**NATIONAL QUALITY FORUM**

**+ + + + +**

**PEDIATRICS PERFORMANCE MEASURES COMMITTEE**

**+ + + + +**

**THURSDAY**
**MARCH 2, 2017**

**+ + + + +**

**The Pediatrics Performance Measures Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 8:00 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, 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**
**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, 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, John's Hopkins Children's Center at JHHS ***
**RAJIV MODAK, MD, El Rio Community Health Center**
**JILL MORROW-GORTON, MD, University of Massachusetts Medical School**
**RICARDO QUINONEZ, MD, FAAP, Baylor College of Medicine**
**JEFF SCHIFF, MD, MBA, Minnesota Department of Human Services**
**CAROL STANLEY, MS, CPHQ, Duke University**

**NQF STAFF:**

**SHANTANU AGRAWAL, MD, President and CEO**

**HELEN BURSTIN, MD, MPH, Chief Scientific Officer**

**KAREN JOHNSON, Senior Director**

**MADISON JUNG, Project Analyst**

**KATE MCQUESTON, Project Manager**

**ELISA MUNTHALI, MPH, Vice President, Quality Measurement**

**ROBYN NISHIMI, PhD, Consultant**

**SUZANNE THEBERGE, Senior Project Manager**

**MARCIA WILSON, PhD, MBA, Senior Vice President, Quality Measurement**

**ALSO PRESENT:**

**KIMBERLY ARTHUR, MPH, Seattle Children's Research Institute**
**RENEE ELLEN FOX, MD, FAAP, Center for Medicare and Medicaid Services**
**NARAA GOMBOJAV, PhD, The Child and Adolescent Health Measurement Initiative ***
**LAWRENCE KLEINMAN, MD, MPH, University Hospitals Rainbow Babies and Children's Hospital ***
**CHRISTOPHER LANDRIGAN, MD, MPH, Boston Children's Hospital, Harvard Medical School**
**SUZANNE LO, MPH, University Hospitals Rainbow Babies and Children's Hospital**
**RITA MANGIONE-SMITH, MD, MPH, Seattle Children's Research Institute**
**KAREN MATSUOKA, PhD, Center for Medicare and Medicaid Services ***
**SARAH REEVES, PhD, MPH, University of Michigan ***
**MARK SCHUSTER, MD, PhD, Boston Children's Hospital, Harvard Medical School**
**JEFFREY SILBER, MD, PhD, Children's Hospital of Philadelphia**
**MICHELLE SOLLOWAY, PhD, The Child and Adolescent Health Measurement Initiative ***

**DAVID STOCKWELL, MD, MBA, Children's National Medical Center**

**SARA TOOMEY, MD, MPH, MPhil, Msc, Boston Children's Hospital, Harvard Medical School**

*** present by teleconference**

CONTENTS

**CONTENTS**

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

8:02 a.m.

MS. MCQUESTON: Welcome to our one-day meeting for National Consensus Standards for Pediatric Performance Measures. We are delighted to have you all here, and thank you so much for coming with this early start this morning.

Just a very few brief announcements from me before we begin. I am Kate McQueston. I am the Project Manager supporting this project. Restrooms, you will go out through the conference area and turn right. They are down the hallway. We will have a few breaks today; unfortunately, not too many. We are keeping you hard at work. We will have one at 10:45 -- or 10:15. We will have lunch at noon, and then another break at 3:30.

Our wireless information, so that everyone has it, you will use the NQF wifi network with the username guest, and the password is NQFguest, NQFguest, NQF in caps.

So we will go ahead and begin the

meeting, and we will begin by passing it off to NQF's new CEO, Dr. Shantanu Agrawal. Thank you.

DR. AGRAWAL: Thank you very much. So this is week five, day four, so I still get to be called new, I think, for quite some time. We will have to figure out how long. We will actually be convening a consensus panel to figure out how long I can be called new, because that's how we do it.

(Laughter.)

DR. AGRAWAL: Thank you very much for being here today. This is going to be a really full day, I know, but I just wanted to underscore the importance of the work. I think you are reviewing 11 measures today. That is going to add to a body of about 100 measures that have already been endorsed in the pediatrics space by NQF. The second phase of work for this committee, so again, just really extremely important.

I want to thank John and Jeff for their leadership of this committee. I also

wanted to just reinforce how personally important this was to me. You know, yes, we are here to improve the health and healthcare of the pediatric population, but I should also tell you that my wife is a pediatric hospitalist. She has made it very clear we need really good measures in pediatrics.

(Laughter.)

DR. AGRAWAL: When -- when I got offered this job and had accepted it, literally, she said, you know, let's get to work. We need pediatric measures. So because I ascribe to the happy wife, happy life doctrine, please do a good job today.

(Laughter.)

DR. AGRAWAL: I need to go home and report some good news.

No, so thank you very much. This is extremely vital. It has been great in the last five weeks to sit in on these committees, really see how they operate. The amount of expertise that we have here is incredible, and I thank you

for donating as much time as you do. Our work would not be possible without the representation we have here and your willingness to participate in this process. Thank you.

CO-CHAIR BROOKEY: Good morning, everybody. I am John Brookey. I recognize most of you in the room. I feel like I was just here like -- I walked in the room, and I was like I have been here yesterday, but I wasn't. It has been over a year now since we were here.

So we are going to go ahead and -- are we going to do the introductions of the whole room later, or are we going to do it --

MS. MCQUESTON: We will combine that with our --

CO-CHAIR BROOKEY: A little later, okay.

MS. MCQUESTON: -- our disclosure after we do --

CO-CHAIR BROOKEY: So I will let Jeff say hello. Then we're going to go to our opening remarks from CMS.

CO-CHAIR SUSMAN: So again, welcome to everybody. It is great to see everybody back, and a few new faces. This is a really exciting time I think for measurement and quality, and particularly in the pediatric space, where we probably have a dearth of measures available, so this should be a great meeting.

Just a couple of words about process. We have a hard time seeing people at the ends, so it will really help us, if you have questions or things to comment on, to put your cards up like this so that we know you want to say something. And I hope this is a really fun meeting. Don't be worried about disagreeing with each other, but do it nicely. Thank you.

CO-CHAIR BROOKEY: Okay. We are going to move to -- I believe that Karen and Renee are on the line, and they are going to -- is that -- oh, Renee is here, okay. Hi Renee.

DR. MATSUOKA: Yes, and this is Karen. Hi, everyone.

I wanted to take just a few minutes

just to welcome everyone to this meeting and also just to really thank you for making the time to be with us. We know how important your time is and how busy you are, so the fact that you are willing to lend us some of your time today is really important, so thank you very much.

And apologies for not being able to be there with you in person. It is by no means any indicator of how -- how important the work is because in fact, the reason why CMS decided to fund a call for measures specifically for pediatric performance measurement is an indication of how important we really do think this work really is.

So we funded this measure call because it really does align and further our mission under CHIPRA to expand and update the portfolio of evidence-based pediatric measures available for public and private use. So the work that you will be doing today in terms of reviewing and potentially endorsing new measures that are going to become available for use not just in Medicaid

and CHIP, but to the commercial population as well, is so important.

And just to give you a sense for why it is so important to Medicaid and CHIP, you probably all know that Medicaid and CHIP serve over 70 million beneficiaries in this nation. That is more than what Medicare and the marketplace combined cover. And of that total enrollment, over 51 percent of our enrollees are children. That is -- almost 40 million beneficiaries are children, and over 40 percent of births in this country are also covered by Medicaid.

And of course, on the CHIP side, where almost all of our beneficiaries are children, we serve a further 8.4 million kids. So having a robust slate of measures that we can use to really assess the extent to which we are providing evidence-based care to our kids is so vitally important.

And so we thank you for being here today because NQF endorsement, as you know, helps

to really increase the visibility and the use of measures, and no matter how many measures we have, people don't know that they exist, and people don't know that they are reliable and credible, useable and feasible, and scientifically rigorously tested. No matter how many measures we might have, that doesn't do us any good if people don't actually use them.

So I want to just -- before turning it over to my colleague, Renee Fox, I want to again thank you for all your work, for lending your expertise to our work, and hopefully together, we can start to build a robust slate of measures that we can use to build a brighter and healthier future for our nation's children. So with that, I am going to turn it over to Renee.

DR. FOX: Thank you, Karen, and good morning, everyone.

As Karen said, this really started with the CHIPRA reauthorization, 2009, and before that there really were no standard measures. One of the reasons that we have these measures was

that we were unable to measure Medicaid performance across states. States maybe had their own measures, but they were not -- they were not reporting them.

And so now there is voluntary reporting on the Child Core Set measures consistent with consistent metrics across five domains: primary care access and preventative care; perinatal health, my specialty; care of acute and chronic conditions; behavioral healthcare; and dental and oral health services. And the Child Core Set, as of 2016, had 26 measures on it, and 50 states and the District reported on at least one measure, and the median number of measures reported by states was 16 for federal fiscal year 2015.

And so as Karen said, this is really a foundation. We don't just want to measure; we want to help improve the care for Medicaid and CHIP, and through this, we also look at quality improvement and CMS has roles in -- through funding and technical assistance to support

states in setting performance goals and implementing improvement projects, either through groups, affinity groups like the Maternal and Infant Health Improvement Initiative, some of the external quality review organizations' improvements.

And so to have effective system reform, we need to have data and performance metrics that include -- that enable health service delivery transformation, data and IT, payment reform, and so we are here in the -- Karen and I sit in the Division of Quality and Health Outcomes, and we have been referring to this as DQ 2.0. And so we have multi-level management measurement.

You are aware of the Medicaid CHIP Child and Adult Core Sets, but we will soon have forthcoming some first-ever nationwide Medicaid consumer assessment of child healthcare -- of healthcare providers and systems, the CAHPS. We are going to have a quality rating system in managed care, and we have provider-level

measures, including the Health Home Core Measures, Behavioral Health Clinics Core Measures, and as many of you know and probably have worked on, the Center for Clinical Standards and Quality AHIP Core Measures, which have released seven adult core sets, and the pediatric set should be forthcoming.

And there are many levels of performance measures that roll up from provider to plan to state to Medicaid. I also want to share my appreciation for the work this committee and the staff at NQF are doing, and really understand -- want to highlight that measures are appropriately designed for different levels of measurement, from practice to payer to plan and state levels, say that not all measures are appropriate for use at each level, but should be reviewed here in the committee for use at the level for which they were tested.

Your expertise will help ensure that these measures are important, scientifically rigorous, useable, feasible at the level they

were tested, and we are very mindful at HHS of the provider burden that results from the proliferation of measures in general. And the decision, just as Karen said, made by you at the committee level will help inform our MAP Medicaid Task Force, which will meet in May 2017 to provide input on how to strengthen the core set for children, and we will look at these measures in the MAP Task Force as discussed as part of this call when developing their recommendations for us, so your work is highly vital to what we do. It is vital to the care of children and measuring quality, and I want to appreciate -- thank you, everybody, for your hard work and volunteering.

CO-CHAIR BROOKEY: Thank you, Karen and Renee. I think we're turning it over to Ann for our disclosure of interest.

MS. HAMMERSMITH: Thank you. Hi, everyone. I am Ann Hammersmith. I am NQF's General Counsel. I will lead you through the oral disclosures of interest.

If you recall, when you were nominated for the committee, you nominated yourself for the committee, you received a conflict of interest form that was a few pages longer than we would like, but it is necessary, and we asked you a lot of information about your professional activities. So what we do at the -- at the first meeting of a committee is we go around the table and you share anything you would like to disclose, but only if it is relevant to the work that you're going to be doing.

One thing that I want -- want to try and reassure you about is, if you disclose something, it does not mean it's a conflict. Part of the reason to do this is for transparency so that everyone understands where you may be coming from. An example of that is if you sat on a committee for a professional society as a volunteer, and it is relevant to the work today, we would look for you to disclose that. It is not necessarily a conflict of interest.

So before we start, I just want to

remind you of a few things. You sit as individuals. You do not represent your employer. You do not represent anyone who may have nominated you for service on the committee.

One thing that is different about our conflict of interest and disclosure process is that we look at things other than financial interests. Many disclosure of interest processes only look at the financial angle. Because of the unique nature of the work that we do, we are looking for financial disclosures, but we're also looking for things that you may have done where no money was involved, but it is relevant to the subject matter today, and I will use the example I just used. If you served on a committee for a professional society as a volunteer and it is relevant to the work today, we would like you to disclose that.

The other things we are particularly interested in are speaking engagements, grants, research support. If you are on the phone, I will call on you at the end of the in-person

disclosures. Right now, we will start with the Co-Chairs -- I always start with the Co-Chairs -- and we will go around the room. Tell us who you are, who you are with and if you have anything to disclose.

Federal government representatives, we do not ask you for disclosures of interest, so you don't need to participate in this portion of the disclosures, so.

CO-CHAIR SUSMAN: How could I forget? I am Jeff Susman. I am the Dean Emeritus at NEOMED, Northeast Ohio Medical University, and I am fortunate to have the best job in the world, which is being on sabbatical now.

(Laughter.)

CO-CHAIR SUSMAN: And I have nothing to disclose in this sphere.

CO-CHAIR BROOKEY: Good morning. I am John Brookey. I am a pediatrician and a Quality Medical Director and a Medical Director for Medicaid in Kaiser Permanente in Southern California, and other than being a user of

measures, I have nothing to disclose.

MEMBER HARPSTER: I am Karen Harpster from Cincinnati Children's, and I have nothing to disclose.

MEMBER FATTORI: Good morning. I am Debbie Fattori. I am a pediatric nurse practitioner, and I serve at the DuPont Hospital for Children as the Senior Director of Ambulatory Nursing. I have no disclosures.

MEMBER FEI: Hi. I am Kerri Fei. I work for Blue Cross Blue Shield Association in Chicago, so we are the trade association for all 36 Blue Plans across the United States. I am an ER nurse by background, and prior to this, I did work at the American Medical Association doing physician-level performance measure development for -- with the Physician Consortium for Performance Improvement, and I have been gone from there for seven years. So I used to sit in that measure developer chair over there, and I am happy to be on this side now.

(Laughter.)

**MEMBER FEI: But I have been gone from there for seven years, and I have nothing to disclose.**

**MEMBER BERGREN: I am Martha Bergren. I am a Director of Advanced Population Health and Health Systems Leadership and Informatics at the University of Illinois Chicago, College of Nursing. I have been a pediatric nurse, both in the acute care setting and in the community setting, for longer than I would like to admit, and I have no disclosures.**

**MEMBER MORROW-GORTON: Jill Morrow, I am a Developmental Behavioral Pediatrician Associate Professor, Pediatrics and Community Medicine and Family -- whatever they call it, at the University of Massachusetts. It has this really long name. I am also a Senior Medical Director at MassHealth, which is the Massachusetts Medicaid program, and the clinical lead for the Long-Term Services and Supports Office, where we choose measures and use measures, but don't develop measures.**

**MEMBER DORSEY: I am Karen Dorsey. I am a pediatrician, and I am Associate Director of Quality Measurement at the Center for Outcomes Research and Evaluation at Yale, and we develop measures in the Medicare fee-for-service elderly population. And I actually do the measure evaluation work for our measures that are in public reporting, so I am also usually in that hot seat over there. Glad to be on this side of the microphone, and no disclosures.**

**MEMBER QUINONEZ: I am Ricardo Quinonez, pediatric hospitalist from Texas Children's Hospital, Baylor College of Medicine, and I have nothing to disclose.**

**MEMBER KELLER: David Keller, the -- I am a Professor of Pediatrics and Vice Chair of Clinical Affairs and Clinical Transformation at the University of Colorado School of Medicine and Children's Hospital Colorado, and if anyone has any ideas how I can get something to disclose, please tell me.**

**(Laughter.)**

MEMBER SCHIFF: I am Jeff Schiff. I am the Chief Medical Officer at the Minnesota Medicaid Program, and I have something to do with selection of measures for our -- for quality for both our managed care organizations and our ACOs. I do -- I have been a co-investigator with Dr. Mangione-Smith on her Center of Excellence for the -- and have not been directly involved in the continuity of primary care for children with medical complexity, but some of my researchers have pulled data and been involved in that.

