families, this may be an area where a stratified measure may be more appropriate. And I think we don't have the evidence to even think about how to stratify it yet. So I concur that it's important, but not necessarily important in the same way for all families. And so really that kind of nuance needs to be better explored.

MEMBER BRISTOL-ROUSE: And I think that's my response, because I absolutely hear that those of us at this table are anomalies, very much, in many ways, but I'll own that, Maureen, but I think it is essential for some families. So I think it's what the -- when you're talking about making providers accountable

it's a different level and so what is the opt out? Are you getting consent before every kind of interaction you would have with the school instead of me saying no, I don't want it, and then my provider getting dinged because they didn't do it? So I think that's where the issue lies.

CO-CHAIR BROOKEY: So we have a few flags up. Are there new comments, then go ahead. Otherwise, we will move towards a vote. Maureen?

MEMBER EDIGER: I don't want to be misunderstood, but I don't think that this is a great idea. I just think that it's not helpful to the provider that it's something else that they are now accountable for. And absolutely should families or kids -- really, the kids that need that communication, absolutely, it should be there. I just don't think this is the right context for it.

CO-CHAIR BROOKEY: I think there are some really good comments made. We're going to vote on the evidence for this measure which

sounds like it was fairly low. And so the question would be whether we can even actually vote this as high, or whether it's going to turn out to be at most a moderate. Is that an individual decision or can we make the decision?

DR. NISHIMI: It was not based on a systematic review.

CO-CHAIR BROOKEY: Correct. So I think the highest should be a moderate in this case. Okay.

MEMBER KELLER: This was based on a systematic review conducted, as was everything else here. It's just that the systematic review didn't find anything, which is different.

CO-CHAIR BROOKEY: All right.

DR. NISHIMI: The evidence supporting it is a single paper, derived from it systematically.

MEMBER KELLER: Derived from that overall -- the same systematic review they did to try and back up everything else they did. So I don't think we can say they didn't do a

systematic -- the fact that a systematic review finds no evidence --

DR. NISHIMI: That's fair.

CO-CHAIR BROOKEY: So you can individually decide whether or not it was sufficient and score it a high. Is that fair? So if you feel the systematic review was thorough enough to score it a high, go for it, otherwise you can score it a moderate, low, or insufficient. And then if it's insufficient, we can go further. So let's go ahead and vote on this measure.

MS. CHAVEZ: Okay, now voting on FECC-14 for evidence. One, high; 2, moderate; 3, low; 4, insufficient. Thirteen, 22, 24. Okay, zero voted high; 2 voted moderate; 15 voted low; 7 voted insufficient. This does not pass evidence.

DR. NISHIMI: So when we go to the next criterion, which is gap, we will not consider this.

CO-CHAIR BROOKEY: Okay, again, we're

changing the topic a little bit on FECC-15, the caregiver has access to medical interpreter when needed. And the leads are Lauren and Jim. Lauren, would you like to go first?

MEMBER AGORATUS: Sure. Under evidence, I had a general comment that there needs to be more on cultural competency, not just professional translation. And another comment that there was no grading system. There were concerns in other areas as well. Another general comment where health disparities should be addressed in all measures, not just this one. Concerns about sample size and reliability and validity, which I guess we'll get to later. That's it.

CO-CHAIR BROOKEY: Thank you.

MEMBER BOST: Just briefly, there were a lot more studies associated with this evidence-based assessment, but they were not all about complex kids and that the item is actually scored always, sometimes, never -- always, usually, sometimes, and never.

CO-CHAIR BROOKEY: Okay. Any comments or questions about the evidence for this measure? Kevin?

MEMBER SLAVIN: Does it matter that this is a legal requirement, at least in the state that I practice? If we don't speak the language, we are legally obligated to find somebody to help us interpret, so in my mind it doesn't matter what the evidence says. If you're not doing this, you're practicing outside of the law.

DR. LION: While that is certainly true in both my clinical practice and in loads and loads of studies, there's lots of evidence that that does not stop people from not using a professional interpreter to communicate with limited English proficient families. So we know that even though it is, in fact, the law and it's a federal law, people still don't do it.

MEMBER AGORATUS: This is Lauren. I have to concur with that. As a bilingual advocate, I hear from Spanish-speaking families

all the time and unfortunately, sometimes they use children as translators. And so the medical information isn't even accurate.

CO-CHAIR BROOKEY: Or housekeepers. Good point, but I have to agree that it's underutilized. The AT&T line is underutilized. Interpreters are underutilized. I think most of us would concur. Any other questions about the evidence of the measure?

MEMBER DORSEY: Sorry, I just wanted to say that this may be one instance to Jim's point where the literature review criteria was too stringent. I mean given that in many, many situations in the healthcare environment, a translation has been demonstrated to be critical. I don't see a reason why we would not extrapolate from that body of evidence to this patient population which even more -- has more of a need for clarity and communication. So I think I just want to advocate that we be a little more lenient in terms of understanding that the entire breadth of the literature probably applies in this

instance.

CO-CHAIR BROOKEY: Okay, let's vote.

MS. CHAVEZ: We're preparing to vote on evidence for -- this is for 15 for those on the phone. 270-15. We're at 18, 20, 21, we're 15 24. Five for high; 19 for moderate; 0 for low; 0 for insufficient.

CO-CHAIR BROOKEY: Very good. I think we have two more of these to vote on evidence. Do you think we'll finish all the rest of them by 3:30? What do you think? The FECC-16 is child has shared care plan. And the leads on this discussion will be -- is Craig here? Craig and Karen.

MEMBER KNUDSEN: In terms of evidence on this one, the reviewers found it had pretty strong evidence, actually. There were seven RCTs done on this; three cohort studies, seven case series studies; and two consensus statements, one from AAP. And they all showed better outcomes with shared care plans. So that's pretty much the evidence there.

CO-CHAIR BROOKEY: Karen?

MEMBER DORSEY: I don't have anything to add.

CO-CHAIR BROOKEY: Any comments or questions? Go ahead, David.

MEMBER KELLER: So I confess I didn't look up all the studies, so this is a question for the developers. Shared care plan is an interesting concept. We're in the middle of trying to define it in my institution, and it's challenging. How much commonality was there between the definition of a shared care plan and how did you -- in these different studies, and is that reflected in the questions that you actually used for this measure?

DR. LION: So we were limited to some extent again by the amount of detail that authors chose to provide. There was some instances where a single study was described in multiple different publications where we could find more details in one of the publications compared to others. There were certainly not enough

evidence, enough detail in the descriptions of the shared care plans for us to be able to identify particular aspects or elements of shared care plans. They were more likely to be associated with better outcomes although we certainly tried.

In thinking about how to actually conceptualize a shared care plan, we had a number of criteria that needed to be met. It needed to be described as a shared care plan or an individualized tailored to that particular patient and/or family. It needed to be developed by the patient and family in conjunction with the primary care provider or a care coordinator and then shared with a primary care provider. It could also incorporate other providers in a multi-disciplinary team.

There was quite a bit of variety in the different studies in who all was involved in the multi-disciplinary teams, or who was being shared with, indeed. But we found a fair degree of variability in how things were described, but

we tried to sort of identify the lowest common denominator in terms of a shared plan developed with a patient or family and the PCP or care coordinator.

CO-CHAIR BROOKEY: Questions, comments? Go ahead, Ricardo.

MEMBER QUINONEZ: So as a recipient of shared care plans often as a hospitalist, I find that a large percent of them are not updated. And so you could argue that that would actually make outcomes worse, since we are likely to put in the -- I mean, obviously, we should ask the family, but sometimes residents will put in the orders for wrong medications, for the wrong dose, etcetera, etcetera. Was there a discussion of adding an updated care plan and what the condition would be? I just see that as an unintended consequence.

DR. LION: Fantastic questions. So the way the measure was initially specified, we did actually include a subpart looking at when -- if it was not developed in the past year, whether

it was updated in the past year. Unfortunately, because -- and this gets some into the performance gap, but because the overall performance on having a shared care plan at all was so low, what -- what we found was -- overall only about 40 percent of kids had one.

If you did have one, your performance on the subparts, including being updated in the past year was actually pretty good. So we saw ceiling effects on the subparts, but a very relatively poor performance overall. So we ended up at this point dropping those subparts from the measure, because it didn't seem, at this point, worth measuring, but in the future when shared care plans have better uptake, perhaps.

CO-CHAIR BROOKEY: And I'm sure your resident's reconciled medication was 100 percent of the time. Okay, can we move to vote on this one? Alright, sounds like we feel more favorably about this one, so let's go ahead and move to vote.

MS. CHAVEZ: Okay, now voting on

evidence for FECC-16. One, high; 2, moderate; 3, low; 4, insufficient. Thirteen, 23, 24. Twelve voted high; 11 voted moderate; 1 voted low; and 0 for insufficient.

CO-CHAIR BROOKEY: Okay, so we're voting on our last one for evidence. This is FECC-17, child has emergency care plan. And according to what my notes are, there was not an empirical evidence review provided. Unless that's changed -- has that changed? Okay, so, developer, go ahead and fill us in on the changes.

DR. LION: So following the work group call, we took another look at the evidence and we identified two, in addition to the two AEP consensus statements related to emergency care plans would strongly endorse their use. We found two additional studies. One was -- where is it? So we identified a randomized control trial. However, it had fairly poor follow-up, so it randomized 170 patients with complex congenital cardiac disease to a program of a web-based

emergency care plan for the ones who were randomized to the intervention.

Unfortunately, 35 percent of the enrolled participants completed both the baseline and follow-up surveys, however, there were approximately even numbers in both the intervention and control groups. And there was an improvement in the intervention group in the parent's perception of emergency care provider's comfort and ability to care for their child.

Interestingly, however, the emergency care plans were only accessed in 13 out of 100-ish actual emergency room visits. So it's unclear whether it just improved parent comfort, knowing that it was there, or whether it actually changed the way care was provided.