MEMBER EDIGER: Hi. My name is Maureen Johnson Ediger. I am a parent committee of the Quality and Safety Committee of the Board for Children's Hospital Colorado, and it is relevant that I have four children, and that is really why I am here, is because two of them live with special needs, one physical and one mental health special needs, so -- and I have nothing to disclose other than I have four children.

(Laughter.)

MEMBER BRISTOL-ROUSE: I am Tara

Bristol-Rouse. I am also here as a parent partner. I am with Patient and Family Centered Care Partners, and also I am Director of Patient and Family Partnership for the State Perinatal Collaborative in North Carolina, and I have nothing to disclose.

MEMBER EINZIG: I am David Einzig. I am a child psychiatrist, and I'm also boarded in pediatrics, so I did come from a triple board program. Work at a large children's hospital in Minnesota, Children's Minnesota. Primarily a clinician and see lots of patients. My primary interests include collaborative care and co-located care models. I am also on the Behavioral Health Steering Committee.

MEMBER MODAK: My name is Rajiv Modak. I have spent the last 15 years working on the front lines at El Rio Community Health Center in Tucson, Arizona, and I have nothing to disclose.

MEMBER FINKELSTEIN: Hi everybody. I am Jon Finkelstein. I am at Boston Children's Hospital, and I'm Professor of Pediatrics at

Harvard Medical School. My -- my day job is as Interim Quality -- Chief Quality and Safety Officer at Boston Children's. I -- I disclose two things. One is that I serve on the Board of the Academic Pediatric Association as the Chair of the Healthcare Delivery Committee, and the APA is a voting member of NQF, though I am not here representing them.

The second is that I am a co-investigator for the Center of Excellence in Pediatric Quality Measurement at Boston Children's Hospital, which has developed one of the measures we are talking about today, 3136, the GAPPS measure. I was not directly involved in that development, but I am part of that Center, so I will recuse myself. Happy to leave the room or focus on my incredibly compelling email when you discuss that.

(Laughter.)

MEMBER STANLEY: Good morning. I am Carol Stanley. I am currently with Duke University Department of Pediatrics, have been

there roughly nine months now, and last year, when I was here, I was with Virginia Medicaid and had been there for about eight years. Also, I have worked with Medicare with one of the QIOs previously, and Anthem Blue Cross Blue Shield. I have no disclosures.

MEMBER KONEK: Good morning, I am Sue Konek, recently now into my second year of retirement from the Children's Hospital of Philadelphia as the Director of Clinical Nutrition. I am the dietician on the group, and within the last month, I have started working at Dayton Children's two days a week in my -- my goal to balance fun and meaningful work. I really have nothing to disclose.

MEMBER BOST: Morning. I am Jim Bost. Last year when I was here, I was Director of the Outcome Center at Children's Healthcare of Atlanta, but as of Monday, I am a Research Division Chief here at Children's National Medical Center in D.C. of the Biostats and Study Methodology Group, and I have nothing to

disclose.

MEMBER HOUTROW: Hello. My name is Amy Houtrow. I am a pediatric rehabilitation medicine physician in Pittsburgh. I have federal funding from multiple granting agencies unrelated to measurement development or assessment. I am the Chair though of the American Academy of Physical Medicine and Rehabilitation Performance and Quality Metrics Committee, and we will eventually be doing measure development and have not started that.

I am also the Chair of an IOM study committee on improving health outcomes, for which obviously quality has a role, but we will not be actively looking at anything in particular related to measure development. And otherwise, I have nothing else to disclose.

MS. HAMMERSMITH: Okay. Thank you. I am going to call on some people who are on the phone. Lauren Agoratus?

MEMBER AGORATUS: Yes, hi. You get two points for saying my name correctly.

MS. HAMMERSMITH: That never happens.

MEMBER AGORATUS: I am the Coordinator for Family Voices, and I am the parent of a medically complex child, and I also most recently have been participating on a UCLA project on medical complexity, and I checked with Suzanne, it is not a conflict, but I am also on the CAMHI well-planner for children with special healthcare needs. We are not developing measures. We are developing a parent guide. Thank you.

MS. HAMMERSMITH: Okay. Thank you. Marlene Miller?

(No response.)

MS. HAMMERSMITH: Is Marlene Miller on the line?

(No response.)

MS. HAMMERSMITH: Are there any other committee members on the line?

(No response.)

MS. HAMMERSMITH: Okay. Thank you for the disclosures.

Before I leave you, I just want to

remind you of a few things. If you are in the meeting and you think you may have a conflict, you think somebody else on the committee may have a conflict, we ask that you speak up in real time. We don't want to get several months down the road and have someone pop up and say, you know, I think I may have had a conflict on that measure. We would like to know now.

You are always welcome to speak up during the meeting. If you don't want to do that, you can go to your Co-Chairs. You can go directly to NQF staff, and we will figure out how to resolve it. Any questions before I leave?

MEMBER FINKELSTEIN: Just to be clear, if we recuse ourselves, would you like us to leave the room, or is it okay to --

MS. HAMMERSMITH: I don't think it is necessary --

MEMBER FINKELSTEIN: Okay --

MS. HAMMERSMITH: -- for you to --

MEMBER FINKELSTEIN: -- fine.

MS. HAMMERSMITH: -- leave the room.

You can look at your compelling email.

(Laughter.)

MS. HAMMERSMITH: Anybody else?

(No response.)

MS. HAMMERSMITH: Okay. Thank you. Have a good meeting.

CO-CHAIR BROOKEY: Thank you, Ann. Before I turn it over to Suzanne, I just want to say how much I appreciate the work of the staff. I mean, it is a monumental amount of work to put this meeting together and all these documents for us to review, so I really do appreciate it. You guys did a great job, and it's just an honor and privilege to be a part of this process. So I will turn it over to Suzanne now.

MS. THEBERGE: Thank you. I am Suzanne Theberge. I am the Senior Project Manager on the team, and do quick team introductions.

MS. MCQUESTON: I am Kate McQueston, Project Manager.

MS. JUNG: I am Madison Jung, Project

Analyst.

DR. NISHIMI: Robyn Nishimi, I am a consultant to the project serving in sort of the Senior Director role, and as well I was the first Chief Operating Officer for NQF.

MS. MUNTHALI: Elisa Munthali, Vice President from Quality Measurement. Welcome.

MS. THEBERGE: And over here, we have Marcia, who is our Senior Vice President.

Now I am going to turn it over to Kate for some slides.

MS. MCQUESTON: Great. Thank you, Suzanne.

So we will begin today with just a brief project overview. So as you know, this project will be evaluating measures related to pediatric populations that can be used for accountability and public reporting for all populations and in all settings of care. This is the second phase of this project.

And specifically, the project will be addressing topic areas including child- and

adolescent-focused clinical preventative services and follow-up to preventative services; child- and adolescent-focused services for management of acute conditions; and child- and adolescent-focused services for management of chronic conditions.

So as all of you know, the project builds on our previous work last year, where NQF evaluated 23 newly submitted measures and one previously reviewed measure against NQF's standard evaluation criteria. NQF has a large portfolio of measures that include the pediatric population. Currently, there are more than 100 NQF-endorsed measures that include the pediatric population. Some of the measures are specific to the pediatric population and only include that group, and other measures are all-patient.

Measures fall into areas that pertain to clinical and cross-cutting areas, including cardiovascular surgery, pulmonary care, cancer, perinatal care, health and well-being, and safety. Many of these measures in the NQF

portfolio are used in public or private accountability and quality improvement programs, but the work group have identified gaps that remain in areas of care coordination, screening and abuse and neglect, and injuries, trauma and mental health.

Next, this is a brief overview of the measures that were endorsed in last year's cycle that many of you will be very familiar with. This is -- we won't go through these measure by measure, but just want to have them here as a reference. Here they are, continued.

And then these are the measures that we will be considering for this cycle, the 2016 to 2017 measures under review. We have 11 measures that we will be discussing today.

And with that, I will pass it over to Suzanne for an overview of the evaluation process.

MS. THEBERGE: Okay. So I know most of you have been through this before, and so this will be familiar. Let's see if I can get the

slide clicker to work. Next slide. Thanks.

So I just wanted to speak briefly about the role of the standing committee and the process that we will be going through this morning. As has been mentioned, the committee members act as a proxy for NQF's membership, and so we expect varied perspectives and values and priorities from you all during this discussion, and so of course we ask you to respect each other's differences of opinion, and -- and keep the conversation collegial, as they -- they always are.

And we -- I also wanted to note, we do have a very full agenda in order to fit everything into a one-day meeting. We packed it quite tightly, so we appreciate your early arrival, and we do really appreciate your helping us keep on track and -- and staying within -- within our time for the agenda.

As you also may remember, once we get into the measure discussion, we will have our measure developer colleagues joining us at the

table. We have a couple of chairs for them, and we really appreciate their presence. We do have several that will be on the phone as well, and we -- we really appreciate their joining us here, as that -- that presence is invaluable.

They will be providing a very brief introduction to their measure or measure set, and then if you have questions, you can ask them to respond to any questions that you might have about anything pertaining to the measure.

We do ask that you -- we know that our committee members often like to offer some suggestions on how a measure could be improved, or ways it could be refined, but we do ask that you consider and vote on the measure that has been submitted to us. That is what we will be evaluating today.

So the -- these are our ground rules. I don't -- see if anybody has any questions. They are fairly straightforward.

(No response.)

MS. THEBERGE: And hearing none, I

will go over the major criteria and the process.

CO-CHAIR BROOKEY: Can you just back up one slide?

MS. THEBERGE: Sure.

CO-CHAIR BROOKEY: Could you talk about being in the room or out of the room during voting? Or are you going to get to that?

MS. THEBERGE: I could talk about that right now.

CO-CHAIR BROOKEY: Okay.

MS. THEBERGE: So we do ask that folks remain in the room, especially when we are getting to a vote. I mean, we understand that you may need to step out for a moment, but if you could try to be at the table and voting while we are voting, we would appreciate it, to ensure that we maintain quorum and that we also maintain a consistent number of committee members voting throughout a measure. If you do need to step away for longer than a couple minutes, please just let a staff member know.

And so now, wanted to speak quickly

about our process and the criteria. So the first criteria is importance to measure and report, and the -- as you know, we first will be discussing evidence, so we will look at the evidence base for the measure, and then we will ask you to vote on evidence. You should all have a clicker, and we will be testing those after I get through this slide.

So we ask you to vote. If the measure passes evidence, then we go on and discuss and vote on gap, and then if that -- if the measure passes gap, we go on and discuss the scientific acceptability, so first reliability, again discussion and vote, and then validity, discussion and vote, and both of those as well are must-pass. The committee must -- must recommend a measure before we move forward.

And then we -- we go to feasibility, discuss and vote, usability, discuss and vote, and those two are not must-pass criteria, so we do continue on after we discuss those, and then we have a final vote on overall recommendation

for endorsement. At that point, if a measure has been recommended and there is a competing or related measure, we would have that discussion and -- and go from there.

So I think we will pull up our practice vote and have you all just do a quick practice vote and see -- make sure everybody's clicker is working so if it is not we have a little bit of time to replace it before we get into the actual voting.

Okay. So the practice question is which of the following philosophical schools was most identified with the Greek philosopher Aristotle? Option one, stoicism; option two, agnosticism; option three, platonism; and option four, empiricism. Voting is now open. You can attribute that to our Project Manager, Kate McQueston.

(Laughter.)

MS. MCQUESTON: It's okay to guess.

(Laughter.)

MS. THEBERGE: So yes, so we -- point

at Madison. She has got the voting slides. And just so folks know, project team will be voting for the people on the phone. They will be sending us their votes via chat, and we will be entering those. So I think we are still waiting for a few more votes.

(Pause.)

MS. THEBERGE: All right. Voting is now closed. Okay. The votes are in. We have 17 percent for stoicism, with four votes; 4 percent for agnosticism, with one vote; 43 percent for platonism, with ten votes; and 35 percent for empiricism, with eight votes. And Kate, what is the answer?

MS. MCQUESTON: The correct answer is 4, empiricism.

(Laughter.)

MS. THEBERGE: But yes, the bottom line is all the votes got in. Thank you.

(Laughter.)

MS. THEBERGE: All right. Thank you, everyone.

So I wanted to just take a moment and look at our algorithms. You do -- next slide.

You do have these printed at your seat, and I realize this is unreadable for those of you sitting at the table, but you do have a printed copy. And as you know, we just -- NQF has created these algorithms to help you rate the measures according to our criteria, so we ask you to look and walk through each of these.

We have one -- one for evidence, and I want to just talk for a moment about exceptions to the evidence. We have outcome measures, which we are not necessarily looking for the quality, consistency, and -- and -- the QQC, quality, consistency, and I am forgetting the other one. What we're looking for is that there is a rationale to support the relationship of the outcome to processes of care or the importance of measuring the outcome.

The other thing that we wanted to just flag is that expert opinion is not considered empirical evidence, but there are exceptional

circumstances in which expert opinion could be considered. In that case, you would vote for an exception to the evidence, and -- and that would be in cases when there is really no evidence available, and there has been a systematic assessment of expert opinion, and there is a really strong rationale for why that specific structural or process measure needs to have a measure associated with it.

So next slide is our reliability algorithm, which again, pretty straightforward, and then we have our validity algorithm, and you should have those at your seat as well.

For those of you in the room, up on the other screens, we do have a quick overview of our -- our threshold on voting, and I wanted to speak briefly about that. Measures must have greater than 60 percent to be considered recommended within our consensus guidelines, and measures that achieve between 40 and 60 percent are considered consensus not reached. We also sometimes refer to that as the gray zone, and we

do continue to review the measure if a measure is in that consensus not reached zone of, say, 50/50.

We keep moving forward up to the final vote, and there has been a change since the last time that you all met. If you don't achieve consensus on a must-pass criterion, then you don't make a final vote on overall recommendation for endorsement. The committee would stop at usability, and we would ask the developer to bring back further information to address the issues you had raised, and then we will ask you to vote again and make your final recommendation on the post-comment call in May. So that is a minor change to our process.

And then anything that does not achieve at least 40 percent is -- is considered not recommended, or does not pass.

And so one last slide from me. Just want to -- we do have appropriate-use measures, so I wanted to just speak briefly about that. We had a request for some more information about

that.

So appropriate use is a type of process measure that is used to evaluate procedures and medical technologies. They are not cost and resource use measures. They are not efficiency measures. And we do have some specific criteria for them, and the evidence should include a systematic assessment and grading of the quality, quantity and consistency of the body of evidence that the measure process does not lead to a desired health outcome, and -- and we ask you to just make sure that there is a -- evidence on the effectiveness or lack of effectiveness or benefit of the test or procedure to patients.

So, any questions?

CO-CHAIR BROOKEY: So just one process question: if we do kind of reach an impasse, are we going to like say at some point we're going to bring it back to the follow-up phone call, or what are we going to do if we get into that situation? Because we do want to be able to

finish the majority of these by the end of the day.

MS. THEBERGE: Sure. So if you are having trouble achieving consensus, eventually, I think if you -- when we got to a point where everything -- you know, people had had a chance to say their perspectives, then we would ask for a vote. You would call a vote, and we would either keep going forward, in the case of consensus not reached, so one of those 50/50 votes, we would have you just move on to the next criteria --

CO-CHAIR BROOKEY: Yes.

MS. THEBERGE: -- or, if the measure did not pass, then it would stop.

CO-CHAIR BROOKEY: Okay.

MS. THEBERGE: And I think our -- we have two committee members who joined us here in the room. If we could ask you just to introduce yourselves and if you have anything to disclose?

MEMBER MILLER: Are you -- this is Marlene. Are you talking to people on the phone?