So while it was an RCT, I gave that a level 3 to 4 evidence because of the low follow-up. And then we also found essentially a cohort study describing an intervention with emergency care plans for children with life-threatening asthma. However, there was no clear comparison

group identified and so even though they reported decreased hospitalizations and decreased deaths associated with asthma, it was unclear what exactly they were comparing that to. So although it was pitched as a cohort study, I think it was probably more of an expert statement. So we did find some -- a description of a nice program.

CO-CHAIR BROOKEY: So mixed findings. Karen and Lauren, do you want to make comments?

MEMBER AGORATUS: Sure. I had the same kind of concerns that it was an outcome based that more information was requested on the American Academy of Pediatrics consensus statement which was provided. Again, the comment of no empirical evidence, but that was a draft.

Also, that the AAP recommends emergency preparedness for natural disasters in addition to emergency care plans. Also, the last comment was that some of the evidence in adult studies isn't applicable to children. That's it.

MEMBER DORSEY: So given the sort of lack of evidence supporting these in this area,

you know, I'm thinking about it as something -- considering the threshold of whether it's so important to measure that we sort of say it's okay, that the evidence is insufficient. And I'm a little torn because I feel like it's an incredibly important thing, right, that having emergency plan is like one of those high level sort of portability of medical information goals. But I fear what you just said, which is that we're not really that portable yet and that even when we have them in the moment of an emergency we're still not at that place where it's easily accessed, consistently accessed. So that's my thinking on it.

CO-CHAIR BROOKEY: Comments? Ricardo?

MEMBER QUINONEZ: Just a clarification, so it sounds like from the evidence we should only be voting on insufficient or insufficient with exception?

CO-CHAIR BROOKEY: Well, I think we have to determine whether that was a systematic review or not, and if you think it is and it's

sufficient, you can vote high or moderate. If you don't buy that that's sufficient, then it would be limited to low or insufficient, but it could still be voted again for with exception. So we'll make it an individual choice as to whether or not this latest information is going to influence your vote. Any other comments or questions? You're all getting kind of tired, aren't you? So let's vote.

MS. CHAVEZ: Now voting on evidence for FECC-17. One, high; 2, moderate; 3, low; 4, insufficient. Fifteen, 22, 24. Zero voted high. Three voted moderate. Nine voted low. And 12 voted insufficient.

CO-CHAIR BROOKEY: That's 50 percent for insufficient. So this does not pass or do we vote on with exception? So this measure will not move forward then. Did we change the rules? Why don't we take a break? We'll look at the rule book. We'll come back in five minutes, and please be back in your seats in five minutes. We'll keep going.

(Whereupon, the above-entitled matter went off the record at 3:31 p.m. and resumed at 3:40 p.m.)

CO-CHAIR BROOKEY: We've huddled about whether we're going to be able to vote for this last measure, for an exception, with insufficient evidence, and the answer is we don't know.

So I'm going to ask whether it's reasonable that -- I'm going to ask if it's reasonable that we go ahead and vote, and then if we have to retract it later we will, but I think it's easier to vote now than to have to vote later, since it's fresh on our minds. Does that make sense?

So if everybody can get your clickers, and get ready to vote, we're going to go back to the last measure, which was, I think, 17, is that right?

And since we have 50 percent insufficient, we're going to go ahead and vote on whether you think the measure is still important enough to move it forward, with exception.

So we're going to vote here in ten seconds. So get ready, set, go ahead. Seventeen insufficient evidence, with exception.

We're voting on exception.

MS. CHAVEZ: Okay, for those on the phone, we're voting on exception. Choices are one, insufficient evidence with exception, two for no exception. Measure FECC 17. I have 16, 20, 20, eight seconds, 22, 23.

Okay. Okay, so eight voted insufficient evidence with exception, 15 voted for no exception.

CO-CHAIR BROOKEY: Okay. And so those of you who missed my earlier comments, we're voting. And we're not really clear if we're allowed to, but if it turns out that we can't vote because of the rules, we'll take this back later. But we just want to go ahead and get the vote in, just because it's all fresh on our minds.

Now, we're going to ask the Committee's permission to do a little change in

schedule. Just -- so I think we've gotten through the hardest piece of these ten measures. Some of these will not go forward for further voting, but many of the elements of GAP and use and usability and feasibility, we can vote en bloc, in terms of just one vote for all the measures. So the only thing we have to go through for each and every measure will be the reliability and validity.

We do have a group here with the Adolescent Measures, the ADAPT, that will not be here tomorrow morning. However, our folks here will be here tomorrow morning, so we're going to ask if we could do --

(Laughter.)

CO-CHAIR BROOKEY: We're going to give them a -- we're going to give them a nice sleep, to get rested up for tomorrow morning.

Is there any objection to postponing the rest of the conversation about these ten measures until tomorrow morning, so we can have the ADAPT conversation now? And we'll do as much

as we can do to get through five o'clock, but any objection to that? So if not, let's move forward to Adolescent Assessment of Preparation for Transition to Adult-Focused Health Care. And we have our developers coming to the table.

So if you could give us a brief overview. And, again, primarily focused in the beginning, at least, on the evidence for the review.

DR. SAWICKI: Good afternoon. I guess we're the last team before you at the end of the day. My name's Gregory Sawicki. I'm a pediatric pulmonologist and health services researcher at the Center for Excellence for Pediatric Quality Measurement, at Boston Children's Hospital, joined today with my colleagues, Doctors Sara Toomey and Mark Schuster, and we represent the team that developed the Adolescent Assessment of Preparation for Transition, or ADAPT Measure.

Health care transition is a process by which adolescents and young adults shift from pediatric-focused to adult-focused health care

delivery.

Multiple guidelines, expert panels, consensus statements, and patient advocacy groups have called on the medical profession to prepare adolescents for developmentally-appropriate care, whether they are changing providers, or staying with the same clinical team.

Transition should be planned, purposeful and have the goal of providing uninterrupted high-quality care. The lack of effective transition may contribute to fragmentation of health care and increased risk for adverse health outcomes, particularly during youth adulthood.

There is broad consensus that preparation for health care transitions should start in adolescents and involve individualized planning. Transition services are, therefore, key aspects of high-quality care for adolescents, particularly those with chronic health conditions.

In a joint 2011 clinical report, the

American Academy of Pediatrics, the American Academy of Family Physicians, and the American College of Physicians provided a consensus-based framework for physicians to implement high-quality, developmentally-appropriate health care services for transition.

This framework includes recommendations for providers to assess transition readiness, develop a transition plan, and document plans and health records as part of the medical home.

Results of a recent randomized study in primary care practices in Washington, D.C. demonstrated that implementing recommended elements of transition preparation improved the quality of care coordination for youth with special health care needs.

Other condition-specific studies, including populations of youth with diabetes, cystic fibrosis, and congenital heart disease, have established a link between structured transition efforts and improved outcomes,

including better transition readiness and improved engagement in the adult system.

Although evidence is emerging to link efforts to improve transition preparation with longer term outcomes, we acknowledge that the level of evidence, currently, is likely to be insufficient at this time, and we believe that our way forward here may be through the exception approach.

There is, however, a large body of evidence that, for many adolescents with chronic health conditions, preparation for transition is seriously inadequate.

In repeated national survey studies, only a minority of youth with chronic health conditions, or their parents, report having discussed transition with their physician, or having a plan addressing transition needs.

As an example, in our development focus groups, we heard, more than once, the youth were only informed of their need to transfer care by administrative staff, rather than their health

care provider.

We developed ADAPT as part of the Pediatric Quality Measures Program sponsored by AHRQ and CMS. ADAPT is a survey instrument for adolescents ages 16 and 17 years old, in which they report on whether specific aspects of care related to transition preparation occurred.

We reviewed the transition literature and conducted interviews with expert researchers and clinicians. Focus groups were then conducted in three U.S. cities, with youth with one or more chronic health conditions, as well as with parents of those with chronic health conditions.

From this formative work, we drafted an initial survey and conducted cognitive interviews in English and Spanish in three U.S. cities.

We then fielded the survey, by mail, in English and Spanish, in one pediatric hospital and two Medicaid health plans. We received over 1,600 surveys and used the data for psychometric testing, composite development, and case mix

adjustments.

ADAPT consists of 26 questions assessing the quality of transition preparation as reported by youth with chronic health conditions, and generates a score for each of three domains: counseling on transition self-management, counseling on prescription medication, and transfer planning. We have presented evidence that the measure is both reliable and valid.

We believe that ADAPT fills an important need for publically-available measures of health care quality for adolescents. It will serve as a valid and valuable tool to assess health system quality and motivate improvements in care delivery. We look forward to your comments.

CO-CHAIR BROOKEY: Okay, I believe it's Marlene, Amy, and Kerri. I don't know who wants to start.

MEMBER MILLER: This is Marlene. I can start. So it's not -- in terms of the

evidence, the report in here, I don't really see that there was a real systematic literature review, there's a couple of quasi-experimental studies that are referenced.

But really, there's not evidence, in terms of -- and I think it's really important. I think it's evidence that physician counseling, which is what this survey solely measures -- and that is an intervention, but that will achieve transition readiness.

I think it's undoubted that preparing teams for a transition to adult care is important. What there's no evidence for, however, is that provider counseling -- which all the questions on this survey say, did your provider, did your provider, are the ways to achieve that.

I mean, I know we're -- in our institution we're looking at things with, you know, the nurse extenders and the -- in simulation, and all other things that aren't necessarily that one provider and are much more

longitudinal than just a counseling session, which could mean one of a million things in terms of how long that actually happened.

So I was -- I don't think the evidence -- the evidence is insufficient, for sure, it's not low. But more particularly, I didn't really appreciate in the write-up the lack of attention to the fact that this is putting forth an intervention that it's all the providers counseling on this, and that that is -- there's no evidence that that provider, alone, will achieve transition readiness at counseling.

MEMBER HOUTROW: This is Amy. Just to give a little bit more context. So this was implemented in a specific age group, 16 to 17, and that came by consensus. And there was a lot of discussion about whether the evidence supports earlier, like the AAP recommends, or potentially later, now that health plans have changed when kids have to transition.

And the domains that they were interested in are, specifically, counseling on

transition self-management, counseling on prescription medications, and then, transfer planning, so it's three different areas. And then they used a survey of 26 questions to get at those, and so numerators and denominators are important here.

The proportion of positive responses on five questions for the counseling on transition self-readiness, the positive response on three questions for medication, and four questions for transition planning.

And then the denominator's respondents, who had valid responses to that -- and in the way that this is intended to be used could be at the individual practice level, or at the health plan level. I think also something for us to consider when we consider the evidence.

Some of the other concerns that were brought up about -- in addition to, if the provider does that, does that lead to a specific change in the patient, whether they feel more comfortable, whether their outcomes are

different, in addition that this may not capture enough regarding specific groups of individuals, who may be in high need of transition services but may not be able to participate in the transition readiness in this way, such as children with developmental disabilities, who couldn't even answer a survey like this, nor could they engage in a -- necessarily, in a dialogue with their provider about those aspects of care.

So I think in that -- Marlene's point about, is this -- if we measure something that a provider does, is there evidence that that will change something at the patient level is a key point.

CO-CHAIR BROOKEY: Kerri.

MEMBER FEI: I don't have a lot to add, except the fact that this is a patient-reported outcome measure and it is measuring how ready these kids feel to start taking care of themselves, and I think that's important.

And I understand the back and forth

about whatever the provider does, does that affect how they feel? As this stands, it's like a survey measure. And given that, with the evidence, when we look at the algorithm, it's going to be a pass/fail.

It's not going to be -- it either will pass, or it won't pass. So we won't go down the list, am I correct?

They say it at the top box. Because this is a -- it's a patient-reported outcome measure.

CO-CHAIR BROOKEY: Which is what we're going to be voting on, and so I think it's important to ask the question, again, just to be clear in everyone's mind, about the denominator.

Because I think, probably, if we go around the room, everyone's going to be challenged with the 16/17. And, can we speak a little bit more to that, so that we know what we're voting on? So, Jenny.

MEMBER MOYER: My question, really, is about the previous comment. This did not strike

me as a patient-reported outcome, this is a survey of the patient to see what their provider did. Did your provider have you talk to this provider? Did this provider talk to you?

And those things that -- if the change that we're looking for is the provider changing their behavior, then this is not a patient-reported outcome, it's a survey of the patients, to find out how they assess their care.

MEMBER FEI: I don't know if you guys want to speak to that? I mean, I view it as any type of patient -- almost like a patient experience measure.

DR. SAWICKI: Well it is a patient experience measure.

MEMBER FEI: Right.

DR. SAWICKI: And whether that's considered a patient-reported outcome versus a patient survey. I mean, this is a -- you're correct that this is a survey asking an adolescent to report on what their provider has told them, or counseled them about, so it is a

patient experience measure.

MEMBER FEI: It was submitted as a patient-reported outcome measure in the materials, though, just so --

CO-CHAIR BROOKEY: Does everyone understand the implication of that? So if we say that it is patient-reported outcome, we only have a yes or no vote. And so if we don't believe it is, it can go down the path of going from insufficient evidence and we can actually go for exception.

But if we do say yes, it's a patient-reported outcome, then it's either going to pass or not pass. So that's why the pathway will be different, depending on what we consider this to be.

So is there any objection to moving this forward as a patient-reported outcome survey? I've heard one objection.

Any other comment?

MEMBER FEI: And here's the thing, I don't know that -- can we decide that. It's the

developers that submitted it that way and if -- I'm not, that's where, maybe, we need NQF's -- that we need staff involvement here, because I'm not sure that we can tell them what to do. I don't think that's the right way to go.

CO-CHAIR BROOKEY: I think you can just go by what they have on the screen, what they've determined it to be, so --

MEMBER THACKERAY: Yes, if I'm reading the box correctly, patient-reported outcome, or patient-reported experience, both follow the same pathway.

CO-CHAIR BROOKEY: Right.

MEMBER MOYER: I'm looking back at the FECC that is a patient-reported measure. That's a patient survey to find out whether things happened.

So I think this one and that one are, conceptually, exactly the same and should be handled the same way.

CO-CHAIR BROOKEY: So let's see, Jeff.

MEMBER SCHIFF: I just had a couple of

questions. This is mail and survey only, so I'm wondering if there's a -- if you have any experience with how many, were they distributed by mail to -- for dispense?

DR. SAWICKI: For --

MEMBER SCHIFF: It's a little unclear, as to what they --

DR. SAWICKI: For our field test, the surveys were sent, by mail, to families with an explanation and cover letter, to ask the families to have the youth complete the survey.

MEMBER SCHIFF: Okay. For foster care -- you said there was an exclusion for court law enforcement. Foster care was included, or not? Was that included in the court?

DR. SAWICKI: I don't recall if we had that come up at all, with the -- the patients were identified in one of two ways. One through administrative health plan data, and the second was through a clinical program giving us information about their patient panels.

MEMBER SCHIFF: Okay. In English

only?

DR. SAWICKI: English and Spanish.

MEMBER SCHIFF: Okay.

CO-CHAIR BROOKEY: So we're going to ask Karen to make some comments about the pathways here, for these two sets of measures, can you just go ahead and speak to that for us?

MEMBER JOHNSON: So the way, I think, we -- oh. I'm sorry. I'm Karen Johnson. I'm one of the Senior Directors here at NQF. The way that we would handle this measure is we would consider this a PRO -- PRO-PM, a measure. It is a patient experience measure, so it would go down that pathway of the green box.

But be very clear, by doing this, we're not looking for the kind of evidence that you were looking for for the earlier measure, so we're not interested in quantity, quality, consistency.

What you're wanting to know here is whether there is something that providers can do to affect this outcome or this patient

experience. So that would be -- that's why you have a yes or no there.

CO-CHAIR SUSMAN: I have a fundamental concern that, I think, Virginia already talked about, which is, this seems exactly like the FECC measures and we've treated it very much differently. It just doesn't compute here, from a consistency and a reliability standpoint.

MEMBER JOHNSON: So the measure that you just talked about, the FECC measure, each of the individual performance measures within, even though it was off of one survey, each of those measures were process measures. Right. Did you have a care coordination? Did they do this? Did they do that? Each of those are process measures and those go down the QQC pathway. These are not process measures. They're not asking them, did the provider do something, this is an experience measure.

MEMBER MOYER: And as I read --

CO-CHAIR SUSMAN: You read the --

MEMBER MOYER: I'm reading the survey

and reading the survey, it's almost the same. I mean, did you and your provider talk --

CO-CHAIR SUSMAN: Yes.

MEMBER MOYER: -- did your provider talk to you about refilling your own prescriptions? How often did you schedule your own appointments? Did your provider talk about your health insurance? These are -- to me, these are materially the same as the FECC questions there. It's a patient experience of care, but it's not a patient outcome. Do you see --

CO-CHAIR SUSMAN: I'm not arguing one way or the other about what it should be, it just seems that we would be treating this measure and the FECC measure in a fundamentally different way when, I think, they're the same exact concept.

MEMBER MOYER: But what are we --

MEMBER MILLER: And this is --

MEMBER MOYER: -- it to?

MEMBER MILLER: This is Marlene and I will just chime in. I also agree, being a reviewer, both primary reviewer on the FECC

measure and these, these strike me as the same, and this little nuance here was lost on me because they seemed identical as surveys of processes of, did these various things happen?

MEMBER HOUTROW: And this could have been, it could have been presented to us as a measure of process, in which we looked in the medical record to see if the physician documented that he had a conversation about medication, right?

And it's the same process that we're looking for evidence of, whether you survey the parent, or whether you survey the child, or you survey the medical record.

But, my question comes back to, what are we obligated to do with number one, if we're being told this is a patient-reported outcome?

MEMBER JOHNSON: So if it is considered a PRO, then the question for you is, is there something that providers can do, at least one thing that providers can do, to affect the -- basically, the results that somebody would

report on the survey?

CO-CHAIR BROOKEY: David.

MEMBER KELLER: So it -- boy, this is an interesting existential question, isn't it?

(Laughter.)

MEMBER KELLER: Because, I'd have to agree that, you know, in the construct of the triple aim, patient experience is an outcome. And, yet, and so I think our confusion is not so much, I get why this is being considered an outcome, our confusion is, really, why was the other not considered an outcome?

CO-CHAIR BROOKEY: Right.

MEMBER KELLER: Because, yes, they were looking at processes. But, again, they were looking at it from the patient's point of view, which would have made it a patient experience, rather. And I'm thinking, for consistency, we -- I'm now confused. Is it just -- is it why we didn't pick up on this before?

I mean, and, you know, and it gives us a little bit of reflection. I get why we -- but

I actually think we should treat this this way. I guess, I'm thinking we made a mistake in not treating the other one this -- that way, in the same manner.

MEMBER HOUTROW: Those aims are better health, which is the patient-reported outcome; better experience, which is patient-reported; and better efficiency of care, right, which is a health systems issue. So you think -- right, health care, yes.

DR. NISHIMI: Please. Mic, please. Microphone, please.

MEMBER MORROW-GORTON: So just thinking about it, wasn't the one we just talked about based on chart review, not based on what the patient, or family, said?

Or was it a survey? Okay.

CO-CHAIR SUSMAN: It's a series of nested surveys.

MEMBER MORROW-GORTON: Okay.

CO-CHAIR SUSMAN: If that's the case, I mean, I guess, one option is for us to review

what we did with FECC and use the same standard. I don't think it's going to change, actually, anything we decided about evidence, which is pass, or no pass.