You said in the room.

MS. THEBERGE: Oh, great, you're on the phone. Well, we will have -- why don't you start, and then we also have some folks who joined us in the room.

MEMBER MILLER: Oh, great, okay. So I am Marlene Miller. I am Chief Quality Officer for Pediatrics for the Johns Hopkins Medicine Health System, and I am I guess a returning member for this Pediatrics Performance Measures Committee meeting.

MS. THEBERGE: Great, thank you. And I think we have just one new committee member. Sorry.

MEMBER KNUDSEN: I am Kraig Knudsen. I am the Chief of the Bureau of Research and Evaluation at the Ohio Department of Mental Health and Addiction Services, and I have nothing to disclose.

MEMBER MILLER: Oh, I am sorry. This is Marlene. I should say I have nothing to disclose either.

MS. THEBERGE: Thank you, both. Okay. With that, I think we are ready to begin our measure discussion, and so we will invite our developers from University of Michigan who I believe are dialing in on the phone.

DR. REEVES: That is true. We are here on the phone.

MS. THEBERGE: Great. Thank you.

CO-CHAIR SUSMAN: So John and I are going to sort of alternate to spell each other, and I have the good fortune to tackle the first one here. We're probably going to take a little longer since everybody is getting back used to the format. We will start with the developer overview, try to keep that relatively brief, but don't hesitate to ask questions. Then I will ask the lead discussants to give their feedback. We will go by each of the criteria, starting at the top and working our way through, remembering the must-pass, and then at the end, we will take a final vote on the measure, assuming it passes the must-pass criteria.

If you have any questions, again, just hold up your cards, and we will be able to proceed efficiently. So without further ado, please welcome our developers to give us an overview.

DR. REEVES: Good morning, everyone. Can everyone hear me okay?

CO-CHAIR SUSMAN: Sounds good.

DR. REEVES: Great. Hi. My name is Sarah Reeves. I am an epidemiologist with the Q-METRIC team at the University of Michigan.

Our measure list sets a percentage of children that are ages three months to five years old with sickle cell anemia that received appropriate antibiotic prophylaxis in the year. Development of this quality measure for sickle cell disease was determined by CMS through a public process to be a priority for the nation, and this measure is supported by strong evidence. Children with sickle cell anemia are at almost 100 times the risk of an invasive pneumococcal infection as compared to children with normal

hemoglobin.

What is even more concerning about this is that these infections are associated with high mortality. One report indicates that case fatality is at least 30 percent. And importantly, these infections can be largely prevented.

In a randomized control trial, children were randomized to receive either a daily placebo or penicillin, and among children in the penicillin arm, the incidence of infection was reduced by 84 percent. So given the success of this antibiotic prophylaxis in reducing infection, an expert panel at the National Heart, Lung and Blood Institute, or NHLBI, strongly recommends that all children with sickle cell anemia should receive twice-daily penicillin prophylaxis up until the age of five.

So in concordance with these guidelines from NHLBI, our measure uses administrative claims, and it assesses the proportion of children ages three months to five

years with sickle cell anemia that receive appropriate antibiotics within the year. As you all know, there is a lot of hemoglobin variations, and our measure focuses on the HbFS cases, or sickle cell anemia, and this is consistent with the NHLBI recommendations.

For the measure, our denominator is the number of children with sickle cell anemia, and we identify these children as children that have at least three healthcare encounters associated with sickle cell anemia within a year. We found that compared to the gold standard of newborn screening records, this case definition is valid, and it has high sensitivity and specificity to identify children that should be in the denominator.

Our numerator is the number of children with sickle cell anemia that received at least 300 days of appropriate antibiotic prophylaxis within the year. Receipt of antibiotics is identified through the presence of administrative claims for specific NDC codes that

are listed within the specification, and we did find also that this method for the numerator was valid, as it was highly correlated with the written prescriptions in the medical records.

So in addition to the evidence that our definitions for the numerator and the denominator were valid, we assessed the reliability of this performance score using a signal-to-noise analysis with Medicaid data across six different states. We found that the reliability of this measure was high. All of the reliability coefficients were over 0.8, and these usually -- and these all range from 0 to 1.

We used the same Medicaid data from six different states to assess the rates of antibiotic prophylaxis, meaning as the measure is 300 days filled within the year, from 2005 to 2010, and across this period, children with sickle cell anemia contributed over 3000 person-years to the study population. To our knowledge, this is the largest study population to ever assess antibiotic prophylaxis among children with

sickle cell anemia.

Among the children in the study population, the proportion that received at least 300 days of antibiotics ranged from 3 percent to 36 percent. This obviously indicates that substantial opportunities for improvement exist. In addition, we were able to identify differences in the performance scores by state.

So overall, we feel that this measure is highly important and has a real potential to make a substantial positive impact in these children's lives. Our measure focuses on appropriate antibiotic prophylaxis. This is an effective method to reduce the incidence of infection among children with sickle cell anemia. We found the measure is highly reliable and valid, and the data elements are readily available and administratively straightforward.

We also found an important and large performance gap, and we strongly believe that endorsement of this measure could have a very positive impact on the health of these high-risk

children. We really look forward to talking to you guys about this measure and addressing any questions or concerns that you might have. Thank you for your time.

CO-CHAIR SUSMAN: Thank you for that very lucid overview. So we will go on and ask Jeff to provide his overview comments and discussion around evidence.

MEMBER SCHIFF: I want to thank the folks from Michigan for the succinct and complete response.

I -- I want to go through this. I think the best thing I can do is maybe talk about how I went through this and looked at some of the evidence and then looked at some of the questions I had without getting -- and then we can get into some issues specifically around validity, reliability, and et cetera.

I think that the -- I guess the -- the first thing I will say, and I, in looking at the literature, I think that the authors appropriately identified a significant issue for

a relatively small population, but one that has, I agree, has significant mortality. I wanted to look at a few things that may alter this, and I think that they addressed these, or the literature addressed them, but I just want to bring them up for review.

One is some of the studies are a little bit old, and they were -- they -- the relationship of the studies to the -- to the pneumococcal vaccination was an issue that I looked at, and they actually addressed that. The polysaccharide vaccines were available, and one of the -- I am not going to go -- I can go dig it out of there, but the -- the literature did show that -- that despite the polysaccharide vaccine, there was still a risk to these kids, and that is I think an important thing to note, that we don't -- we're not out of this problem because of the pneumococcal vaccine or the expansion of the pneumococcal vaccine.

They also looked at -- they also addressed, or some of the literature addressed,

the resistance issue, and I think that -- you know, the issue of whether we are going to create more resistant organisms, and I think they have addressed that as well, and I will let the authors discuss that.

And so those are just I think two of the most important things in the literature that I think were addressed that make this still a viable and important measure. The other things then I just wanted to talk about with regard to the measure itself, the -- a couple things I thought were interesting that may be worth discussing, and one is that I thought it was interesting that it took three visits to the -- to -- to reliably identify someone with sickle cell, and that was the -- the increase in reliability as we went through from one visit to three visits was -- I guess that was concerning to me, like I just -- seems like that is a diagnosis that should be fairly straightforward, but if you look at the numbers, they went up pretty significantly as you went along, so I

thought that was -- I thought that was important.

I liked the idea of 10 out of 12 months because I think that if we are looking at people who are actually using the drug -- you know, there's a number of steps here. The prescriber has to prescribe it. That is not what we are measuring because they could prescribe it, and it couldn't be filled, but if you look at for the -- the jump from prescribing to actually having the medicine filled, which is where the claim comes in, I think that that is a reasonable thing, and I think it is -- I don't know if it is reasonable to assume, and I don't think we really know from this measure whether or not if someone prescribed it and they filled it, they are actually getting the med, but I think that if they are filling it over -- for 300 out of 360 days, I think that that is a relatively good thing.

Most of these children are in the Medicaid program, so I think that it is a reasonable measure for Medicaid as well. And

then -- I think I will stop there. I think those are the key points that I got from reviewing their -- the measure as it is developed and some of the -- I guess the questions that were raised for me as I thought about what they were putting forward.

CO-CHAIR SUSMAN: Thank you very much. I am going to go then to each of the other reviewers, and then open it up for further. So Rajiv?

MEMBER MODAK: Thank you. So a couple things. I agree with much of what Jeff said. I wanted to go to that point again about the three visits because that stuck out to me as well. So a couple of things with that: by excluding patients with less than three visits, what I was concerned with was the possibility that that could be a significant group of people that you're missing.

In my practice, when we do quality improvement measures, for example something simple like immunizations, you know, there's two

steps. One is to -- to take care of those kids that you are actually seeing in the office, and then the next step is also to take -- reach out to those kids that don't come in, and that is where I potentially see you are missing the kids who might be at highest risk for sepsis because of their lack of compliance with care by not addressing them with revisits.

And then, but the other side of this is that the fact that we are dealing with kids who often are -- 80 percent or so I think who had Medicaid at some point, significant opportunity to address -- to address a gap in care for -- for the potentially most underserved population, so I will put that out there.

CO-CHAIR SUSMAN: So maybe it would be appropriate to ask our developers to very briefly discuss the rationale for the three-visit criterion?

DR. REEVES: Sure. I would be happy to.

The Table 4 that we have in our

testing form lays out the sensitivity and specificity of three different case definitions that we looked at to identify children with sickle cell anemia, and as you can see, the area under the ROC curve is really best for three -- having at least three claims for sickle cell anemia within the year.

Our concern with loosening that definition any more was appropriately including children in the denominator that were not even eligible to receive antibiotic prophylaxis, and we felt like that would then dilute the measure inappropriately, so that is why we decided to go with the three healthcare visits for sickle cell anemia within the year.

CO-CHAIR SUSMAN: Thank you. I think hopefully folks can look on the screens in the center aisle and -- or their own materials, so the ROC curve analysis that was done, a tradeoff of sensitivity, specificity.

Okay. There was a question or a comment from Jeff.

MEMBER SCHIFF: Oh, I just wanted to further what the developer said. They -- from one visit, they went from true positives of 409, three visits was 374, so they lost a lot of false positives, but they didn't lose very many true positives, so I thought that was an okay tradeoff.

CO-CHAIR SUSMAN: Okay. Good. So I haven't heard any qualms about the evidence. We will go to Lauren and Maureen. What I would ask you to do is not repeat. If you agree, you can just indicate. And Lauren?

MEMBER AGORATUS: Yes. I agree with a lot of what was already said. Under evidence, I was a little bit concerned that it was not -- there was no way to clarify if additional, for example, not prophylactic antibiotics were administered for breakthrough infections. Also, the sources of evidence were related to penicillin, and yet there would be other antibiotics that could be used such as erythromycin, sulfamethoxazole, and trimethoprim.

And then under the performance gap, there was a concern noted about the small sample size, and I didn't see anything on racial and ethnic disparities. Under the specifications for reliability, I was also concerned with the three health encounters. Also, perhaps consideration of exclusion for children who could have suppressed immune systems due to comorbid conditions such as organ transplant, cancer or other immunosuppressive medications such as steroids, Humira, et cetera.

Under reliability testing, again, the generalization was questionable due to the small sample size of the empirical evidence that was presented. Under the validity testing --

CO-CHAIR SUSMAN: We are going to try to just stick to evidence right now, and I think --

MEMBER AGORATUS: Oh --

CO-CHAIR SUSMAN: -- everything you are --

MEMBER AGORATUS: -- okay, okay --

CO-CHAIR SUSMAN: -- raising are -- are --

MEMBER AGORATUS: Yes.

CO-CHAIR SUSMAN: -- great comments, but we will plow through evidence first, and then we will make sure we get to the other concerns you're raising.

MEMBER AGORATUS: Okay.

CO-CHAIR SUSMAN: Okay.

MEMBER AGORATUS: I am done.

CO-CHAIR SUSMAN: Okay. Sorry to cut you off. Maureen?

MEMBER EDIGER: I don't have anything else to add.

CO-CHAIR SUSMAN: Okay. So let's open it up for the panel to ask any questions or clarifications. When you are through talking, if you can remember to turn off your speaker. I think after a while, there's only so many people that can speak at once.

MEMBER MILLER: This is Marlene on the phone. I don't -- can't raise my little hand, so

I will -- put me in the queue, though. When you have questions to go around, I have some.

CO-CHAIR SUSMAN: Why don't you start off?

MEMBER MILLER: Oh, great. My question had to do with just the reality of who is providing this care for these children. A colleague that I know well was doing a lot of research on this and found there's discrepancies sometimes between whether it's the primary care provider that is providing the sickle-cell-specific care, or at times the hematologist, and so I have two questions for this.

The first is does the algorithm to identify three visits -- you identify whether they go to the primary care provider for these sickle cell visits, or to a pediatric hematologist? And my second related question has to do with the realities of care for children, particularly when the children's hospital, which is where pediatric hematologists are going to be congregated -- they are usually not in private

practice, per se -- what if that is across a state line, so it's in a different Medicaid database?

For example, what if the child lives in Maryland and gets their primary care in Maryland, but happens to see a pediatric hematologist at D.C. Children's? Will this methodology pick up those kids across borders? Because the reality is there's only about 80ish, you know, kind of full-service children's hospitals in this country, and then many are going to be across state lines for some children.

CO-CHAIR SUSMAN: So it sounds like a great question for a developer to clarify.

DR. REEVES: Great. So first I would like to address your first question about the algorithm and if it distinguishes between visits with primary care physicians or hematologists.

So the algorithm is actually even broader than being types of visits. It is any claim whatsoever that is associated with sickle cell anemia. This could be even durable medical

equipment, or a laboratory claim, or any -- it is a count, just a count of three claims that deal with sickle cell anemia, so it is not limited to those that are seen by a primary care or a hematologist. Does that make sense for that question?

CO-CHAIR SUSMAN: Sounds --

MEMBER MILLER: Yes, that is good.

(Simultaneous speaking.)

CO-CHAIR SUSMAN: -- yes.

DR. REEVES: Okay. Great. And the second question, if it can pick up kids across state lines, so that would really depend on what data set you were actually using. If you were using the state-specific Medicaid data, it really wouldn't pick up those kids, but if you had maybe the MACs data set which we use that goes across states, you could tie children across states, or perhaps a private insurance company could track the children irrespective of the state that they receive, so I think that that question would really depend on the data set that you were using

to identify the children.

MEMBER MILLER: So given that the majority of these kids are likely in a Medicaid system, the first one you mentioned that you use that crosses states, what states are included in that?

DR. REEVES: So for the purposes of this, we used Illinois, Michigan, South Carolina, Texas, Florida, and -- what was -- Louisiana.

MEMBER MILLER: So you have a tool that links all those as one seamless large data set?

DR. REEVES: Yes.

MEMBER MILLER: But all the other states, would they be able to be linked? And, you know, this is the perpetual holy grail of linking all the Medicaid databases so we could get a perspective on children's care across the country.

DR. REEVES: It could do that. I mean, to be totally honest, the MACs data set that we are talking about is also relatively

delayed, so there is a several-year gap before we can look at that, so really, the state Medicaid databases are more current and are available more quickly and would be able to assess improvements in care on a quicker basis.

CO-CHAIR SUSMAN: I am going to go --

MEMBER MILLER: But --

(Simultaneous speaking.)

CO-CHAIR SUSMAN: -- down -- oh, go ahead.

MEMBER MILLER: I was going to say, but then to use them, you would have to be able to link visits, so my example, you know, of a child in Maryland and a -- that goes to D.C. for their hematology care, or you might miss visits and ordering of prescriptions and ordering of durable med equipment.

DR. REEVES: That's a possibility.

CO-CHAIR SUSMAN: Good.

(Simultaneous speaking.)

DR. REEVES: Thank you.

CO-CHAIR SUSMAN: Thank you. Let's go

down the row here and start with Jill.

MEMBER MORROW-GORTON: So maybe we are not supposed to do this, but I am a state Medicaid Medical Director. All Medicaid programs pay for over-the-border services. I think the -- it is 50 miles, usually, and -- and many Medicaid programs will approve visits at specialty places out of state if that is the only place you can get it, so that would -- that claim would come back to the Medicaid program.