CO-CHAIR BROOKEY: Yes, I think so. I would recommend that we go ahead and consider this one to be a health patient-reported outcome and take it though the top of the algorithm.

I think that we can huddle about the others, but I think that it wouldn't be very difficult to flip those into the same questions and we've already had the discussion, so I don't think it would take very long to go ahead and go through those, tomorrow, while we're -- maybe, if we finish early, today.

So that would be one recommendation. But I think, maybe, we should go ahead and consider this one to be a patient-reported outcome. So we have a bunch of people here, so, Kevin, why don't you go ahead and start.

MEMBER SLAVIN: Yes. I think --

CO-CHAIR BROOKEY: There you go.

MEMBER SLAVIN: Sorry. Guess I shouldn't hold it. I think if we were to flip the FECC measures into this same patient-reported outcome, they would have to be taken en bloc, rather than as individual pieces.

Because we're no longer so concerned about the individual evidence and whatever and it would have to be, sort of, looked at the same way that this is.

CO-CHAIR BROOKEY: Let me just comment that the discussion we just had is still relevant and important for future tweaking of those measures, so we would take that into consideration, whether we would just go ahead and vote them all, as ten, ten at once. So good point.

MEMBER FEI: The other thing, for the FECC measures, since they were submitted as process measures, we'd have to get the developer's agreement that they would want them considered as patient-reported outcome measures, as well.

CO-CHAIR BROOKEY: Okay.

MEMBER QUINONEZ: Yes, I think there's no absolute objective way to handle the question that we're arguing right now, and so I would agree with the fact that we should consider this one a patient-reported outcome, but we should leave the other one as it stood. Because, it was a robust conversation, and I think, you know, we voted based on that robust conversation.

CO-CHAIR BROOKEY: Okay, we're going to take all these comments into consideration. So thank you. Who else has their hand up here? Jeff.

MEMBER SCHIFF: I have to make up for lost time. I wanted to -- I had, I wanted to ask you for the NQF staff, if there's a difference in the quality of the endorsement coming in as a patient-reported outcome versus a process measure? I'm getting a no. Okay.

MEMBER JOHNSON: For the record, there is not.

MEMBER SCHIFF: Okay. And then, I

just want to be -- I just, this is to David's existential comment about this. I think we have to be clear that there's a lot of patient-reported things that are process, but because they're reported in document and the medical record, they become a process measure of care, be it a depression screen or anything like that. If it's -- it seems like what we're really talking about is whether the collection mechanism occurs outside of the office to assess the quality of the care in the office, or the patient's outcome.

CO-CHAIR BROOKEY: Carol.

MEMBER STANLEY: So I'm a little confused. With this particular measure, it seems like some of it's patient experience and some of it is outcome. Because you ask about if the physician talked about your prescription drug use. But then there's another question that asks or about making your own appointments. And then it asks, specifically, your behavior, have you scheduled your own appointment.

So one of those is an indicator, because the behavior has -- there's an actionable behavior, which is an outcome, versus the process measure is, were you counseled about making your own appointment?

I also have concerns about the age group that's targeted for this survey, age 16 to 17. Is it realistic for a 16-year-old to call and make their own appointment and to -- and call and get their prescriptions refilled? So that's part of the questions that we're talking about and --

CO-CHAIR BROOKEY: Right. Karen.

MEMBER DORSEY: I just want to, you know -- it just occurred to me, just hearing the conversation, that this is a little bit different from the FECC, because -- I mean, one way to think about it is that the transition from pediatric to adult care is inevitable, right. It's an inevitable part of the structure of our health care system.

And so I think you can reasonably

think about -- whether they be processes or behaviors around that transition, as being patient experience, because it's an experience they're going to have, one way or the other.

Whereas, care transitions, we'd have to sort of make an analogist argument, perhaps, right? And say, suggest every child with complex medical conditions requires the level of care transitions as -- I mean, care coordination that's described.

We may all agree to that, right, but it's a little bit qualitatively different to me, because, you know, pediatric to adult transition is inevitable. Nobody's in control of that. That's going to happen. And so this can be thought of, even if it describes processes, as an experience of that transition, no matter what.

CO-CHAIR BROOKEY: So we can take a couple more comments and then we're going to ask for a decision about voting for this, based on the top box.

I think what we're going to do, first

of all, we're going to go back to the FECC measures, we're not going to lose any of the conversation that we had, or any of the voting that we've made that will actually help the developers go back and look at the measures.

So everything that we did has been captured and will go back to the developers for future discussion. But, I believe we will go ahead and vote on them tomorrow, yes/no, according to the top box.

But, I think our conversation will help inform that vote. So I'd like to get the approval to move forward with voting yes, or no, on this particular measure. But are there any other comments, before we move in that direction? So, David.

CO-CHAIR BROOKEY: Carol.

DR. SAWICKI: Just a few points to respond to some of the comments that have been made. I think the comment that pediatric to adult transition is inevitable is, certainly, an appropriate and valid one. But, the literature

does tell us that there's a lot of deficiencies in how it happens and it's very haphazard. And I think that we would be remiss to not have some role for providers and health care systems to try to make it better. And it doesn't have to be, you know, the same level of care for everybody, but I think that ignoring that and just saying it's inevitable is, probably, missing the point.

And the comment around the role of providers, I think that, to the point that a provider could actually impact change, we spend a lot of time in developing this. So about specific wording, were we asking the adolescents to report on what their provider did, or whether it was somebody else in the office, we had a lot of comments that said, it's the social worker, the care coordinator, the nurse.

And, fundamentally, we agree that there are lots of different ways that transition care can be planned. But if the physician and provider is not part of that discussion, is not initiating or having these discussions, these

kind of programs are, probably, going to fall apart.

So even though we don't expect the provider to be doing everything, we do expect them to be having conversations. And so from an accountability perspective, we really did feel that it was important to ask about what the provider was doing.

And then, some of the things here about the age range, the measure simply asks whether you received some counseling. And there are some questions about whether you are doing certain actions, and there are some teenagers who are doing some of these things.

It's not an expectation of the measure or the measure score, but we didn't want to penalize someone, from a score perspective, if they had a teenager, or a teenage population that was doing really well and were filling their own prescriptions. For instance, for birth control for an adolescent female, they are responsible for doing that, in many situations.

We didn't want them to get a lower score, because they were told that they're not talking about it. Because they maybe mentioned this when someone was 15 and was asking for birth control.

So that was sort of part of routine adolescent care, and so the age range that was chosen didn't go as low down as the AAP, which says 14, and some consensus guidelines down to age 12, but we felt that, by age 16, some of these conversations should have been started.

CO-CHAIR BROOKEY: Just another question, before we vote. We have a couple more comments. Is there any consideration to adjusting the age range? Was that discussed, or is that sort of a final decision, based on consensus?

DR. SAWICKI: I think we talked about the appropriate age range. We felt that going below 16 would be inappropriate, from an adolescent-reported measure.

But we have, certainly, talked to

certain health systems, and there are, certainly, 18, 19, 20-year-olds -- these questions apply to a 25-year-old as well, and so I think there could be some discussion around increasing the age.

CO-CHAIR BROOKEY: Jenny.

MEMBER MOYER: Yes, my only question is that, I request that it be very clear what it is we're voting on, because I'm having difficulty reading the box, knowing how to interpret that.

MEMBER HOUTROW: That was my question, as well. So if I were to read this, I would say, rationale supports the relationship of the teenager reporting receipt of transition counseling to at least one health care structure process intervention or service, in this case, the process of a physician delivering counseling. Is that right?

CO-CHAIR BROOKEY: The green box says -- if I take out the extraneous words and just put in the PRO piece of it, it says, does the measure assess performance, from a PRO, because it's an or PRO, so it's -- whatever it's looking

at, it's just a PRO, and is that what we're assessing performance on? But that's not what the words on the screen are. Just look at the green -- do they all have this?

MEMBER HOUTROW: It seems a little bit circular, right? So the PRO is the patient's report of receipt of services, which is related to the process of physicians providing those services.

CO-CHAIR BROOKEY: But it's still a perception of the patient, because it's a patient-reported outcome, so it's not a definitive process measure where we go into the chart and see if they actually documented it, it's a patient's experience. So which may differ from what the chart would reflect, so that's why it's different.

MEMBER MOYER: But this is not a health outcome. And the first question is, is it a health outcome, or mortality --

CO-CHAIR BROOKEY: Or, or --

(Simultaneous speaking.)

CO-CHAIR BROOKEY: -- or a PRO.

Yes, so strike out outcome, we're looking -- it's just a PRO. Does this assess performance, based on patient's reported outcome? You have to look at the green box and not the screen up there.

MEMBER MOYER: And it's -- okay. And you're interpreting experience to mean the patient's experience of care, not the patient's experience of his health?

CO-CHAIR BROOKEY: Right.

MEMBER MOYER: Okay. I interpreted that as the patient's experience of his health. Okay, so patient's --

CO-CHAIR BROOKEY: Right. I would agree that that green box does not reconcile with that white screen. So we're voting on, does the measure assess performance from a patient-reported outcome survey, period.

MEMBER QUINONEZ: So, I -- and this might be just because I'm new at this, but I feel really uncomfortable voting on this top green

part, because it feels almost like that is way too low of a bar to set for an outcome, for a measure.

It basically ignores the evidence. And if you can prove that what you're trying to measure is measured by what you're trying to do, you pass. I mean, you could, basically, have no evidence for a measure and still pass.

CO-CHAIR BROOKEY: Right. So look at the white screen again. So the rationale supports that this measure assesses performance from a PRO. I think we need to change that question, otherwise, it's quite confusing.

So the rationale is all this discussion around the evidence, or lack of evidence, so you have to believe that it's strong enough to support a yes or a no -- a yes response on this. Yes.

DR. BURSTIN: Just a little bit of context. It's a good question, regard when it comes up all the time. And I'll say, when we had our evidence task force, I don't know, five, six

years ago, one of the key questions was, is the evidence requirement the same for an outcome measure versus a process instructional measure?

And the ultimate decision was that you could move an outcome measure forward, even if you didn't yet know the evidence-based processes that it's linked to, because sometimes having the outcome measure out there first drives the work to figure out the processes.

And the classic example is central-line associated bloodstream infection where, you know, that outcome measure was out, reported, and then the processes began to emerge.

I think there was a lot of concern of not holding back outcome measures that are otherwise important, while waiting for the --

MEMBER QUINONEZ: Except that it --

DR. BURSTIN: -- development of processes, yes.

MEMBER QUINONEZ: Except that example had a lot of --

DR. BURSTIN: Yes.

MEMBER QUINONEZ: -- evidence basis behind it.

You know, there was a lot of evidence that checklist helped and that --

DR. BURSTIN: Not at the time the measure was endorsed. Not initially.

MEMBER MOYER: The argument I would make is not that, because I completely agree with that. It's that what this outcome is -- this patient-reported outcome, is the patient reporting whether a process occurred, so it's a measure of whether a process occurred.

DR. BURSTIN: All of our PRO work to date, we have included patient experience within that bucket of PROs, which, I think, is what's confusing.

Most patient experience measures do report on, did you get the information in a way you can handle it, did you get -- they're often did you get, dot, dot, dot, so it's not that far off.

In Europe, for example, they'll separate PRO measures from PREMs, Patient-Reported Experience Measures. We've, at least -- the work we've done determined that they should be held to the same standard, but it's a fair question.

CO-CHAIR BROOKEY: I think what the Committee is struggling with is that there's a huge gap between voting for the evidence, or voting just for this rationale. And, it seems like we're holding this particular one to a different standard, which is why we talked about revisiting those earlier measures, tomorrow.

So I guess, the response would be that you have to listen to the rationale, what evidence there is or is not to support moving this forward, before voting yes on this. And it's just a completely different way of thinking about this measure.

There is no evidence. Insufficient evidence. We stated that this would go -- if this were to go for, down the linear -- the

vertical path, it would go down to whether or not this is limited, or insufficient evidence with exception.

DR. SAWICKI: The specific question that's being asked is -- and I think was raised with the first comment was, is there evidence to connect what a physician counsels an adolescent and an outcome. So that's one.

That's a specific question. There's plenty of evidence that says that physicians and care teams and health systems are doing a disservice to adolescents with chronic health conditions in not preparing them for adult care.

So I think, from a level of evidence perspective, this is an important topic to capture in a patient experience measure. That is -- that's the argument that our team will make.

This is, also, a novel measure in that it is actually asking the adolescent, so we're pediatricians, you know, a lot of us in the room, and a lot of the measures on transition that have been used by MCHB, by other organizations, have

been parent-reported measures and proxy measures.

And there is a place for proxy measures, particularly in children and young adults with cognitive delay and disabilities and things that they couldn't report on their own processes of care.

But there is a rationale to think about how adolescents are reporting on their care, and there is an evidence base to suggest that there is sub-optimal care being provided to adolescents.

CO-CHAIR BROOKEY: Just one more comment. And, at least in Kaiser, in California, our attorneys will not allow us to survey 16-year-olds. So we, can we -- there's this period between 12 and 18 that they're -- we cannot survey them. So --

DR. SAWICKI: In our field test, we worked with two Medicaid plans in Texas and in Pennsylvania and neither have that concern.

CO-CHAIR BROOKEY: It may be state-by-state, depending on their interpretation, I don't

know.

DR. SAWICKI: The mailing was directly to the parent, not to the adolescent, so we weren't allowed to mail directly to the adolescents.

CO-CHAIR BROOKEY: Okay. Any other questions, comments? Amy.

MEMBER HOUTROW: I just want to go back to the clarification of what outcome we're talking about and what process we're talking about. So the process of interest that underlies the outcome is that the physician did something to talk to the patient about transition. And the outcome is the report that the patient was talked to about transition. Not that they were ready for transition, or it changed their readiness, right? It's about, whether they just reported that that thing happened, as the outcome, which you have to then say that they are linked, because there's a clear relationship. If you do something, as a physician, and the patient reports that you did it, there's obviously a

link. If you didn't do it and they report that you didn't do it, then there's obviously a link. But that puts us out of talking about any of the kind of evidence that you were just speaking about, which there is a wealth of evidence that we're not doing well transitioning kids.

PART 4 Section B 4:22:03

CO-CHAIR BROOKEY: All right. So we're still talking a patient-reported outcome survey. So we're going to take it through the green track. And the question, and I'll let you go, David, but the question we're going to ask is, whether or not you feel that the rationale supports this particular PRO in, basically, assessing performance? David.

MEMBER KELLER: This is, actually, a comment from a while ago. Just back to the age. I would argue that the age that was selected by the measure developers was actually the perfect age to do this.

And I'm very, because what you are

trying to do is capture this before the children turn 18, is because what we're really trying to do is measure readiness for the transition and the transition happens at 18, the common one being around mental health and mental health medications.

And if you haven't prepped for that, at age 18, all of a sudden they get, just, all sorts of bad things happen. So I'd applaud that measure. I just wanted to put that out there, as a thought on the age criterion.

CO-CHAIR BROOKEY: Do people feel like they have enough information to vote, whether they support the rational? Hold on for just one second. Does the Committee feel like they have enough information to vote?

So I'm going to ask the question, you're going to tell me whether I got it right, okay? We're voting on this patient-reported outcome survey, as to whether or not the rationale, provided by the developers, supports the fact that this measures, that this survey

assesses performance on a patient-reported outcome survey. And I'm just putting all the words together, because it's not on one screen. Is that clear?

(Off microphone comment.)

CO-CHAIR BROOKEY: Well, let's talk about what's a factor. It is a patient-reported outcome survey, right? So according to the algorithm, it would go in the green pathway. And then, even though the green pathway doesn't state it, the question on the white screen is, we're really supporting the rationale that is supporting this survey. So based on everything you've heard, whether there's evidence, indirect evidence, or direct evidence, is the rationale sufficient to support this survey? That's what we're voting yes, or no.

(Off microphone comment.)

CO-CHAIR BROOKEY: Can you use your mic?

MEMBER MOYER: I think Amy and I are struggling with the same thing. That, if the

question wants to know, if there's a relationship between the patient-reported outcome, which is, did your doctor do this, and the process, which is, did your doctor do this? Well, there's the relationship between those two things, theoretically, is one to one, either your doctor did it, or your doctor didn't. Now the patient may not report that it was done. We've all had the experience of having people report that you didn't do something. Your spouse is pretty sure that you didn't do something that you're pretty sure you did do. But, so I mean, we all know that that can happen, but the health outcome that it's, the patient-reported outcome is whether counseling occurred and the process is counseling, so there's a one-to-one relationship between those two things, there's no need for a rationale.

MEMBER HOUTROW: There's no need for evidence.

MEMBER MOYER: Or evidence.

CO-CHAIR BROOKEY: Right. So that's

the second box, we're just, we're on the first box.

MEMBER HOUTROW: We're on the second box, though.

CO-CHAIR BROOKEY: But I'm on the first box.

(Laughter.)

CO-CHAIR BROOKEY: So we have to get out of the first box to get to the second box.

MEMBER MOYER: But you said that this was a patient, I mean, and Helen confirmed that this --

CO-CHAIR BROOKEY: Yes.

MEMBER MOYER: -- we've answered the first box, there's no question about that.

MEMBER HOUTROW: So the first question is yes, we accepted that this experience is a patient-reported outcome.

CO-CHAIR BROOKEY: But, do we need to vote on the first box, because that's, that's what's on the screen right now? Why don't we vote on the first box, because we -- all right,

so we're going to assume it's a PRO, so we're going to go to the second box now. So now we're voting on, and again, the words are not on the screen, right?

So the second box says, does the Committee agree that the relationship between the patient-reported outcome survey and, at least, one health care action structure process, intervention, or service, is identified and supported by the stated rationale?

CO-CHAIR SUSMAN: Isn't part of this that the denominator population values this patient-reported outcome, isn't that, really, the question we're trying to answer here?

And I think there's some good evidence, or at least, a suggestion that they do value this information, as they've reported. And that, to me, clarifies the question a little bit that we're trying to vote on, at this point. Or not.

CO-CHAIR BROOKEY: Go ahead, Jeff.

MEMBER SCHIFF: I, while I agree with

Jenny and Amy, I think the issue here is about how the data's collected. Because when we looked at the FECC, we were talking about the novel way in which we were collecting information on these processes of care.

And I think that, to me, I'm not sure if it's on the screen correctly, but I think what we're really saying is, to me, it's, does the novel way of asking the adolescents, in that time frame, about whether or not they got that care, warrants a pass/fail on a patient-reported outcome, because of the way the data's being collected.

You could get the same data documented in the record, but that's not what we're actually measuring. What we're trying to measure, I think, is whether or not the patient felt they received it.

CO-CHAIR BROOKEY: Right. So going back to the question. It's a survey. We know it's a survey. We know what it's measuring. And does it, is it related to, at least, one health

care action, as stated in the rationale? So medication management would be just one aspect of it.

CO-CHAIR SUSMAN: Or alternatively, is it related to outcomes, or processes that the adolescents, themselves, value? I mean, is that experience of import to them?

CO-CHAIR BROOKEY: Anymore clarifying questions for the developers?

(No audible response.)

CO-CHAIR BROOKEY: Any objection to voting?

(Pause.)

CO-CHAIR BROOKEY: You're objecting, Ricardo?

MEMBER QUINONEZ: Yes.

CO-CHAIR BROOKEY: Karen.

MEMBER DORSEY: So I don't, I'm feeling like we're complicating things, a little bit, because it seems to me that what the second box is asking is, can the, sort of, intended target in the health care system have an impact

on the patient's experience?