So -- so if you are from Maryland, and you get seen at National Children's, the claim that National Children's submits goes to the Medicaid program in Maryland, so it would be reflected in their data.

CO-CHAIR SUSMAN: Okay.

MEMBER KELLER: Hi, David Keller. So the -- I may be dating myself, but way back when, when I was a resident in Baltimore, we brought these kids in once a month for shots of long-acting bicillin, and it is still listed in the NHLBI guidelines as an approved method of

prophylaxis. Is that captured? I was hunting through that really long file of which drug codes were captured, and I could not find it in there. Is it in there and I have just missed it?

CO-CHAIR SUSMAN: Sounds like --

DR. REEVES: No, we --

CO-CHAIR SUSMAN: Go ahead.

DR. REEVES: I am sorry. We did not include that in the measure for the oral antibiotics.

CO-CHAIR SUSMAN: Okay. Any other questions or comments regarding validity? I mean evidence. I am sorry. Yes, thank you. Yes?

MEMBER BRISTOL-ROUSE: This is Tara Rouse. So Jeff had brought up the issue of resistance, and I was wondering if the primary reviewers could say anything more about the benefits outweighing the potential risks.

DR. REEVES: I am sorry. Was that a question for the developers?

MEMBER BRISTOL-ROUSE: Sure. Whoever.

DR. REEVES: Okay. Thanks. When we

thought about this measure, we did think about antibiotic resistance, and felt that it wouldn't be any more likely in this population to develop antibiotic resistance than any other population.

CO-CHAIR SUSMAN: Were there other complications or negative consequences considered?

DR. REEVES: We did consider the possibility that a child could be allergic to penicillin, and that would not be prescribed, and therefore actually in our definition of antibiotics in the NDC codes, and this also is a question that I believe Maureen asked earlier, we included penicillin, but we also broadened the definition just a little bit to include erythromycin for children that could be allergic to penicillin.

CO-CHAIR SUSMAN: Jeff, you had a comment?

MEMBER SCHIFF: Well I just wanted to ask the developer, you had a number around resistance in this population I thought in your

literature that sticks in my head. It was relatively low, but I didn't know if you -- that was nine percent, but is that correct?

DR. REEVES: I am looking at the literature. I actually can't remember the exact number. I might get back to you about that in just a minute. I am sorry.

CO-CHAIR SUSMAN: I think the --

MEMBER MILLER: This is Marlene --

CO-CHAIR SUSMAN: -- the --

MEMBER MILLER: -- I have a question that is related to the unintended consequences. Was there also consideration of post-secondary infections such as C. difficile, you know, the rate of, you know, the complications because of the constant use of antibiotics?

DR. REEVES: No. We did not look into that.

CO-CHAIR SUSMAN: Okay. Any other questions or comments? And yes, we are on evidence.

Looks like we are probably ready to

vote here. So again, we're looking at the evidence, the outcome, and I think this can be high, moderate, low or insufficient. And we will vote.

MS. JUNG: Voting is now open for Measure 3166 for evidence. Option 1 is high, option 2 moderate, option 3 low, and option 4 insufficient.

MEMBER MILLER: How do I vote? This is Marlene.

MS. JUNG: Please submit your vote via the chat function, and the staff will enter it for you.

MEMBER MILLER: You've got to help me find it. Oh, I see chat here.

MS. JUNG: Voting is open, yes. We are still waiting on a few votes.

MEMBER MILLER: Can you confirm you have my vote?

MS. JUNG: We do now, yes. Waiting on one more.

Voting is now closed. The results are

39 percent for high, we have nine votes for that; 57 percent for moderate, 13 votes; 4 percent for low, one vote; and 0 percent for insufficient, with zero votes. And the criteria has passed for evidence.

CO-CHAIR SUSMAN: So that is great. Just remind you that this is a must-pass criteria, and it easily passed, obviously. So we are going to go ahead, and I will ask Jeff to just comment briefly on gap.

MEMBER SCHIFF: I don't -- the developers talked about a gap that I -- a gap that I pulled out of the -- of their stuff was 5.7 to 36 percent state range, so there is a lot of variability in states. Illinois must -- appears to be doing better than some of the other states, and I am glad Minnesota, I don't think was measured. But -- but I think there is a -- a significant variation and a significant room for improvement.

CO-CHAIR SUSMAN: Okay. Rajiv, anything different?

MEMBER MODAK: Just to add that I was shocked how low that was.

CO-CHAIR SUSMAN: So two votes sounding like there is a significant gap. Lauren, on the phone?

MEMBER AGORATUS: Nothing else.

CO-CHAIR SUSMAN: Okay. And Maureen, anything else to add?

MEMBER EDIGER: No.

CO-CHAIR SUSMAN: Okay. Opening up for any committee comment, and if not, we can vote on gap.

(No response.)

CO-CHAIR SUSMAN: Okay.

MS. JUNG: Okay. Voting for Measure 3166 for performance gap is now open. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

Lauren, can you submit your vote? There it is. Thank you.

MEMBER AGORATUS: I did. Okay.

MS. JUNG: Voting is now closed. The

results are 78 percent high, 18 votes; 17 percent for moderate, with four votes; 4 percent for low, with one vote; and 0 percent for insufficient, with zero votes. And it has passed for performance gap.

CO-CHAIR SUSMAN: Okay. We are now going to move on our discussion concerning reliability and validity, so if we can talk. This first, maybe I will give -- Rajiv, would you like to take a first shot about reliability?

MEMBER MODAK: So as far as reliability, it seemed to me that there were -- I noted that there were hundreds of patients from each state measured, and the reliability was consistently greater than 80 percent for all the states, so to me, that sample size does seem sufficient to generalize for widespread reliability.

CO-CHAIR SUSMAN: Jeff, further comments?

MEMBER SCHIFF: I don't think anything besides what we have already talked about.

Thanks.

CO-CHAIR SUSMAN: Lauren, any --

MEMBER AGORATUS: Me neither.

CO-CHAIR SUSMAN: -- commentary? No? And Maureen?

MEMBER EDIGER: No.

CO-CHAIR SUSMAN: So it looks, in briefly reviewing, that there was -- oh, John? I am sorry. Go ahead, please.

MEMBER FINKELSTEIN: So I have one question for the developers. In claims data, in pharmacy claims data, the days supplied field is -- is, in data sets that -- that I have used, is notoriously finicky, sometimes not populated. The agent is always populated well, but the days supplied sometimes isn't, so I wondered in your analysis if in this data set it was always populated, if it had extreme values that didn't make sense, and what you did when effectively the prescriber wrote quantity sufficient, which is not so uncommon. How did you -- how did you count -- how did you count those? How did you

come up with 300 days?

DR. REEVES: Thanks for the question. We came up with 300 days because, I mean, as you guys had said, it was about 10 out of the 12 months, and we felt that that allowed for a little flexibility.

As far as the specific claims, our days supplied was populated relatively well. We -- we did look at a lot of different methods to try to make sure we were not having very low-quality data, so we had very very few kids that had zero days' supply, or even under 50 days' supply. Actually, the average days supplied was 186 I believe, so we felt that even if there was some incompleteness within those claims, that it was relatively well reflective of the number of prescriptions that were filled through that time.

MEMBER FINKELSTEIN: So I -- I am surprised at 186 days supplied. So where I work, I think there is a three-month limit, so -- so really people were writing single scripts for -- or was that with refills? Did you count refills?

I am just trying to get at how you did this, if --

DR. REEVES: Yes, I did count the refills. That would be 186 days throughout the entire 12-month study period.

MEMBER FINKELSTEIN: I see. So yes, I understand. Yes, plus -- I mean, yes, but it's going to average out. But these are dispensings. These are pharmacy claims, so they are dispensing, not prescriptions, so these are dispensing events that totaled 186 days, is that what you're saying? It is not one dispensing event of 186 days.

DR. REEVES: Exactly, you are correct. It is all of the dispensing events across the year summed.

MEMBER FINKELSTEIN: Right.

CO-CHAIR SUSMAN: Okay. Other clarifications about reliability?

(No response.)

CO-CHAIR SUSMAN: If not, I think we can move on to a vote on reliability.

MS. JUNG: Voting is now open for Measure 3166 for reliability. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

MS. MCQUESTON: Thanks, Laura and Marlene. We have your votes.

Voting is now closed. The results are 30 percent for high, with seven votes; 70 percent for moderate, with 16 votes; 0 percent for low, with zero votes; and 0 percent for insufficient, with zero votes. The measure has passed reliability.

CO-CHAIR SUSMAN: Okay. Now we move on to validity, and maybe just to mix things up, I will ask Maureen to comment first.

MEMBER EDIGER: I defer to my other discussants.

CO-CHAIR SUSMAN: Okay. And Lauren?

MEMBER AGORATUS: The only thing I had was that the validity testing of the measure, it was reassuring to see that in addition to just face validity.

Again, some concern over the sample size because those that were used were smaller than the ones in the evidence, which were considered small, but I know other folks considered them sufficient, so we will see.

CO-CHAIR SUSMAN: Okay. Rajiv?

MEMBER MODAK: So I just was reassured by the high correlation between the antibiotics that were prescribed and the antibiotics that were dispensed, based on administrative claims because I was skeptical of that beforehand.

CO-CHAIR SUSMAN: And finally, we will move to Jeff.

MEMBER SCHIFF: I -- I don't think I have anything significant to say. I thought the changes in -- in odds of dispensing 300 days over time changed a little bit. It is Table 7. So I don't know if the authors want to address that. That bounced around a little bit, and I didn't know if I could explain why that is -- why that is the case. That is probably the only concern I had.

CO-CHAIR SUSMAN: So is this a question maybe briefly for the developer --

MEMBER SCHIFF: Right.

CO-CHAIR SUSMAN: -- on Table 7?

MEMBER SCHIFF: Yes.

DR. REEVES: Sure, thanks for that question. As you all know, we also had submitted the TCD measure last year, and we did see increases over time in TCD screening rates, so we were really hopeful that we would see increases in antibiotic prophylaxis over time also in this population, and we really didn't. And so I think that this is, to be totally honest, just reflective of a lot of opportunity for improvement and a lack of improvement efforts over the last few years in antibiotic prophylaxis.

MEMBER SCHIFF: Thanks.

CO-CHAIR SUSMAN: Any other questions? I don't see any cards up -- or on the phone?

(No response.)

CO-CHAIR SUSMAN: Then let's go ahead

and move to vote on validity.

MS. JUNG: Voting on validity for Measure 3166 is now open. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

Voting is now closed. The results are 39 percent for high, with nine votes; 61 percent for moderate, with 14 votes; 0 percent for low, with zero votes; and 0 percent for insufficient, with zero votes. The measure has passed for validity.

CO-CHAIR SUSMAN: And now on to feasibility. So if I can ask Jeff maybe to briefly comment on feasibility?

MEMBER SCHIFF: We in Medicaid like claims-based measures because we don't have to go to our providers and ask for them to be calculated because we get -- we get beaten up at retirement parties for --

(Laughter.)

MEMBER SCHIFF: -- and at other times when people -- when people say why are you making

us do chart abstractions for this stuff? So I think this is a measure that actually gets to a significant quality issue that is -- that is feasible to do, and with enough reliability and validity, so --

CO-CHAIR SUSMAN: Okay.

MEMBER SCHIFF: -- comment --

CO-CHAIR SUSMAN: It's going to be hard to top that. Rajiv?

MEMBER MODAK: I agree. In fact, just as an anecdote, we in my health center actually have direct log-ins to Medicaid, so we can actually receive our own data directly, which is immensely helpful for outcomes.

CO-CHAIR SUSMAN: That's cool.

MEMBER MODAK: Yes.

CO-CHAIR SUSMAN: Wow. All right. Lauren?

MEMBER AGORATUS: No concerns.

CO-CHAIR SUSMAN: And Maureen?

MEMBER EDIGER: Nothing additional.

CO-CHAIR SUSMAN: Okay. Any other

comments or questions?

(No response.)

CO-CHAIR SUSMAN: Seeing none, we will move on to a vote on feasibility.

MS. JUNG: Voting for feasibility for Measure 3166 is now open. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

Voting is now closed. The results of voting are 74 percent for high, with 17 votes; 26 percent for moderate, with six votes; 0 percent for low, with zero votes; 0 percent for insufficient, with zero votes. And the measure has passed for feasibility.

CO-CHAIR SUSMAN: And that leads us to usability. How about, Rajiv, you want to talk a little bit about usability and use?

MEMBER MODAK: Yes, and I just noted that the measure is not currently being used as a quality measure, and -- but it obviously has the potential to improve high-quality care by reducing the burden of preventable pneumococcal

infections. And I just noted that -- no, I take that back. I noted that the antibiotic resistance was not quantified, but that actually was, so that is it.

CO-CHAIR SUSMAN: Jeff?

MEMBER SCHIFF: I think at the level of either health plans or the state, this is a -- a very viable measure because I don't think it is designed to be a measure for individual clinicians, but I think it assesses the ability of -- of organizational structures to get these kids on these meds.

CO-CHAIR SUSMAN: Okay. And then we will go to Lauren?

MEMBER AGORATUS: No additional concerns.

CO-CHAIR SUSMAN: And Maureen?

MEMBER EDIGER: Nothing additional.

CO-CHAIR SUSMAN: All right. Let's vote on usability and use, unless there are any questions?

(No response.)

CO-CHAIR SUSMAN: And again, seeing, hearing none, we will go to vote.

MS. JUNG: Voting is now open for Measure 3166 for use and -- usability and use. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

CO-CHAIR SUSMAN: Are we having voting issues?

MS. JUNG: Yes, we are having some voting issues. Just a moment and we will try and reload this slide here.

CO-CHAIR SUSMAN: While we are correcting the technical difficulties, I just want to thank everybody for really being succinct and on point with comments. It really helps us to move along, and you evidently have been studying how to do this, because everybody has been quite up to par. Now, if we can get our computers to work, we will be doing fine.

Any questions about process or comments about process while we are waiting here?

(No response.)

CO-CHAIR SUSMAN: Okay. Thanks.

Just to highlight, we're going to go on although we were supposed to break after this first measure. You guys have been so wonderful we get to do another one. But looking ahead, it is probably a good idea that we use our time to get -- plow through this.

MS. JUNG: Okay. Let's try that again.

CO-CHAIR SUSMAN: Okay.

MS. JUNG: Voting is now open for Measure 3166 for use and usability. Oh, it is working, great.

CO-CHAIR SUSMAN: Yay.

MS. JUNG: Waiting on -- oh, voting is now closed. The results are 52 percent for high, with 12 votes; 48 percent for moderate, with 11 votes; 0 percent for low, with zero votes; and 0 percent for insufficient, with zero votes. The measure has passed for use and usability.

CO-CHAIR SUSMAN: Okay. And now, if I am not mistaken, we get to vote on the overall

measure.

MS. JUNG: Okay. Voting for Measure 3166 for overall suitability for endorsement is now open. Option 1, yes; option 2, no.

Voting is now closed, and we have reached consensus with 100 percent, with 23 votes. This measure is now endorsed by NQF.

CO-CHAIR SUSMAN: Well done, folks.

MS. JUNG: Recommended for --

CO-CHAIR SUSMAN: Excellent.

MS. JUNG: -- endorsement.

CO-CHAIR SUSMAN: Yes. Give a round of applause for yourselves.

(Applause.)

CO-CHAIR SUSMAN: So, thank you. We are going to do a switch, and John is going to take on the next couple, but I just want to thank everybody for being really efficient and directed with their comments.

CO-CHAIR BROOKEY: Do we have a developer here for the -- the GAPPS preventable adverse event measure?

Great. So we're going to NQF Measure 3136, and I will ask you to introduce yourself and give us a brief overview of the measure, if you would.

DR. LANDRIGAN: There it is. Is that it? Great, thanks.