Is there, can we imagine that, something the health care provider does, can impact the patient experience? That seems like a pretty low bar. That seems like a pretty clear yes. But it's a, to me, it's a separate question to say, is this a valuable measure?

I mean, is it contributing something valuable, such that we think it's appropriate for endorsement, and I don't think that's the question we're trying to answer with this first vote.

I don't know where that comes, downs the line, maybe it comes when we talk about validity, or use and usability, but it seems, to me, that this is a very straightforward question to say, do health care providers have some ability to impact this patient experience? And that seems like a pretty clear yes, to me.

And then, at some point, we're going to talk about, whether this is a valuable contribution to this particular area of medicine

of transitions.

CO-CHAIR BROOKEY: Ricardo.

MEMBER QUINONEZ: So the, the thing that makes me the most uncomfortable is that, we're supposed to be voting on the evidence, for this part of the -- and that algorithm ignores the evidence.

I would feel much more comfortable, and I, kind of, agree that it's a, I'm try, I've been trying to imagine ways in which you could answer yes, and the answer to number two be no, and it's very hard for me to find it.

And so I would be much more comfortable, if we voted on the green algorithm, but we separately voted on the strength of the evidence, you know, and whether that, you know what I mean? That, that would be, to me, that would, that would make a lot more sense, when you're assessing --

DR. BURSTIN: I have plenty more examples of --

MEMBER QUINONEZ: -- quality of

evidence.

DR. BURSTIN: -- outpatient rationale, by the way.

(Simultaneous speaking.)

DR. BURSTIN: There are numerous outcomes that go down in rationale, because there are no clearly processes, or anything related to them, where committees have been very uncomfortable putting an outcome forward, without even a rationale.

So it, you know, again, this is an issue we've had, lots of, you know, some members get uncomfortable about this. I think it's very much the sense of trying to not hold outcomes hostage, until you've got, so the processes instructors in place around them, is the logic of it. But I understand where you're coming from.

CO-CHAIR BROOKEY: So my recommendation is that we move it forward to a vote, and we have very clear comments that we're going to document about the concerns, not only about the process, but as well as, the fact that

this process does not allow us to, sort of, vote on the evidence, the strength of the evidence, or even to have an option for insufficient evidence with an exception.

Because I don't think that we're all going to all feel comfortable with this, in any other way. Does that make sense? I mean, it's, it's, I think we need to move on to the next aspects of this measure, which, I think, will be, probably, an easier conversation. So any objection to moving forward with voting?

(No audible response.)

CO-CHAIR BROOKEY: Okay, let's go ahead and vote.

MS. CHAVEZ: Okay, we're voting on Measure 2789, Adolescent Assessment of Preparation for a Transition to Adult-focused Health Care.

Does the rationale support the relationship of the PRO, to, at least, one health care structure, process, intervention, or service, yes, or no? One for yes, two for no.

Voting's open.

(Pause.)

MS. CHAVEZ: We are expecting 24 votes. There is one recusal. Seventeen, 18, 22, 24. Twenty-two voted yes, two voted no. The measure passes evidence.

CO-CHAIR BROOKEY: So let's move on to performance gap, and I'm going to ask Marlene, Amy, or Kerri, to comment.

MEMBER MILLER: This is Marlene. I heard everything. I don't supervise PRO, but I, I thought there were certain minimal data on a performance given compared to adolescents who, you know, comparing, you know, their transition score, you know, how he did on it versus teens that weren't well-prepared.

I couldn't really understand if the survey, actually, therefore, detects transition readiness. You want some control, not to say that kids that aren't well-prepared, this is how they scored, versus kids that are well-prepared, this is how they scored, but I didn't see

anything like that, so I have, I really couldn't comment, as to whether there is a gap, in a sense that this measure can actually address it.

CO-CHAIR BROOKEY: Amy, or Kerri.

MEMBER FEI: I don't know that there's anything, actually, from the survey that would help us with gap. If you guys want to chime in on what you found? There was a little provided in the documentation, but I'm not sure that it demonstrates, numerically, that there's a huge gap.

DR. SAWICKI: Well, the goal of our field testing was to quantify, at a health plan level, or at a hospital level, the performance on these three scores. It was not, we did not use multiple, or other, instruments to then differentiate whether youth were prepared, or had readiness on other measures.

There's, certainly, other measures that can evaluate transition readiness, in other ways, but the data that we got, uniformly, across the three sites show that there was a gap in

performance, meaning that the scores were quite low in all three, particularly, for the transition planning domain where less than ten percent, on average, of any of the youth stated that they even had a discussion about transferring care.

In terms of the, the highest scores were in the medication, prescription medication domain, about 60 percent on average. And in the middle, with on the transition self-management. And we looked at it, the individual item, at the composite level.

And, you know, from a perspective of someone who thinks about adolescent medicine, the one question in the transition self-management domain is, did your provider meet with you, without your parent in the room, and only 30 percent, across the board, said yes.

So I think all, if you're looking at the of a, is there a performance gap? We didn't look at differences, based on different populations, in that way, but we identified that

there is a gap, at a population level.

CO-CHAIR BROOKEY: Other questions, comments, about performance gap? Ricardo.

(Off microphone comment.)

CO-CHAIR BROOKEY: Okay. Oh, you got your voter? Okay, you're ready to vote? All right, we'll vote.

MS. CHAVEZ: Okay. We're now ready to vote on gap for Measure 2789. Options are one for high, two moderate, three low, four insufficient, and voting's open.

(Pause.)

MS. CHAVEZ: Eighteen. Twenty-two. Twenty-three, 24. Two voted high, 16 voted moderate, five voted low, one voted insufficient. This measure passes gap.

CO-CHAIR BROOKEY: So let's move on to reliability. And I'm not sure who's going to tee up the conversation about reliability.

MEMBER MILLER: This is Marlene. I'll just, I'll say that on my sense of reliability, I didn't really so many kinds of things, like

repeated testing with same population, the testing that was done was three, sort of, geographically disbursed areas.

But, I don't have, I couldn't see any information on how similar these populations, test populations, in these areas were, because they were, literally, just spread across the country, so I, I didn't think their, their reliability was very hard to determine.

CO-CHAIR BROOKEY: Developers want to comment?

DR. SAWICKI: So I think there are several ways to think about reliability for a survey measure. One that we did do was look at the interim reliability when developing the composites and that did show quite strong ordinal alphas, from a perspective of, do the questions hold together, as a construct?

And so that's one measure of reliability for a survey instrument that I think was done, in terms of the populations, themselves, two were for Medicaid health plans,

they were geographically disbursed.

So they were, somewhat, different, in terms of race and ethnicity, but we were limited in the amount of data that we had from claims to, to do much more, in terms of looking at very many differences.

We didn't have the numbers, really, to be powered to look at race, ethnicity differences, for instance, as an example. But from a survey development perspective, our contingent is, that the reliability is there from a, from a statistical perspective.

MS. MUNTHALI: Sorry, we were just discussing the number of items that might be in this measure, so perhaps you can elaborate, a little more, on that? We just want to make sure that, in terms of process, we're following the same process, as we did with the fact, which had multiple items in a measure.

DR. SAWICKI: So the survey, itself, consists of 26 questions, not all of them are used in development of the score. And then, for

each of the three domain scores, there are either four or five items that fall within each of the domain scores.

MEMBER JOHNSON: So do you consider each of the domains a separate performance measure?

DR. SAWICKI: So they each encompass a different aspect of care around transition counseling. Each individual item I do not consider, as a separate measure, as many of them are intimately linked and our factor analysis that we did, also show that there was a good link between the composite questions.

They do have three different constructs that they capture, when it comes to thinking about, sort of, validity and face validity of what these different constructs are. They are, somewhat, separate, albeit, potentially, related, in a global sense.

CO-CHAIR SUSMAN: So when you finally end up scoring this, is it scored as one composite score for all three domains?

DR. SAWICKI: It's one, one score for each domain, separately.

CO-CHAIR SUSMAN: Okay, so --

DR. SAWICKI: But the survey --

CO-CHAIR SUSMAN: -- there will be three scores?

DR. SAWICKI: -- was administered, as a -- that would be three scores.

CO-CHAIR SUSMAN: Okay. Thank you.

MEMBER JOHNSON: So if it's three scores that's what we would consider three separate performance measures. So like the FECC that had ten different things in it, we're seeing this one, as having three different things in it.

CO-CHAIR SUSMAN: Do you have the information on reliability and validity, for each of those domains, then? I know you have --

DR. SAWICKI: For the first two domains we do have, well, we have reliability data on all three domains, in terms of the ordinal alpha, in terms of the factor analysis, we were able to do a factor analysis on the first

domains, the third domain of transition planning, the numbers were insufficient, by nature of the responses.

So the actual question that the domain refers to is, did you and this provider talk about whether you may need to change to a new provider, who treats, mostly, adults, and if someone answered no to that, they skipped out of the remainder of the questions, because if they didn't have that first conversation, they can't comment on anything else. And so that was, by nature, a score of zero.

And so less than 15 percent in each of the field test samples answered yes to that question, so the numbers were insufficient to do true validity testing for that particular domain.

CO-CHAIR BROOKEY: In terms, going backwards to performance gaps, would there be any difference between the three domains?

DR. SAWICKI: I think the gap is present for all three.

CO-CHAIR BROOKEY: So I'm asking the

Committee, whether we need to go back and re-vote, or can we just have one vote for all three domains?

CO-CHAIR SUSMAN: I'm okay with the one.

CO-CHAIR BROOKEY: Any objection to leaving the vote, as is, for all three?

(No audible response.)

CO-CHAIR BROOKEY: Okay. It sounds like, for the reliability, and I'm going to have Carol make a comment, reliability, we may need to vote for three different domains. So, Carol, comment.

MEMBER STANLEY: Yes, getting back to the question about age and the scoring that you're talking about, as a composite. So am I understanding correctly that, if one of the respondents says that they were counseled about prescription use, and also responded that they haven't made their own doctor's office visit, scheduled it themselves, in the past year, won't that reflect poorly in the scoring?

DR. SAWICKI: The questions you cited were, there's a scheduling of appointments, which is not in the same domain, as the prescription medicines, so in terms of the ones around scheduling appointments, the way that we handle it, in terms of the score, was that if someone said that they had scheduled their appointments, they got credit, full credit for that part of the score, those two questions.

So the first question is, did you and your provider talk about you scheduling your own appointments, instead of your parents? If you said no to that, but said that in the last 12 months you did schedule your appointments, you still got credit, as a, as a, you didn't get penalized for that.

MEMBER STANLEY: But what if they talked about scheduling their own appointment, but because of --

DR. SAWICKI: But they never did it, correct.

MEMBER STANLEY: Because they're 16

years old, they --

DR. SAWICKI: So they still got credit for that, because the conversation was had and the counseling was recorded.

MEMBER STANLEY: Oh okay. Okay. All right, got it. Thanks.

CO-CHAIR BROOKEY: So just to clarify, the three domains are on the screen. The first of which is counseling on transition self-management, number two is counseling on prescription medication, and number three is transition planning. Number three is the one that didn't have sufficient reliability testing?

DR. SAWICKI: It did not have sufficient validity testing, I would --it has --

CO-CHAIR BROOKEY: Oh yes, I know enough to --

DR. SAWICKI: It has face validity testing, for sure, in terms of our focus groups and cognitive interviews, as well as, sort of, expert consensus. But it was unable to go through the factor analysis, in terms of validity

testing.

CO-CHAIR BROOKEY: Okay.

DR. SAWICKI: In terms of reliability, there was enough numbers to create the ordinal alpha, coefficient for all three domains.

CO-CHAIR BROOKEY: For all three, okay. Any further questions from the Committee, before we vote? Go ahead.

MEMBER THACKERAY: You had touched, briefly, earlier, on the idea that this had to be the health care provider, the physician, or nurse practitioner, can you expand on that, a little bit, why it has to be that specific individual, and I guess, not having gone through the studies that are referenced, do the studies also support that it needs to be the physician, or is there recommendations, or evidence to support use of a social worker, use of a care coordinator?

DR. SAWICKI: The consensus guidelines, certainly, do state that care coordinators and other ancillary staff can be involved in transition planning.

As part of our validity testing, we did do focus groups with young adults, parents, and teenagers, and all of those situations, it was very clear that, despite what we may think, as health care practitioners, the youth really identify their doctor, as their provider, as their main point of contact to the health system, if they, in fact, had a doctor that they went to.

And so we really felt strongly that, from this perspective of measuring an experience, we wanted to make sure that they were anchoring it to their provider.

DR. SCHUSTER: Yes, I'm going to just add to that. That, the survey, in no way, suggests that you wouldn't, in a practice, have a social worker, case manager, or someone else help with a lot of this.

But the idea is, this is a major transition, one of the scariest parts of these kids' lives, and they want and the profession is saying that it's the primary provider, who has to introduce the topic, or at least discuss it.

That the message has to come from the provider and not just others, but others can still play a role.

CO-CHAIR BROOKEY: Jeff.

CO-CHAIR SUSMAN: Vote.

CO-CHAIR BROOKEY: Jeff wants to vote. All right?

(Laughter.)

CO-CHAIR BROOKEY: Okay. So we're voting on three separate, it sounds like the reliability testing would be sufficient for all three to be done en bloc, are we okay with that? So let's vote for all three in one vote. I mean --

MS. CHAVEZ: So one vote for all three domains?

(Simultaneous speaking.)

CO-CHAIR BROOKEY: Are we okay with that? Okay.

MS. CHAVEZ: Okay? So we're voting on reliability for all three domains for the ADAPT survey measure. One high, two moderate, three

low, four insufficient. Voting's open.

(Pause.)

MS. CHAVEZ: Twenty-four.

CO-CHAIR BROOKEY: Moderate, yes.

MS. CHAVEZ: Okay.

CO-CHAIR BROOKEY: Okay.

MS. CHAVEZ: Zero voted high, 19 voted moderate, five voted low, zero voted for insufficient.

CO-CHAIR BROOKEY: Great. So we talked, a little bit, about validity already, do our members, expert members want to comment on validity? And just to remind you, that would be Marlene, Amy, or Kerri. If not, do the developers want to comment?

DR. SAWICKI: I think that, in terms of validity, if you're going to think about them, as three separate domains, the only difference being that the third domain didn't undergo the confirmatory factor analysis and other types of voting, either the focus groups, the cognitive interviews, the expert interviews, all the things

that, sort of, went into face validity is there for all three.

CO-CHAIR BROOKEY: Okay.

DR. SAWICKI: The third, the third domain, it's about specific transition plan and was a transition plan discussed and given? And we know that that's a strong recommendation from many groups, professional groups and organizations, and we understand that there's a similar to a previous discussions today around care plans. We know that there is some rationale for thinking about care plans. And in the end, for an adolescent who is transitioning to adult care, having a care plan is important for them and may not be, as important, for their parents, particularly, in certain situations, and so even though we're unable to do the, sort of, construct validity testing, because of the numbers, that to us, actually, indicates that it's even more important to have such a measure in place.

CO-CHAIR BROOKEY: So I would recommend that we vote -- Amy, do you have a

comment?

MEMBER HOUTROW: I have a comment about validity. There were a number of populations that were excluded from this, but one that, for me that brings up a question about validity for a very important population, again, is those individuals who can't participate adequately, who really do need transition services, but can't participate in this survey, and so those would be individuals, who are, can't communicate, for whatever reason, or are otherwise intellectually impaired.

DR. SAWICKI: We 100 percent agree with that comment and we recognize that there is a need of quality measurement around that population and that population may, in fact, be the most vulnerable. But, when we started this process, we realized that there really isn't one way to capture patient experience around all adolescents and young adults with chronic conditions and we would be remised to ignore those that are cognitively able to transition,

they're a very important part of the population, understanding that future work needs to focus on those that have other developmental and intellectual disabilities.

MEMBER HOUTROW: Do you see that, as a threat to validity, at all? That you keep, are missing -- they're not an exclusionary population.

DR. SAWICKI: I think that, well, the exclusionary population's inability to complete the survey. And so it is, I don't think it's a threat to validity at a, at a population level, understanding what the population that you're serving, is.

MEMBER HOUTROW: They're not excluded.

DR. SAWICKI: They are excluded, meaning, if they get the survey and they cannot complete it, they are, the parents are instructed to not complete the survey.

CO-CHAIR BROOKEY: Yes that was my, I wanted to clarify that. It's being mailed to the parents, but the parents are being instructed not

to complete it, although, you can't really prevent them from completing it?

DR. SAWICKI: In our field test, we did have several families mail back the survey with their comments saying, I'm completing this form for my child, because they cannot, and we excluded those from out analyses.

There, also, is a question, at the end, which is very similar to other patient experience surveys asking, if the individual receives any help, so that kind of ability for stratification could be done.

CO-CHAIR BROOKEY: That, I think that helps. What I recommend -- oh. Jenny, go ahead.

MEMBER MOYER: The FECC question. So the denominator here was kids, who were pediatric care?

DR. SAWICKI: So the denominator is slightly different for the three domains, but the denominator for who was actually sent the survey, are children, who are 16 and 17 years old. And so the assumption is that they're in pediatric

care.

MEMBER MOYER: Okay, because 40 percent of kids that age see family practitioners --

DR. SAWICKI: So, so it --

MEMBER MOYER: -- not pediatricians.

DR. SAWICKI: So they could be in family practice care, or other types of general practices. It was, the way the sample is constructed is by using health plan data to identify children with medical complexity.

An organization could choose to field surveys in different ways, but it's not restricted to just the pediatric, or pediatrician.

CO-CHAIR BROOKEY: Ricardo.

MEMBER QUINONEZ: Just some guidance. So since validity does take into account some evidence, does the assessment of the evidence have to be different, because of the pathway we went through when we assessed the initial evidence?

(No audible response.)

MEMBER QUINONEZ: I mean, is it more limited, is the bar higher, how do we --

MEMBER JOHNSON: I don't think it's any higher.

CO-CHAIR SUSMAN: Yes, I don't see that there's any difference in the way we've considered validity in the past. You know, basically, is this measure valid?

MEMBER QUINONEZ: Right, but one of the things that I'd asked for measure specifications are consistent with the evidence, so that's why I'm asking about, whether our assessment --

CO-CHAIR SUSMAN: Yes.

MEMBER QUINONEZ: -- of the evidence should be any different?

CO-CHAIR SUSMAN: I would say no.

CO-CHAIR BROOKEY: I think --

MEMBER QUINONEZ: Like --

CO-CHAIR BROOKEY: I think you're raising a process question that we need to take

back to NQF. I think we've already, kind of, gone through this enough times, I think we need to, sort of, voice our concerns about this, but I, we're taking this pathway now and I think we vote on the validity, as been reported.

DR. SAWICKI: One more clarification for Dr. Moyer. In the third domain of the transition planning, there's a specific question that says, does your provider take care of mostly children and teens, and if they say no, because it's a family practice, they don't get a score for that. So even though transition planning needs to happen, from a developmental perspective, the actual transfer is not expected in that case, and we built that into that survey, specifically.

CO-CHAIR BROOKEY: Thank you. I'm going to recommend that we vote on one and two, together, since they're similar, in terms of validity, and then, 3 separately, if that's okay? So it will be two different votes. We're now voting for one of the first two domains. And

just to remind you, I've lost it on my screen here.

MS. CHAVEZ: Okay.