So thanks very much. My name is Chris Landrigan. I am a pediatric hospitalist and health services researcher at Boston Children's Hospital, and I am joined today by Mark Schuster and Sara Toomey from the Boston Children's Center, as well as David Stockwell from Children's National here in D.C. We represent the core team that developed the GAPPS, or Global Assessment of Pediatric Patient Safety Measure.

As many of you know, adverse events, events in which medical care causes harm, have been recognized as a leading cause of death and injury in the U.S. since the 1990s. Although there is now excellent awareness of the problem of patient safety in U.S. hospitals, systematic measurement of AE rates in hospitals is lacking.

The large majority of hospitals in the U.S. currently identify preventable adverse events through voluntary reports by staff.

Unfortunately, a series of studies have shown that only between 2 and 10 percent of all adverse events are reported through these systems. As a result, comparisons across hospitals are not reliable, as it is unclear if any differences represent true differences -- excuse me -- or simply differences in reporting patterns.

Over the past decade, work has been conducted to develop a more systematic approach to measuring rates of adverse events using so-called trigger tools. Triggers are not themselves adverse events, but they help to identify them because triggers are often associated with an adverse event. When a trigger occurs, review of the medical record is required to confirm whether an adverse event did or did not in fact occur.

An example of a trigger is the

administration of Naloxone. If a chart is identified because Naloxone was given, a review might find that an adolescent was brought to a hospital because of a heroin overdose and was given Naloxone to save her. If that is the case, the trigger would not have identified an adverse event. However, if Naloxone were given because the hospital gave the patient a narcotic overdose as part of treatment, then there was a preventable adverse event. A review of the record is required to sort these things out.

Studies of trigger tools have shown that they are far more sensitive and reliable than voluntary reporting systems or screening tools that reply on administrative billing data alone. Trigger tools identify adverse events at 10 to 100 times the rate of these other methods, and with better specificity.

To develop GAPPS, we began with a systematic review of the pediatric and adult literature to identify candidate triggers for possible inclusion in GAPPS. In addition, we

spoke with trigger developers. From our review, we compiled a list of 78 candidate triggers for possible inclusion in GAPPS.

Using the RAND/UCLA Appropriateness Method, we convened an expert stakeholder panel to review candidate triggers and assess their validity and feasibility. Panelists were chosen as -- or nominated rather by professional societies from the U.S., including for example the Academic Pediatric Association, the American Nurses Association, Consumers Advancing Patient Safety, and the National Patient Safety Foundation.

An initial list of 54 triggers were determined by our panel to be valid and feasible and were included in our draft tool. To rigorously evaluate the performance of GAPPS across a range of pediatric inpatient settings, we identified 16 hospitals to participate in the GAPPS national field test through the Pediatric Research in Inpatient Settings, or PRIS, Network. In each hospital, we reviewed approximately 240

records per hospital from randomly selected admissions for a total of 3814 records in total.

Nurses served as our primary reviewers. Primary reviewers presented all suspected adverse events to two secondary reviewers, who were physicians working in the study hospitals. These physicians independently made final determinations about the presence, severity, and preventability of any suspected adverse events. Secondary reviewers discussed and resolved any cases for which they had initial disagreement.

Of note, physician reviewers do not conduct reviews of the entire medical record. They only review suspected incidents identified by the nurse primary reviewers. These secondary reviews take about five minutes each. Since only a fraction of charts contain a suspected adverse event, the time burden for physicians is minimal.

GAPPS proved to be a reliable tool for detecting preventable adverse events. Primary reviewers agreed 92 percent of the time on the

presence or absence of a suspected adverse event, with a kappa of 0.69. Secondary reviewers verifying adverse event presence or absence agreed 92 percent of the time, with a kappa of 0.81.

We refined the GAPPS tool at the end of our study based on an analysis of each individual trigger's incident and so-called positivity rate, that is, the frequency with which a particular trigger indicated the confirmed adverse event. Following removal of triggers that did not yield many adverse events, the tool was streamlined to include 27 triggers in the Manual GAPPS Trigger List.

To further decrease the time burden on nurse reviewers, we also developed an automated trigger list that allows hospitals to scan for triggers through the electronic health record. Using the automated list, the process of nurses investigating these triggers and then reviewing suspected adverse events with physicians is unchanged. The automated trigger list simply

speeds the initial step of looking for triggers.

In order to measure rates of preventable adverse events, we recommend that hospitals conduct 60 chart reviews per quarter. This requires a maximum of 30 hours of nurse reviewer time per quarter and an estimated maximum of one to two hours of physician time per quarter to conduct reviews.

We believe that GAPPS fills a critical need for a public measure of pediatric healthcare safety. It will serve as a valid and valuable tool to assess health system quality and motivate improvements in patient safety. Thanks for the opportunity to discuss the measure. Look forward to questions.

CO-CHAIR BROOKEY: Thank you very much. I apologize. I forgot in reading this, this is done retrospectively and not concurrently, is that correct?

DR. LANDRIGAN: It is done retrospectively, that is correct.

CO-CHAIR BROOKEY: Okay. So it is not

-- okay.

DR. LANDRIGAN: It is not done in real time --

CO-CHAIR BROOKEY: Ideally --

DR. LANDRIGAN: -- in hospitals.

CO-CHAIR BROOKEY: -- concurrent would be great, so very good. So we have -- I want to remind everyone about the algorithm, as your cheat sheet for this is an outcome measure, and so if we go along the top of the sheet here, which means that the -- the relationship between the measured health outcome should have -- there should be one health -- relationship with one healthcare action.

So when we talk about the evidence review, we need to look at -- from that perspective. So let's see. I think I am actually one of the discussants here, but I am going to defer to my colleagues here, and I believe Jill is the primary discussant, if Jill would like to begin.

MEMBER MORROW-GORTON: Sorry. I was

going to say this is an outcome measure.

(Laughter.)

MEMBER MORROW-GORTON: And I think that you did a really nice description and actually answered a couple of questions that I had reading the -- the information, that it was very helpful to hear that the nurses were your primary reviewers, because there was some question about that, and that the physicians were in fact only reviewing the ones where there was a suspected event and not reviewing the entire record. That I think answers some questions down the road in terms of the rest of the discussion.

I -- I think that most of this evidence is in the adult literature in terms of trigger -- using -- using triggers and that sort of thing. I think the concept of, you know, do you do kind of a sentinel reporting versus looking at every single record or a good number of records. This seems to be a fairly good middle ground in terms of being able to use your resources efficiently and still identify sentinel

events that can lead to processes that need to be improved to try to prevent those from occurring.

I had a couple of questions about community hospitals versus academic hospitals, and I wondered if -- if you could talk a little bit more about that, and then -- and then a couple questions about your -- your automated list versus your manual list because I noted that your -- your automated list had some elements in it that your -- one of them had some elements that --

DR. LANDRIGAN: Sure.

MEMBER MORROW-GORTON: -- the other one didn't because your hospital didn't measure that, but I just was curious why that was not there, so -- .

DR. LANDRIGAN: Sure. So -- so to answer the first part of your question about the community versus academic centers, so the way that the GAPPS field test was conducted is through the PRIS Network, we identified eight community and eight academic centers and rolled

the -- rolled the -- the GAPPS measure out in both, really in order to try to assess how well this performed at a range of different institutions that had varying degrees of electronic health records, for example, and different patient populations, obviously, and found that it was really able to be used pretty reliably in both. You know, the rates of harm, as we can get into if folks are interested, were certainly different in those two settings, but -- but it functioned well in both.

And then with respect to the automated versus the manual trigger lists, you know, really the -- the idea here was that, you know, manual review, while it is not enormously burdensome if the sample size is not huge, still certainly represents some burden for institutions, and there has been a bit of a movement, to some degree in the adult side and to some degree in pediatrics as well, to begin to automate some of these triggers as the past few years have gone by, since many of them are based on laboratory

values and other things that are pretty easily scanned for in the electronic medical record.

And so we did that in our institution and were able to automate the large majority of them. In addition to providing some efficiencies in terms of even the common measures, the other advantage of doing this in an automated fashion is that you can include things that are relatively rare and probably were not worth the effort of a manual comb-through, if a nurse is kind of going through the charts one by one to look for -- look for triggers, but, you know, that might add some degree of value: for example, elevations in relatively rare drugs levels, things like that that can provide clues about adverse events that -- that add a little bit of depth and richness to the measure.

CO-CHAIR BROOKEY: Okay. Who else is -- Karen?

MEMBER HARPSTER: So I guess to comment from that question, how many -- of the 16 institutions that were in the sample, how many

were electronic, and how many were paper? Because that I think would be -- make a huge difference.

DR. LANDRIGAN: Yes. So the large majority of the institutions actually had some degree of electronic health records there, but, you know, as -- as I am sure many of you are aware, that it is sort of this constantly moving picture at this point where there are institutions like those that are now on Cerner and Epic that have relatively robust electronic health records, and the large majority of their functions, including note-taking as well as labs and pharmacy and everything else, is sort of under one system.

Lots of others within this study set, particularly as we stretched back to 2007, when we began the review process, many of them were sort of in a mixed state, where labs for example or pharmacy might have been electronic, but much of the rest of it was paper.

MEMBER HARPSTER: Sure.

DR. LANDRIGAN: And we found that, you know, really the measure functioned pretty equivalently either way.

MEMBER HARPSTER: Okay. I would say the only thing else I would have to add is that I was astounded by the number of adverse events. The one that, you know, stuck out, 11.1 adverse events per drug event per 100 in patient -- pediatric patients, 74 adverse events per 100 NICU patients --

DR. LANDRIGAN: Yes.

MEMBER HARPSTER: -- and 203 per 100 PICU patients, which that was really astounding.

DR. LANDRIGAN: Right, yes. It -- I think it really is pretty astounding, and particularly some of our sicker patients who are in the neonatal ICU, or in the, you know, pediatric intensive care units, the frequency with which they suffer relatively minor adverse events is tremendous, as David knows. David is an intensivist.

And, you know, these things are -- are

just daily events, unfortunately, in our ICUs. Many of them are minor and transient, but certainly a subset of them are much more serious than that, and --

MEMBER HARPSTER: Yes.

DR. LANDRIGAN: -- important to tackle, I think.

MEMBER HARPSTER: Thank you.

CO-CHAIR BROOKEY: Okay. Ricardo?

MEMBER QUINONEZ: So you have identified, or you presented evidence both from adult and pediatric literature that -- that you can do a better job of identifying these adverse events, but really, what we are trying to prevent is the events.

DR. LANDRIGAN: Right.

MEMBER QUINONEZ: So is there -- is there evidence from the adult literature or pediatric literature that the identification itself improves the ultimate outcome?

DR. LANDRIGAN: Well, I mean, I will -- I suppose I would almost think of it in

reverse, in that if we don't identify the events, it is hard to take actions to improve, I think for a couple of reasons.

First, it is if these things are not transparent and the data about them is not there in the public domain there before the leaders of hospitals, it is difficult for them to commit resources and prioritize efforts to improve patient safety, and I think that that is probably reflected most clearly in the fact that if you look at adverse event studies in adults, particularly in -- and to some degree we now have some data in pediatrics from this GAPPS measure, over the past 6 to 12 years, the rate of improvement has been relatively slow.

And I -- and I think it is because there tends to be patchy focusing on discrete patient safety problems over time, but often, there is not a very well-coordinated, unified institution to drive the total rate of these things down. And so my sense is that, you know, like many other measures that the NQF endorses,

the first step is really to make these things -- make these things apparent to everyone, and then, you know, the hope is that that stimulates improvement.

I don't think that -- that measuring all by itself stimulates improvement, but I -- but it is a critical first step, in my mind.

DR. STOCKWELL: If I could add, the -- there is -- the adult literature is a little bit further ahead than the pediatric literature. The pediatric literature is developing, though, there is no question.

Specifically, to answer your question, there are -- there is evidence that hospitals that have used a manual global trigger tool approach have been able to alter their overall harm rates. That is certainly the case. I would echo what Chris says, though: if it is not measured, it is not managed, and so we think this is one of the great disparities in our measure field, that we are not more robustly measuring safety events in pediatric inpatient setting.

MEMBER MILLER: This is Marlene. I also have some questions I want to get into.

MEMBER HARPSTER: Thank you. Sorry. I am wondering if you could speak to the admission time frame. I -- you did -- you excluded patients that were in the hospital less than 24 hours, and though I appreciate that that does speak to acuity, we are seeing more and more in my hospital patients being, you know, moved out of the hospital --

DR. LANDRIGAN: Sure.

MEMBER HARPSTER: -- quickly, and still having somewhat I would consider high-risk procedures. For example, an appendectomy oftentimes without a rupture is discharged in less than 24 hours, but that patient probably is at risk for some adverse events, and I am wondering if you could speak to that a little bit?

DR. LANDRIGAN: Yes. I mean, I think that that is absolutely right. This measure is -- in a sense, it is broad. We are trying to

capture adverse events across the entire medical center, and really a wide range of adverse events, but it is narrow in the sense that it really is restricted to patients who would traditionally have been considered inpatients as opposed to observation state patients or, you know, the emergency departments transitioning to the outpatient and so forth.

And I -- I completely agree with you that adverse events occur at a very high rate in those settings as well, but just in trying to define the measure and keep things as consistent as possible across settings, that is sort of where we drew our lines.

CO-CHAIR BROOKEY: Marlene?

MEMBER MILLER: Yes. I have two questions. I am wondering if the developers can speak to the feasibility of automating this in IT. As an institution that recently moved from, you know, the Sunrise Eclipsis kind of system to Epic, it astounds me how even though we all -- everyone on Epic, or everyone on Cerna may want

to do the same thing, each one of us has to de novo rebuild it, which is an enormous burden. So I am wondering if they can speak to the ease of the IT automation, how easy that is to -- does it still need specific institutional development work to make that active and live? Because obviously I am trying to get at burden here.

The other one I am confused a bit on the last statements on the development, is the recommendation for this as a sampling strategy over every quarter? And if that is, because that is how I sort of heard what you were saying in the development, how -- what is the recommended sampling to ensure removal of bias of charts?

DR. LANDRIGAN: So I will answer the second question first, and then David will tackle the first one.

So with respect to sampling, our recommendation is that there are 60 charts sampled per quarter across the hospital, or 20 per month, in order to get a sense of -- of what is happening within the institution as a whole,

and we recommend that it essentially be random sampling across the entire institution, where, you know, the simplest method that most of our institutions in the study used is they lined all of the discharges up, the numbers up in a row, and then just used a random number generator to select 25.

DR. STOCKWELL: So in terms of the feasibility of automation, I think we should think about the question that we had earlier in terms of how many of the hospitals were fully electronic. Just like the answer to that question was there are many that are in transition from paper to automation, that certainly is some of the challenge as you work towards automating a full list of -- of triggers.

However, it can be done on -- certainly done on an institutional level. That is where I had gotten my start with looking at automated triggers. We had run reports from -- we have Cerner here in town at Children's National -- from Cerner reports that would -- we

would do daily pulls that would identify the specific data elements, Naloxone in Chris's example, or there are also organizations that can help to extract those data files from any organization with things like HL7 fees or flat file extractions, things like that.

So it certainly is feasible. It is typically a one-time organization to -- to have the local hospital prepare their data to be extracted, and then once that -- once that is done, those things are -- are pretty much in place. So it is fairly feasible.

CO-CHAIR BROOKEY: Jeff?

MEMBER MILLER: I guess I am wondering on the magnitude of that. Anything is feasible with electronic health records. You know, for example, when we try to take algorithms from other children's hospitals that we have learned, it is, you know, many days of a programmer to actually put that up and actually make it live, and so it adds significant burden when you start talking of that programming fee ahead. Did you

in your -- in your tests, did you actually kind of give those automated codes to someone and then measure how long it took them to build that out in their own system de novo?