MS. ALLEN: Before --

MS. CHAVEZ: Now --

MS. ALLEN: One second. Before we start voting on validity, I just wanted to clarify something, for the record, a vote was miscalculated. So for reliability, the votes are actually zero high, 18 moderate, five low, and one insufficient.

CO-CHAIR BROOKEY: Okay, thank you.

MS. CHAVEZ: So we're voting on validity for the first two domains, counseling on transition self-management and counseling on prescription medication. And this is for validity and our options are one high, two moderate, three low, four insufficient, voting is open.

(Pause.)

MS. CHAVEZ: Twenty. Twenty-two. Twenty-two. Twenty-three.

CO-CHAIR BROOKEY: One more.

MS. ALLEN: Lauren, we're waiting on your vote.

(Pause.)

MS. CHAVEZ: Okay. So one voted high, 20 voted moderate, two voted low. The first two domains pass validity.

CO-CHAIR BROOKEY: Thank you. All right, the next vote will be on feasibility and this is -- I'm sorry. I'm sorry, the third domain. I got ahead of myself. So third domain.

MS. CHAVEZ: Okay, we're ready to vote on the third domain, transfer planning on validity, one high, two moderate, three low, four insufficient.

(Pause.)

MS. CHAVEZ: Sixteen. Twenty-two. Twenty-three. Twenty-four. Okay, zero voted high, 15 moderate, six low, three insufficient, and this domain passes validity.

CO-CHAIR BROOKEY: Okay, we'll move on to feasibility, and I think we've talked a little

bit about the logistics, about the way the survey is mailed out. We did talk about the way the denominator would be populated, in terms of getting health plan data.

This could, either, be a group, or practice measure, or a health plan measure, I believe, so are there other comments, or questions, about feasibility, from the expert group?

(No audible response.)

CO-CHAIR BROOKEY: Any concerns about feasibility? You wanting to vote? Yes, Jenny.

MEMBER MOYER: Just the survey response rate?

DR. SAWICKI: So in the three different field tests in our pediatric hospital, we had about a 45 percent response rate. And then, in our Medicaid plans, it was 22 percent and 28 percent.

MEMBER MOYER: Thank you.

CO-CHAIR BROOKEY: Which is very good. So --

MEMBER FEI: I guess, my question is, was there thought given to other methods, besides mail?

DR. SAWICKI: So that came up on the phone call, as well.

MEMBER FEI: Yes.

DR. SAWICKI: And that, we all know that, for adolescents, coming up with electronic ways of capturing data is, probably, a good way for the future, and I think that moving forward, thinking about capturing patient experience measures, electronically, at the point of care, in some other way that, particularly, for this population of adolescents, I think it's appropriate. We didn't do it in our field tests, so we can't comment on how it --

MEMBER FEI: Right.

DR. SAWICKI: -- would have been different, but certainly, it's something to be considered.

DR. SCHUSTER: And I'll just add to that, that we're doing other patient experience

work and we are, currently, testing another survey using email, which feels like it would have been a decade ago, but it's still a very new idea in the patient and family experience fields.

MEMBER FEI: It's moving in the right -- hey, it's all in the right direction.

CO-CHAIR BROOKEY: Are we ready to vote? Okay. And are we going to vote, I don't see any reason not to vote all three at once, is that all right?

MS. CHAVEZ: So we're voting on feasibility on all three domains?

CO-CHAIR BROOKEY: Yes.

MS. CHAVEZ: Okay. All right, one high, two moderate, three low, four insufficient, and voting is open.

(Pause.)

MS. CHAVEZ: Twenty. Twenty-two. Twenty-four. One voted high, 19 voted moderate, four voted low, and zero for insufficient.

CO-CHAIR BROOKEY: What's the use in feasibility? All right, we're at the top of the

hour, and so I need to ask your permission, if we can go a few more minutes. We have two more votes. Can we go a few more minutes?

Okay. So we've got to use --

(Laughter.)

CO-CHAIR BROOKEY: Otherwise, you guys have to come back tomorrow, right? So --

(Laughter.)

CO-CHAIR BROOKEY: Okay, we're going to usability and use. And are there, from the Committee Members, are there any comments, or questions about use, usability and use?

MEMBER FEI: I don't, I don't remember, from our call, but I know it hasn't been used outside of your testing, is that correct?

(No audible response.)

MEMBER FEI: Do you have a plan to use it more widely? I know you did, you got pretty wide results, so I thought that was good, but beyond those settings, is there thoughts of using --

DR. SAWICKI: So --

MEMBER FEI: -- small settings?

(Simultaneous speaking.)

DR. SAWICKI: Since this was developed, as part of the AHRQ CMS, PQMP, it became publically available. We've had over 80 groups inquire about use of the tool, you know, this happened within the last six months, so we, right now, don't have any way to know who is, or is not, implementing, or using it, at this point.

CO-CHAIR BROOKEY: Comments, questions? You want to vote? Okay, let's vote. And, again --

MS. CHAVEZ: Okay, we're voting --

CO-CHAIR BROOKEY: Again, we'll go ahead and vote for all three, simultaneously, unless there's any objections?

(No audible response.)

CO-CHAIR BROOKEY: Okay.

MS. CHAVEZ: Okay. And we're looking for 23 votes. One high, two moderate, three low, four insufficient, voting is open.

(Pause.)

MS. CHAVEZ: Thirteen, 18, 21. Two more. Twenty-three, thank you.

CO-CHAIR BROOKEY: All right, very good.

MS. CHAVEZ: Three voted high, 16 voted moderate, two voted low, two for insufficient, and those are the votes for all three domains on usability and use.

CO-CHAIR BROOKEY: Okay, so the final vote is on overall. And, I'll say it again that, we've had a lot of discussion, I think, it's been captured and will be taken back.

And so lots and lots of good comments and suggestions. I appreciate all of you being so keen on the words, because it really is, it's challenging, because all of us, this isn't our full-time job, but I think it's been a very fruitful discussion, and we should move on to this last vote of the evening for overall.

MEMBER MOYER: So this would be, if we vote positively, then this would become a quality

measure, if it goes, if it gets through everything else it has to get through?

CO-CHAIR BROOKEY: If it gets through everything else.

MEMBER MOYER: Yes.

CO-CHAIR BROOKEY: If it gets through everything else, it would be endorsed, but I can tell you, there's many endorsed measures that never really get put out there. So it, this is the first place to go --

MEMBER MOYER: Certainly.

CO-CHAIR BROOKEY: -- to get put out into the public. But, yes, there are other hoops it has to go through, before it becomes endorsed, and then, before it actually becomes --

MEMBER MOYER: Yes.

CO-CHAIR BROOKEY: -- in use.

MEMBER MOYER: And I, I want to point out that, as much as I do, I actually like the survey and I like the idea and I think transitions are incredibly important, having spent a lot of times in special needs clinic, I'm

a little uncomfortable that, that we're endorsing a very specific survey, without having looked at any other surveys that do the, that are aimed in this same direction.

DR. SAWICKI: I don't know that there really are any other experience --

MEMBER MOYER: I --

DR. SAWICKI: -- surveys.

MEMBER MOYER: I know of several others, they just haven't ever been presented, nationally. So and I, which isn't to say they've tried to, tried to get them out there, but we had one that we used at Texas Children's that was, you know, and I didn't particularly like it. That doesn't mean I like it any more, or less, but I have a concern about, about endorsing a very specific survey, rather than endorsing that, so --

DR. SAWICKI: We, certainly, looked at a lot of other measures that had been looked through, a lot of them were parent report measures and not youth report measures and we

adapted a lot of the questions in our, sort of, development of this and so, you know, particularly, the National Survey of Children with Special Health Care Needs was one. And so, so I mean, it's, I think there haven't been, I don't think there have been any others that have been directly adolescent reported.

MEMBER MOYER: Okay.

MS. MUNTHALI: So we just wanted to clarify something. NQF does not endorse surveys. What you are recommending for endorsement is a measure that's based on this specific survey. So we just wanted to clarify that, again, for the record.

CO-CHAIR BROOKEY: And voting for all three domains, at once, unless there's any objection? Oh, I'm sorry, go ahead.

PARTICIPANT: He was voting.

CO-CHAIR BROOKEY: Oh, he was voting?

PARTICIPANT: He was already, yes, he was eager to get out.

CO-CHAIR BROOKEY: Okay let's vote.

MS. CHAVEZ: Okay, is Measure 2789 suitable for endorsement? One yes, two no, voting's open.

(Pause.)

MS. CHAVEZ: Eighteen. Twenty. We're looking for 23 votes. Twenty-three, thank you. Sixteen voted yes, seven voted no.

CO-CHAIR BROOKEY: All right, so thank you very much. We're not done, yet, we have one more decision to make.

We have a lot of work to do tomorrow, and there's been a suggestion from the staff, not from me, but from the staff, that we get an early start. Would people be okay starting, we were supposed to start at 8:00 a.m., is that correct?

MS. MUNTHALI: At 7:30 a.m., so I don't know how much earlier we can start. But there is breakfast at 7:00 a.m., if I'm not mistaken.

MS. ALLEN: Breakfast is at 7:30 a.m.

MS. MUNTHALI: At 7:30 a.m.

MS. ALLEN: And the meeting starts at

8:00 a.m.

CO-CHAIR BROOKEY: Can we, can we start breakfast earlier and start the meeting at 7:30 a.m.?

MS. ALLEN: Yes.

CO-CHAIR BROOKEY: Is everybody okay with that? We're going to start the meeting --

(Off microphone comment.)

CO-CHAIR BROOKEY: Breakfast at 7:00 a.m., the meeting will start, promptly, at 7:30 a.m. And we will be out, promptly, by 3:00 p.m. All right, good night, everyone.

(Whereupon, the meeting in the above-entitled matter was concluded at 5:05 p.m.)

M

T

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Pediatric Measures
Steering Committee

Before: NQF

Date: 12-01-15

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R. Gross

Court Reporter