I know you have done a lot of work on this with D.C. Children's, so you have had the benefit of kind of developing this gradually over time. I am very interested if there was a de novo institution that got the automated codes, and how long did it take them to really get it up live, and what were the FTE costs?

CO-CHAIR BROOKEY: So Marlene, I wonder if we can just hold that thought until we get to feasibility since we are focused now on just getting through the evidence question, and we can come back to that if that would be -- if that would be okay.

MEMBER MILLER: Sure.

CO-CHAIR SUSMAN: Jeff, did you have a --

MEMBER MILLER: Sure, that --

CO-CHAIR BROOKEY: -- question?

MEMBER MILLER: -- would be fine.

MEMBER SCHIFF: I think this is an evidence question. I am curious about -- and I have been trying to -- I looked at this list yesterday, and I can't seem to find it at the moment, but I am curious about what you -- what is not on the list that you thought had high -- that you could not capture through the triggers?

And then I am also curious about -- specifically, I am curious about the diagnostic errors, which I think are harder to capture, and whether or not they --

DR. LANDRIGAN: Sure.

MEMBER SCHIFF: -- whether or not there is anything to say about the importance of those versus some therapeutic errors.

DR. LANDRIGAN: Yes. So in terms of, you know, what -- what's on the list and what's not on the list, we -- we went through a pretty extensive process when we started -- started developing this measure to, you know, really review the literature and get broad input and

consider I think kind of everything under the sun that was possible.

We looked at both published lists as well as, in speaking with developers, things that they used individually in their own institutions and so forth, and tried to develop a sense of, you know, what were going to be good triggers, what were going to be bad triggers, and then, as part of our RAND process, got folks to -- to vote on these, and then sort of subsequently tested them one by one to see which ones played out.

So I think that, you know, at the end of the day, the list that we have here is not terribly different than the list that exists in the adult world or some of the prior kind of pilot pediatric tools that were developed, but it is streamlined. It is things that I think have been more extensively tested where we developed more evidence on the performance of each of these individual triggers, and we are pretty confident that it is a good list that is going to give a really broad picture of what is going on in the

hospital.

I completely agree with your sentiment that there are some -- you know, there are some limitations to this method. It does not capture everything, even though it is pretty broad, and I think one of the areas where it does tend to miss things is diagnostic errors. I don't think it's a great tool for picking that sort of thing up.

This is really a much better measure of errors of commission, where there is an explicit thing that is done wrong that causes harm in a patient, than -- than a measure of omission, like a diagnostic mistake, for example. It does at least capture some of those as well --

(Simultaneous speaking.)

CO-CHAIR BROOKEY: So Marlene, did you have another question about evidence?

MEMBER MILLER: Me? No.

CO-CHAIR BROOKEY: Oh, okay. I think we should move on. I think that we are -- we are really answering the question of whether or not there is one outcome related to this particular

measure, and the outcome would be improve patient safety.

And I am going to call on myself as a discussant, since this is one of my two measures to review, and just to make a comment that in Kaiser, we have had some experience using trigger tools, so we -- we have been collaborative with data classing and so forth, and so I have a little bit of -- of firsthand knowledge of how good it could be, and how difficult it may be.

But as a pediatrician, I would have to argue that we probably should have started these studies with pediatrics instead of adult medicine because the opportunity for error and medication error is just far greater with dose-related dosing of medication, so this is really, as a pediatrician, something that I think is, you know, really fabulous for improving patient safety, even though we are inferring it from the adult data, from what I understand, what you said here.

So the question for the group is do we

-- does this particular outcome measure have a relationship to a desired outcome, which would be inferred here to be improved patient safety, I would guess? And are there any other questions about whether there is sufficient evidence for this outcome measure before we vote?

(No audible response.)

CO-CHAIR BROOKEY: Are we ready to vote? Okay.

MS. JUNG: The voting for Measure 3136 for evidence is now open. Option 1, pass; option 2, not pass. And just to note, we will have a total of 22 votes for this because we have one --

CO-CHAIR BROOKEY: We have one recusal.

MS. JUNG: Yes.

(Pause.)

MS. JUNG: Voting is now --

CO-CHAIR BROOKEY: Press really hard.

MS. JUNG: -- now closed.

(Laughter.)

MS. JUNG: The results are 95 percent

pass, with 21 votes, and one vote -- or 5 percent for not pass, with one vote. The measure has passed for evidence.

CO-CHAIR BROOKEY: Okay. Very good. So we are going to move on to gap. And who would like to start, either Jill or Karen? Karen.

MEMBER HARPSTER: So I think as we kind of talked about, there is a gap. We don't have a pediatric tool of the measure at this point, so this would be the first, you know, taking a stab at it. The one thing again I mentioned before, the number of adverse events, but in the sample of the 16 hospitals that they looked at, there -- they identified 414 events, and 50 percent of them were preventable, so that is, again, a pretty large number, so I think that they have established that there is a gap.

And they also had -- they talked about disparities, too. I should mention that, and a gap in the disparities of racial ethnicity, and the number of chronic conditions had a number of input in the number of adverse events too.

CO-CHAIR BROOKEY: Jill, anything to add?

MEMBER MORROW-GORTON: No.

CO-CHAIR BROOKEY: And I have nothing to add. Any comments from the group, questions?

(No audible response.)

CO-CHAIR BROOKEY: Anyone on the phone?

(No audible response.)

CO-CHAIR BROOKEY: Are we okay to vote on gap? All right.

MS. JUNG: The voting for performance gap for Measure 3136 is now open. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Pause.)

MS. JUNG: Voting is now closed. The results are 73 percent voted for high, with 16 votes; 27 percent for moderate, with six votes; 0 percent for low, with zero votes; and 0 percent for insufficient, with zero votes. The measure has passed for performance gap.

**CO-CHAIR BROOKEY: Great. So we are going to move to reliability, and I think we actually touched on some of these points in the earlier discussion, but I will look to Jill to see if you want to just comment on reliability.**

**MEMBER MORROW-GORTON: I -- I think there was a question about whether or not the RAND methodology does empirical validity, and I would like to hear a little more about what that -- how that works, and -- and how you used that.**

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

**DR. LANDRIGAN: Sure. So, you know, our approach to trying to come up with a valid and reliable tool really took a few steps. Obviously, we started with a literature review, as I mentioned, and then sort of moved from there to taking things that were extracted from that literature review to a review by -- by a RAND appropriateness panel.**

**And really what that was is we had a group of nine organizations, professional**

organizations from around the U.S., nominate members to consider these -- these measures, where we essentially put before them trigger by trigger what -- what we thought the relationship was between that particular trigger and risk of an adverse event, and had them rate, do they think that this thing is a valid measure of that, and then do they think it would be feasible to extract it from the medical record if it went live?

That generated for us really this list of 54 draft triggers, but the ultimate measurement of -- of validity and reliability of the measure went forward a step, where in the national field test, we -- we really looked at how well each of these recommended things had performed, and so I think of it, in this case, it is not the RAND method sort of operating in isolation, but for us, it was really a first step to getting towards more robust measures of reliability and validity that really were based on a direct assessment of the data.

MEMBER MORROW-GORTON: And I think the other -- the other piece that probably needs a little clarification here is that -- that your -- your data showed high specificity but low sensitivity for all of the reviewers, whether they were physicians, or, you know, your -- you considered some of them to be novices and some of them to be experts --

DR. LANDRIGAN: Sure.

MEMBER MORROW-GORTON: -- and they -- and they all had sort of low sensitivity, and as well, they had -- they may have identified similar numbers of events, but they didn't identify the same events, so -- so the question is are there really more events there, and why is the sensitivity low?

DR. LANDRIGAN: Sure. So, you know -- you know, we tried to cross-check this in several different ways to look as carefully as we could at the reliability and validity of the measure, and our primary measure was really within an institution, how consistently can both nurses at

the primary review stage and then physicians at the secondary review stage rate these events? How frequently do they come to the same conclusions about what is going on, in other words, was there an adverse event in a particular chart, and then if so, was it confirmed preventable, what the severity level was, and so forth?

And in doing that, you know, we -- we found that -- that there was really high reliability at that stage, with -- with 92 percent agreement at both the primary and the secondary stage with event identification. Then, as sort of a -- a secondary check of robustness, we then took those reviewers and put them to a pretty difficult standard, I think, which was comparing their reviews against the reviews of experts who had been doing this stuff for many years, and what we found in doing that is that the specificity of what our newly trained hospital-based reviewers found was quite high, you know, sort of in the 90 percent range.

Their sensitivity as compared with expert reviewers who had been doing this for many years tended to be low. They missed, you know, a relatively large proportion of the total events that were there. And just as you're suggesting, that to us indicated that there's probably more in the chart. Sometimes it goes, you know, undetected, and that experience makes some difference.

And we did take -- in trying to kind of further delve into that and understand well what are the limitations in our method for that, we looked at what happened over the course of the study as our reviewers gained more experience and found that their sensitivity actually improved significantly over the course of the study, and furthermore, that those institutions that had done this before -- you know, remember, again, we were a mix of community and academic centers. Some of them had used trigger tools before, some of them had not.

When we subdivided and looked at those

that had a little bit more experience, their sensitivity as well looked much better as compared with the expert reviewers than did the notices', and so we walked away from the process thinking that -- that really what we need to do is have a slightly more extended training program, which we developed. We actually created a series of five videos as the program ended to try to get people to a more expert level to begin. And then -- and then recognize that we think we do need a ramp up period once people begin to really use this where there is some feedback that happens as they begin to go through this process to get them up to that expert level of sensitivity before this really goes live.

CO-CHAIR BROOKEY: Karen?

MEMBER HARPSTER: I have nothing to add.

CO-CHAIR BROOKEY: So that does affect the reliability, though, if there is a learning curve or a -- I mean, that is obvious.

DR. LANDRIGAN: Right.

CO-CHAIR BROOKEY: So I think --

DR. LANDRIGAN: Right.

CO-CHAIR BROOKEY: -- we have to just understand that that is kind of a threat to reliability, right? Any other questions or concerns about reliability?

MEMBER AGORATUS: This is Lauren. I had a -- I understand that the rationale is to prevent events once someone is admitted, except that I was concerned that there was no follow-up for those that were in the ER and not admitted because sometimes lack of treatment could lead to an adverse event. Also, a question as to whether or not this includes mental health, and lastly, I looked at the -- it was a fairly comprehensive trigger list, but it was unclear whether UTIs from prolonged catheterization, which is a fairly common event, was included.

DR. LANDRIGAN: So I will -- I will answer the last question first. Yes, UTIs are included. Any type of a hospital-acquired infection would be included, or an infection

acquired as a consequence of healthcare intervention would be one of our captured types of events.

In terms of mental health admissions, the way that -- the way that the -- the tool is -- or the measure is structured, we are measuring, you know, again sort of really acute hospitalizations where patients are emergently coming into the institution, and that could include coming in for an overdose, you know, a -- a psychotic break, for example. Any type of a really severe mental health condition that led to an acute hospitalization would be included here. It does not include sort of more chronic psychiatric care, if you will, rehab care, newborn nurseries, those types of not-quite-acute-care admissions, even though you might be in a facility for those.

And then I think there may have been a third facet to your question that I am forgetting. Did I miss something?

MEMBER AGORATUS: The -- the ER, the

people that are not admitted --

DR. LANDRIGAN: Oh yes.

MEMBER AGORATUS: -- and that lack -- yes.

DR. LANDRIGAN: Yes, and again, I agree with that comment, that those are certainly an important population to consider as well. It really wasn't quite within the scope of what we were trying to do here, but I certainly agree with you that that is an important patient safety arena as well.

CO-CHAIR BROOKEY: John?

MEMBER BOST: James.

CO-CHAIR BROOKEY: I am sorry, Jim.

MEMBER BOST: Obviously, most hospital systems have a self-reporting like RL Solutions or something like that for adverse events. Was that data consulted at all when doing any of these reliability or validity assessments?

DR. LANDRIGAN: Not -- not for this specific study, no. There have been a number of studies in the past that have tried to cross-

check those things against one another, and they have found that those types of voluntary reporting systems really only pick up a fraction of what this tool picks up, something in the ballpark of 2 to 10 percent depending on which study you look at.

In fact, there was one study by David Classen that I am thinking of now that I think it picked up 1 percent of what their trigger tool picked up, so we really do think that this is a much more robust way of capturing that. Not to say that those systems are not valuable; I think that what those systems provide is a really detailed focused insight I think into the nature of -- of harms that occur in hospitals and help to focus institutional attention on trying to prevent these things, but if you are using it as a measure to try to track epidemiologically over time the frequency with which adverse events are occurring, I think this is much stronger than that sort of approach.

MEMBER BOST: Yes. I mean, I agree.

I was just wondering, just in the sense of -- just in the sense that -- did they maybe happen to just pick up any that the testing process didn't?

DR. STOCKWELL: So I had an earlier study that predated the GAPPS work where we did look at that. A very common number that seems to come out in the adult and pediatric literature for the overlap is about 8 to 10 percent, and the six hospital study that -- that we performed, then the overlap was 9. And so that -- as Chris is saying, I think that the way to -- to look at these two approaches to harm identification is that what we are presenting today is patient safety outcomes, and many of the voluntary admit reporting events are near-miss events, so have not reached the patient yet.

You need both. I mean, as a -- for operations in a hospital, you need -- you need both so that you can get those two lenses on patient safety, but we feel like this is the -- the more robust, as Chris said, more reliable way

to measure patient safety outcomes.

CO-CHAIR BROOKEY: Any other questions about reliability before we go to vote?

(No audible response.)

CO-CHAIR BROOKEY: Are we ready to vote? I wanted Suzanne to talk a little bit about -- going to your algorithm for evidence, this can only get a moderate, and so that is why you can't vote for high, and I want her just to explain that for us for this and future measures.

MS. THEBERGE: Sure. So for this, the reliability testing is at the data-element level only, which, according to our algorithm, the highest eligible is moderate. To be eligible for a high, we would need the measure score testing as well.

CO-CHAIR BROOKEY: Okay. Should we vote?

MS. JUNG: Voting for Measure 3136 for reliability is now open. Option 1, moderate; option 2, low; and option 3, insufficient.

(Pause.)

MS. JUNG: Voting is now closed. The results are 73 percent for moderate, with 16 votes; 27 percent for low, with six votes; and 0 percent for insufficient, with zero votes. The measure has passed for the reliability criterion.

CO-CHAIR BROOKEY: Great. We can move on to validity. Either Jill or Karen want to volunteer any comments about validity?

MEMBER HARPSTER: I think some of the comments that I had we discussed already. So, you know, one of the comments I had was about the lower sensitivity which we had already talked about.

They go on to talk about how we don't have a gold standard that exists, so we don't have that to use for that. But, they do talk about the base validity with the RAND methodology which they have discussed with us here today.

Other than that, I'm trying to see if I have any other comments here. I feel like we talked about a lot of the comments that I had on validity.

CO-CHAIR BROOKEY: Yes, I think we kind of blended these two together a little bit.

MEMBER HARPSTER: Yes. So, I think we -- and the length of stay of 24 hours, we kind of talked about already, too.

CO-CHAIR BROOKEY: Yes.

MEMBER HARPSTER: I think we're good as far as all the comments I had.

CO-CHAIR BROOKEY: Jill?

MEMBER MORROW GORTON: I agree, I think we've covered a lot of this.

CO-CHAIR BROOKEY: I think we did.

Jeff?

CO-CHAIR SUSMAN: So, this is a question for the developers. It seems that this tool, I'm sure will be helpful. But, that you have equated relatively minor events with some things that are relatively major or egregious.

And, to me, that's a little threat to your face validity as a clinician. And, thinking about some of the apples and oranges sort of gives you a fruit cocktail.

So, when you're reporting rates at an institutional level, I mean, maybe people are dying like flies at one institution and they're, you know, electrolyte imbalances or something in another.

DR. LANDRIGAN: Sure. And, so, we do have a means of rating the severity of these events as well. We use the NCC MERP scale where these things are put on this fairly standardized now five-point scale that gets you right from the electrolyte disturbances up to preventable death.

And, as really a submeasure of this, those would be reported as well. I agree with you, I think it's very important.

DR. STOCKWELL: And, the other point I might make is that the focus of the measure is decidedly preventable adverse events.

And, so, as is often the case, when we discuss patient safety or patient harm from those -- from the patient perspective, from the family perspective, if -- those more serious events can be just as unsettling to the patient or the

family as those that may resolve.

And, so, the scope of all of those events is something that we feel like we should be able to track and hopefully identify, improve and, at one point, eliminate.

CO-CHAIR SUSMAN: Yes, fair enough.

CO-CHAIR BROOKEY: I think that -- I mean, it goes -- it may get into the usability or not so much a feasibility but the usability of the measures in terms of any unintended consequences of trying to compare hospitals because the rates may or may not truly be telling the same story.

And, I think we may have to put that aside if we think there's still benefit for trying to get this to happen in hospitals.

And, I think we have to understand that that's going to be a threat. But, I guess we would have to deal with it like we do with any other measure that it's -- especially if it becomes publically reported.

I think we all intuitively understand

that this is going to be challenging to compare hospitals, even with a really solid measure.

CO-CHAIR SUSMAN: Can I just ask one other question?

When you looked at your community hospitals and the academic referral centers, it appeared that there were greater rates, if I remember correctly, at your academic centers, as you sort of expect.

And, do you have any way that you've corrected for severity of illness de novo that might help tell more of a story about rates of adverse events?

DR. LANDRIGAN: Well, when we did our initial analyses, we did adjust for the presence of chronic complex conditions. So, you know, adjusting in a sense for chronicity of illness and longstanding problems.

And, while certainly, those things occurred more frequently in the academic medical center, it did not, by any stretch, explain all the differences between the academic and the

community.

I image a lot of that is due to the fact that the academic centers are receiving the sickest patients from the community hospitals and so there's a real disparity there in the types of patients that they're seeing and that that really is an apples and oranges type of a comparison.

Our recommendation at the end was that, really, community hospitals and academic hospitals should, for purposes of comparison, be stratified. They really are sort of in different buckets and I don't think at some level it's appropriate to compare, you know, a large children's hospital with a smaller community center that is seeing a very different population of patients.

CO-CHAIR SUSMAN: Okay. So, you do recommend stratification?

DR. LANDRIGAN: Yes.

CO-CHAIR SUSMAN: That's probably -- okay, thanks.

CO-CHAIR BROOKEY: Any other comments

about validity before we go to vote? Okay.

MS. JUNG: As with the previous criterion, for validity testing this is only the highest eligible rating is for moderate, just to note, since it is a patient-level data element.

The measure -- the voting for Measure 3136 is now open for validity testing. Option 1, moderate; option 2, low; and, option 3, insufficient.

Voting is now closed. The results are 73 percent with moderate with 16 votes, 6 percent -- I mean, 27 percent for low with 6 votes and 0 percent for insufficient with 0 votes.

The measure has passed for validity.

CO-CHAIR BROOKEY: Okay, moving on to feasibility, I think that we clarified one concern we all had about the number of physicians that would be involved in the review.

But, they're secondary reviewers, so the first level of review are by RNs and then the follow-ups would go to clinicians or physicians in the hospital, which can present challenges

unless they're staff physicians paid for by the hospital or for the medical group or whatever.

So, that would be the one comment about feasibility.

The other issue would be in terms of having electronic system, maybe SPOE or, hopefully, most hospitals are going electronic and so they would be able to use electronic triggers.

But, that certainly would be a threat if that wasn't available.

Jill, do you want to go ahead and make your comments about feasibility?

MEMBER MORROW-GORTON: Yes, and I think I talked about this earlier when we were talking about evidence, so I probably put it in the wrong place.

But, just kind of looking at what was available in some places and not in others, what was in an electronic record, what was not in the electronic record, what was able to be gained by a manual review.

But, I think you've really touched on that.

Can you give us a little sense of the cost, both time and money, related to training the reviewers, both the -- because you would have to train both your nurse reviewers and the physician reviewers?

DR. LANDRIGAN: Sure. So, the -- in terms of the training, the program that we have is really, it's five one-hour videos that we ask folks to go through that requires a little bit of homework in between, where, particularly, towards the end of that video series, they're going into their own real charts extracting, you know, practicing extracting some cases and then coming back to the training with, you know, really the opportunity to vet questions amongst themselves and sort of think through how these types of events should be rated.

So, it's -- I think it's not -- it's certainly not cost-free in terms of time, but I don't think it's enormously burdensome. And, the

hospitals that participated in this did not find it to be too difficult.

Do you want to add to that, David?

DR. STOCKWELL: I do. Just, and also, to tie in a little bit about the physician piece. I think that -- I appreciate that we've been able to clarify that, because there was clearly some confusion around how much intensive physician work there was to do this.

And, if this relied on physicians digging around in charts, we all know this wouldn't go anywhere. Right?

DR. LANDRIGAN: Right.

(Laughter.)

DR. STOCKWELL: So --

DR. LANDRIGAN: Unless we pay them.

DR. STOCKWELL: All the nurses in the room are like, yes, I know that's right.

(Laughter.)

DR. STOCKWELL: So, this time spent by a physician by reviewing these events is incredibly valuable. And, I'll tell you this

from experience at my own organization, when I was leading safety and quality at Children's National, I was often the physician authenticator for this process.

That process itself was some of the most valuable time that I ever spent learning about how the practice of medicine was executed at our organization.

And, so, we've estimated four hours a year at a minimum of physician time. But, I would argue that that time is some of the best that you can spend of learning where your vulnerabilities are. Where are the problems that the organization needs to address?

And, so, I think it's incredibly valuable time and, as we've constructed it, it's actually a minimal impact on physicians as well.

MEMBER MORROW-GORTON: Do you offer CME and CEUs?

(Laughter.)

DR. LANDRIGAN: That's a great idea.

DR. STOCKWELL: That's a good idea.

DR. LANDRIGAN: That's a great idea.

(OFF MICROPHONE COMMENT AND LAUGHTER)

CO-CHAIR BROOKEY: Do you have any comments for --

MEMBER HARPSTER: The only other comment or question I might have is that, you know have these videos and so, you know, hopefully, that after watching the videos they're really good at looking for what they need to look for.

But, do you have some kind of process that's set inside if you have a reviewer that is maybe not where you want them to be after your training period and what that looks like?

DR. LANDRIGAN: Yes, and I think it's a great question.

In the study itself, we didn't have really any problems with our reviewers and with the training videos and a little bit of coaching as they were getting started, things worked out fine.

I think that, as this rolls forward

going forward, I do think it would be important for institutions to keep an eye on how their reviewers are performing.

We typically recommend that they hire at least two people to fill this role and they each have a bit of an opportunity to do both the primary reviews as well as the secondary reviews, which obviously allows for some diffusion of work, but also gives us the chance to make sure that they're well-calibrated with each other and that one person is not problematic, for example.

DR. STOCKWELL: The other point that may be useful to bring up is that, in addition to the videos, we have standardized charts that really have guides that walk you through the process and, if you miss this, this is what you missed and the nature of the event itself and describing those sort of things.

And, we recognize that there will certainly be flux in the reviewers' core team from -- for any organization. So, keeping that calibration, as Chris has mentioned, is, I think,

a very important aspect of it.

CO-CHAIR BROOKEY: So, tagging on to Marlene's earlier question, you know, my experience is that not all hospital electronic health records are the same. And, things that are very important may be missing in terms of like, for instance, we had an issue for a period of time where ABGs were not connected.

So, there was not -- we could not create a concurrent review of triggers based on ABG until we had that integrated into the system.

And, so, I guess my question is, is this going to be highly reliant on integrated, very sophisticated systems or is it going to be different from hospital to hospital in terms of being able to utilize the tool?

DR. LANDRIGAN: Yes, you know, it doesn't have to be. Obviously, if you have a well-coordinated, easy to read electronic health system, it makes life easier for the nurse reviewers.

But, this is really designed around

the idea that you could just be going through an old paper record where, retrospectively, everything is printed out and stuck into a chart to make it work.

CO-CHAIR BROOKEY: Okay.

DR. LANDRIGAN: It speeds things if you don't have to do that, but it works regardless.

CO-CHAIR BROOKEY: Okay. So, I think that we talked about feasibility. I think we understand it requires people.

So, I mean, when we talk about -- that's usually what we're talking about. It's not something that we can just pull out of the system, administrative data.

So, I think we all recognize that. And, so, I think the question will be whether the group feels this is -- meets feasibility. And, I guess we can rate it high, moderate, low or insufficient.

And, David, you wanted to make a comment before we vote?

MEMBER KELLER: Only the, you know, all of this discussion of relying on the -- on institutional integrity brings to mind several learning collaboratives that we've been involved in where we've found ourselves to be holding ourselves to a much higher standard than our peers.

And, I'm wondering if you considered the unintended -- potential unintended consequences of, well, I think the process that you're describing is exactly the right one that should be undertaken within an institution by making it a measure.

And, you know, are we setting ourselves up for an unintended consequence of some folks gaming the system and then not having adequate external controls to be able to look for that kind of variation in implementation when we take this sort of a measure out into the field?

DR. LANDRIGAN: Yes, I think it's an important point. And, depending on where this goes in terms of public reporting and so forth, I

mean, I think there may very well be a role for having occasional random audits of institutions by external reviewers as well as the internal folks.

We certainly, even in the context of study, use that to make sure that we had good reliability within the institutions and what they were detecting and reporting and so forth.

And, you know, I think if there's fear as time goes by about gaming the system here and there's consequences that are associated with having certain rates of reported adverse events using this measure, then I would think that some type of a check might be important.

CO-CHAIR BROOKEY: Okay, I think we all understand the feasibility of this measure. Are we ready to vote?

(NO RESPONSE)

CO-CHAIR BROOKEY: Okay.

MS. JUNG: Voting on feasibility for Measure 3136 is now open. Option 1, high; option 2, moderate; option 3, low; and, option 4,

insufficient.

Voting is now closed. The results are 5 percent for high with 1 vote, 68 percent for moderate with 15 votes, 23 percent for low with 5 votes and 5 percent for insufficient with 1 vote.

The measure has passed the criterion for feasibility.

CO-CHAIR BROOKEY: Great. So, usability follows feasibility and I think we kind of touched on some of those issues including unintended consequences if this gets out as a publically reported measure.

I think that we have also talked about this is probably an important thing to do to look at potential harm in pediatric population.

So, I'm going to ask either Jill or Karen if you have any specific comments about usability?

MEMBER HARPSTER: I think we touched on them. I guess the only thing I would say is that, you know, it's not currently being used now. So, they mention that, you know --

CO-CHAIR BROOKEY: For children?

MEMBER HARPSTER: For children, exactly.

But, that there's a potential for it to be used.

MEMBER MORROW GORTON: So, I had two questions. One is, what's the experience of adult hospitals using this?

And, the second question is, and you talked a little bit about differences between hospitals, but my question would be, is this as or more useful as an internal quality improvement measure versus being compared across hospitals?

DR. STOCKWELL: So, my review of the adult literature and just from some of the experiences that I've had in working with David Classen and understanding where many adult organizations have been, there are -- there is extensive use of, obviously, not gaps, but the IHI Global Trigger Tool itself.

And, there is not necessarily national use of the results of that, but there are many

organizations, large hospital organizations that will compare hospital to hospital within those organizations and they, by their accord, have found that quite useful.

And, so, I think that the same thing could be extrapolated to the pediatric environment with the caveats that we've already discussed.

DR. LANDRIGAN: And, with respect to your second question about its use as a benchmarking tool versus a quality improvement measurement, I think it certainly fulfills both roles.

I think that most institutions that began using early versions of this were primarily using it for quality improvement purposes. And, I think that's kind of the history of this in many respects.

But, you know, now, as we've begun to demonstrate some reliability and validity and so forth to the measure --- excuse me --- you know, our hope is that it could be used in a broader

sense as well for hospitals to benchmark against one another.

CO-CHAIR BROOKEY: Okay, Jeff?

MEMBER SCHIFF: I just want to speak to this from the point of view of a state program because I don't think we could ever use this successfully at the state program without getting skewered.

(LAUGHTER)

MEMBER SCHIFF: No, I mean, I'm just being really honest about it. I think that this is exactly what needs to happen at the hospital level for the basis of quality improvement at the hospital.

And, what we could do at the state program, we do this sometime is insist that institutions put themselves through these processes and we just have to, at the state program, know that the process is happening.

So, in terms of usability, this is probably more a comment for NQF than for the developers is, I think there needs to be a place

for measures that we can endorse that are usable at the -- as a process measure that we should understand are being done so we, at the state level or the national level endorse them as confidential process measures or quality improvement measures that we don't ever ask for the results of but we know that they're actually happening.

CO-CHAIR BROOKEY: Yes, your point it's exactly mine that I'm more interested in this hospitals doing this rather than reporting it. That's why I struggle with this as a measure because I'm much more interested in the process because of the potential harm that could happen if these results are shared publicly without really good interpretation.

And that's the issue that I have in terms of usability. But I would love to see this happen at every hospital that admits children.

Amy?

MEMBER HOUTROW: I basically have that same concern, especially with overlay. So,

there are certain health systems that are really invested in doing exactly this sort of thing in potentially a very robust way.

And, then there are high -- hold themselves a very high standard. Then there are others for which the systems they have in place would be very difficult to actually use this.

And so, we have a spectrum of potential usability and I'd like to ask the developers about the overly issue and duplication that might occur for some systems that are already pretty advanced in their --

CO-CHAIR BROOKEY: And I'm going to ask for a brief discussion because we're running a little bit over time on this topic. Okay?

DR. LANDRIGAN: You know, sure. I think that certainly is the case that, if you can do this in an automated fashion, it becomes a little bit easier for the institution.

But, we really did design this with the notion that it should be able to work in hospitals that have minimal resources along those

lines and minimal electronic health records to try to make it as easy as possible.

CO-CHAIR SUSMAN: So, it's sort of like rape reporting on campus, those that are really diligent are punished because their rates are higher. And, you know, it may be an apt analogy in some ways.

So, this idea of dissociating accountability from process measures, I think, is really important for NQF to get a hold of. So, I'm talking to this corner of the room because I think, frankly, there are many measures that are ready, just wonderful for improving internal processes and identifying issues quality improvement, but really aren't ready for prime time at a national accountability level.

CO-CHAIR BROOKEY: Jim?

MEMBER BOST: I just want to make sure that somewhere in they are you're really saying it's for system improvement and not any type of punitive use.

CO-CHAIR BROOKEY: Yes, I think that's

what we're talking about.

You can go ahead.

MEMBER MILLER: This is Marlene, could I chime in?

CO-CHAIR BROOKEY: Go ahead, Marlene.

MEMBER MILLER: I agree with this whole conversation and I just have had experience of many other measures that get put forth saying they're not ready for accountability and we don't think they should be used there and then they still get used there.

And, so, I just have a lot of trepidation because there are no safeguards, no processes to prevent that from happening and using measures that aren't ready for accountability in such a punitive way.

CO-CHAIR BROOKEY: Go ahead.

DR. AGRAWAL: So, very quickly, I think you've tapped into a conversation that we're having actively right now in the background.

I certainly, in discussions with our

stakeholders, see this as a gap or a need. Lisa and Marcia are well aware of this.

I think we have to think through what that means for, you know, our -- the public agencies that are involved in the work, what means for the evidentiary base behind these measures.

Because, of course, we also know there are gaps in measures with an extensive evidentiary base that we want to be able to elevate for accountability purposes.

So, if we're not doing that, I think we have to be really thoughtful about it. But, you know, in the spirit of being responsive, I do think this is a really important topic.

DR. STOCKWELL: Can I clarify one thing? I think that the rape analogy is an interesting one, although, I think it's more akin -- the rape reporting is more akin to voluntary reporting.

And, the vulnerabilities that are -- that we've talked about with voluntary reporting.

What we're trying to present to you is a more systematic, reliable process to review patient records in a way that you can actually decrease a lot of that variability from hospital to hospital in terms of just the reporting method itself.

CO-CHAIR SUSMAN: It's really the element of gaming and assiduous review along the specified pathway that you have defined. I mean, I think your measure is really quite robust in many ways.

CO-CHAIR BROOKEY: Thank you.

I'm going to try to move us along. I think that we have good conversation here. This is not a must pass element, usability.

But, I think that when you're voting for usability and overall, individually, you'll have to consider all of these points about accountability, and, so, that'll influence your vote for the overall voting for this measure.

Are there any other points about usability that haven't been made already or

should we move to vote?

(NO RESPONSE)

CO-CHAIR BROOKEY: Vote? Okay, so, we're voting for usability.

MS. JUNG: Voting for Measure 3136 for usability and use is now open. Option 1, high; option 2, moderate; option 3, low; and, option 4, insufficient information.

Voting is now closed. The results are 0 percent with 0 votes for high, 45 percent for moderate with 10 votes, 50 percent for low with 11 votes and 5 percent for insufficient with one vote.

The measure has not met the criterion for use and usability. Oh, consensus not reached, apologies.

CO-CHAIR BROOKEY: Okay. Are there any other points before we go to overall?

(NO RESPONSE)

CO-CHAIR BROOKEY: Okay, so, we're going to vote for overall suitability for endorsement.

Oh, I'm sorry, Carol.

MEMBER STANLEY: Just one minor correction. I was wondering --

CO-CHAIR BROOKEY: Can you use your mic?

MEMBER STANLEY: I was wondering, I noticed you have no exclusions and that comes during the process. But, I think one question that would come up is, why aren't patients who are transferred from another facility excluded, you know, if they're admitted with a UTI already from another institution?

DR. LANDRIGAN: Sure.

So, because that would get into, if this were to go to public reporting, those institutions that get a lot of transferred patients would essentially be punished.

DR. LANDRIGAN: Right. Yes, we do, again, as a submeasure of this, track which of these preventable adverse occurred within the studied facility as opposed to having come in from outside. So, it's easy to exclude that for

measure purposes.

CO-CHAIR BROOKEY: Jeff?

MEMBER SCHIFF: I just want to be clear of the question, if we're voting for suitability for endorsement, is that under the overall open accountability framework of NQF?

CO-CHAIR BROOKEY: Yes.

MEMBER SCHIFF: Okay. Can we abstain?

(LAUGHTER)

MEMBER SCHIFF: I hate to cause a rebellion, but I just think it's really -- I don't want to disrespect the developer, but I don't think this is appropriate for that kind of measurement. So, I guess the answer would be we should vote against it.

DR. NISHIMI: You are voting for accountability.

MEMBER SCHIFF: Okay.

CO-CHAIR BROOKEY: So, if that is a concern, you would vote no.

MEMBER SCHIFF: Okay.

CO-CHAIR BROOKEY: Okay?

Are we all clear what we're voting for?

(NO RESPONSE)

CO-CHAIR BROOKEY: Okay. So, let's vote for overall suitability.

MS. JUNG: Voting for overall suitability for Measure 3136 is now open. Option 1 is yes; option 2 is no.

Voting is now closed. The results are 64 percent with 14 votes for yes, 36 percent with 8 votes for no.

The measure is recommended for NQF endorsement.

CO-CHAIR BROOKEY: Great.

Well, I think we're warmed up now, so that was a great discussion.

(LAUGHTER)

CO-CHAIR BROOKEY: Congratulations.

(APPLAUSE)

CO-CHAIR BROOKEY: Thank you.

So, we're going to take a break for how long?

MS. THEBERGE: Fifteen minutes.

CO-CHAIR BROOKEY: Fifteen minutes, be back in your seat in 15 minutes. Thank you.

MS. THEBERGE: 10:45 folks.

(Whereupon, the above-entitled matter went off the record at 10:30 a.m. and resumed at 10:46 a.m.)

CO-CHAIR BROOKEY: Okay, we need to keep on schedule, so if I could ask everyone to return to their seat and we'll get started.

And, Rita, do you want to introduce yourself and your team member and we'll get started?

DR. MANGIONE-SMITH: Hi, I'm Rita Mangione-Smith and I am the Director of the Center of Excellence on Quality of Care Measures for Children with Complex Needs at Seattle Children's Research Institute.

And, this is one of our Center members and research team members, Kim Arthur.

MS. ARTHUR: So, I will be presenting Measure 3153, Continuity of Primary Care for

Children with Medical Complexity. Good morning.

We began our measure development process by creating a conceptual framework to identify care coordination processes that are related to short and long-term health outcomes for children with medical complexity.

Continuity of care emerged as foundational to care coordination because a continuous relationship with a single primary care provider or a small group of providers with in depth knowledge of a patient and family's needs could potentially increase efficiencies in care coordination, reduce avoidable utilization and improve health outcomes.

We conducted a literature review of continuity of care, defining continuity as the extent to which a patient's visits are concentrated in a single provider or a small group of providers. And, we were focused on primary care, to clarify that.

We found that the association between continuity of care and better outcomes has been

observed in multiple pediatric studies.

Better outcomes include lower ED utilization, lower risk of ambulatory care sensitive hospitalizations and greater primary care provider involvement and care coordination activities such as communication with other providers.

All of these studies use the Bice-Boxerman Continuity of Care Index which I will refer to today as the Bice-Boxerman COC Index.

This Index has multiple advantages over other measures of continuity of care. First, it is sensitive to continuity with a small group of providers if a patient sees a few providers frequently.

Whereas, other existing measures assess continuity with only one main provider.

Second, it uses administrative data. The Bice-Boxerman COC Index is, therefore, more feasible than parent reported measures and is not subject to recall bias.

All of the studies from our literature

review were Level II evidence according to the Oxford Center for Evidence-Based Medicine Levels.

We presented our literature review to a multistakeholder Delphi panel of nine panelists which included representatives of state Medicaid agencies and caregivers of children with medical complexity, among others.

The continuity of care measure we're reviewing today met face validity criteria. So, it was then operationalized and it underwent field testing in a sample of nearly 1,500 caregivers.

Those caregivers were children of medical -- with medical complexity and they were from two state Medicaid programs and that was using 2012 data.

We had compelling results at that time showing an association between higher continuity and lower emergency department utilization as well as greater likelihood of caregivers reporting that their child's care coordination needs had been met during the prior 12 months.

So, we then conducted the analyses that you have seen in our measure submission documentation to assess validity and reliability using 2008 Medicaid analytic extract for MACs data from 17 state Medicaid agencies.

We found that our measure had excellent reliability in this larger sample. And, in our empirical validity testing, we found that children who passed our COC quality measure had lower odds of having an emergency department visit which was consistent with our previous findings.

I look forward to discussing this measure with you today and addressing any questions you may have.

Thank you.

CO-CHAIR BROOKEY: Great, thank you.

And, this measure is a structure measure, if you're following your algorithm. And, I will ask, let's see here, we have Jill again, no, different, Marlene, I'm sorry. Marlene, you're on the line, do you want to go

ahead and kick off?

MEMBER MILLER: Sure, focusing on the evidence, I think this measure correctly cites seven different articles linking the continuity of care to better outcomes.

I think the one thing -- I don't think even beyond these seven articles, if you just talk about, you know, gestalt and wisdom and what makes sense, there is a lot of also logic there behind the more that you are connected, the more you are less likely to have other inadvertent outcomes like unnecessary ED visits, et cetera.

The more you are actually connected and counseled and made sure you're, you know, compliant on your therapies, et cetera.

So, I think it has a lot of evidence going for it.

In addition, I think the one thing that is interesting is that all of the evidence behind it of this particular tool which, it was interesting to me, that it was something developed back in 1977.

All the evidence of those seven articles uses this one tool. And, so, you know, it did raise some questions of, you know, are there other better ways, maybe easier ways?

You know, should we look at other evidence other than a lot of this research, not only using the one tool, but came from one group about this link on continuity of care.

And, I say that, while in the meantime, I'd say that, you know, I think even on a -- whether you'd call it evidence or just pure face validity, I think everyone agrees that this one should make sense.

I guess the one thing that gave me pause in the evidence was using this rather older published tool and that all of this cited evidence uses only this one tool and it's predominantly from one research group in terms of the breadth of the evidence on it with this particular tool from 1977.

CO-CHAIR BROOKEY: Okay, good points.

I should mention that Jeff Schiff is

recusing himself from this particular measure.

David, do you have comments?

MEMBER EINZIG: Yes, just a couple comments.

I agree that, on the surface level, this absolutely makes sense.

Just a couple of questions that I would raise, though, would be can it be a direct correlation that seeing one provider or a small group of providers, is that a direct correlation with improved patient outcomes?

And, then second thought is, if you're in the world of shared care and collaborative care now, if you're in a large pediatric practice where that one patient may not see that one or two providers, if they're in the same office or in the same location, they may not bill. But, if they're still present, that's not -- they may not get accounted for with looking at the claims data.

So, those --

CO-CHAIR BROOKEY: Do you want to put

that into a question for the developer?

MEMBER EINZIG: So, how would -- yes, so how would -- yes, so a question would be, direct correlation with improved care?

And, the second question is, how would that get picked up if they're not scheduled with that provider but that provider is still there?

CO-CHAIR BROOKEY: Go ahead.

MS. ARTHUR: Should I address the previous comments as well or --

CO-CHAIR BROOKEY: I would focus on his questions ---

MS. ARTHUR: --- start here?

CO-CHAIR BROOKEY: --- yes.

MS. ARTHUR: Okay. So, it is true that the studies we found were studies of association and not causation. But they were consistent across multiple studies.

And, in terms of, let's see, the second question about, if the provider was present, but it wasn't billed for that provider, it's true that this measure would not capture

that.

I think that would be a reason to be trying to make sure that the billing reflected the provider who was -- the providers who were there.

CO-CHAIR SUSMAN: So, I guess in follow up, some of the most advanced systems are providing continuity based on team, maybe a nurse or other care coordinator and at least, as I understand this index, it's really focused, isn't it, on physician or is there a broader definition?

MS. ARTHUR: We define primary care clinician as a physician, nurse practitioner or a PA.

CO-CHAIR SUSMAN: But, a nurse who might be very well qualified to do care coordination who's not an NP or advanced practice nurse wouldn't be included, is that correct?

MS. ARTHUR: Oh, a nurse practitioner would.

CO-CHAIR SUSMAN: Good, but one who is

not?

MS. ARTHUR: No.

CO-CHAIR SUSMAN: One who is not would not be included as a continuity provider?

CO-CHAIR BROOKEY: Or a social worker or a therapist?

MS. ARTHUR: This measure isn't looking at that continuity.

CO-CHAIR BROOKEY: Right, right, right, right.

CO-CHAIR SUSMAN: Right. So, I mean, again, I'm just thinking about, and maybe it's better discussed when we talk about validity, that this is a threat where some system, a Kaiser for example, that might have advanced systems of care might be poorly represented by this measure?

MS. ARTHUR: One thing to say is that that, if you're looking at social workers or nurses, a healthcare plan would not be unfairly, like would not be punished for that because that's not measured in this measure, like, we're not looking at that.

CO-CHAIR BROOKEY: Right. Well, I guess the question would be, so the evidence here, and we're talking about evidence right now, the evidence and all the research that has been done is looking at this relationship with these providers and not with the extended team.

So, we're looking here at a very limited range in terms of providers. But, that is what you're evidence is suggesting leads to better outcomes even though intuitively know that all these others are important as well. Is that a fair statement?

DR. MANGIONE-SMITH: I think part of what the measure is getting at is the idea of the quarterback, right? That you have a team, but you should have one consistent person who's coordinating not only care for that child but also the team that's caring for that child.

CO-CHAIR BROOKEY: Okay.

DR. MANGIONE-SMITH: So, I think like Kim was saying, you're not going to get penalized by having that team in place. What you'd be

penalized for is the child seeing somebody different every time they come in --

CO-CHAIR BROOKEY: Right.

DR. MANGIONE-SMITH: -- as a primary care provider.

The one other thing I wanted to add in relation to Marlene's question about the evidence, we did review evidence more broadly. We looked at several different types of measures of continuity of care that have been used and trialed.

We ended up staying with Bice-Boxerman because it does actually give you credit for seeing the same two people several times as well as giving credit for seeing the same one person every time.

CO-CHAIR BROOKEY: Right, right.

DR. MANGIONE-SMITH: The other measures that are available strictly look at your continuity with a single provider. So, you get no credit if there's like a team of providers, maybe they both work part-time and you always see

one of those two, you will have a higher score on Bice-Boxerman than somebody who is seeing those same two providers using one of the other measures.

CO-CHAIR BROOKEY: Okay, understood.

Tara, do you have any comments?

MEMBER BRISTOL-ROUSE: Yes, you know, I think one thing that stuck out for me is the issue of, you know, communication being such a vital part of continuity from certainly the parent perspective that I have.

So, I know that there's, you know, this relationship between the two. But, just in thinking about practical lived experience, I think, you know, so much of it is less about, I mean obviously, I would love to see the same providers all the time, but it's also, even if it's a different provider, if it's a provider who knows my child and is able to speak competently with me, like, that's just as important as having someone I've seen before, maybe even more important if it's someone I've seen before who's

not a great communicator or is not really engaged with my family.

So, just, I guess that point in terms of the communication piece of continuity.

CO-CHAIR BROOKEY: Good comment.

David?

MEMBER KELLER: I will apologize if I'm just reading this wrong, but I want to make sure I'm clear on the denominator, because, in the version I have, it says --

MS. ARTHUR: It's wrong.

MEMBER KELLER: Okay. So, could -- okay, good. Because, I read that like 15 times trying to figure out what I was missing.

MS. ARTHUR: Yes, it's oh boy, we're going to really trip people up with that.

MEMBER KELLER: Okay, so, it's supposed to be the most complex kids, not the other two categories with the lesser --

MS. ARTHUR: Right.

MEMBER KELLER: -- two? Okay. Thank you for that.

CO-CHAIR BROOKEY: That's why you're on the group, Jeff, very good.

So, Jill?

MEMBER MORROW-GORTON: Did I read right that they have to have five or more visits, sort of more than --- four or more? And, I guess my question is, where did you come up with that number?

And, in thinking about kids that have lots of complexities, if they don't need -- so, adding more visits just for the purpose of getting credit for this would not be -- would not necessarily be a quality thing to do when you're talking about primary care visits. So, just a question of where that all fits?

MS. ARTHUR: Thank you for the question.

So, the minimum of four visits was based on the research that was done by Christakis, et al, and he found that the Bice-Boxerman Index was less stable if you had fewer visits.

So, just like with any calculation, if you have a small denominator, whatever you calculate is going to be -- not going to be stable and it's going to be highly variable. So, that's where the four visits came from.

Also, I think it's important to think about, if you think about this population of children with medical complexity and you establish that four visit minimum, these are the children who are going in to primary care frequently compared to others and there's more opportunity to build that continuous relationship with those children who are going in more frequently.

And, I'll also say that, in our sample, the mean number of primary care provider visits was 4.63. So, the kids in our sample were coming in just over four visits and this was kids age 1 to 17, so across the age range, that was what our mean was if that helps give a sense.

CO-CHAIR BROOKEY: And, that was sort of stable across the age range, even with the

older?

MS. ARTHUR: No, older kids had fewer.

CO-CHAIR BROOKEY: Yes, yes.

MS. ARTHUR: Yes.

CO-CHAIR BROOKEY: Okay. I thought I saw somebody, oh, go ahead, Deborah.

MEMBER FATTORI: I'm just curious, as you looked at your evidence, did you look at generational differences with parents?

One of the things that we're struggling with, maybe struggling is too strong a word, but certainly being thoughtful about is, in our millennial parents, where seeing the same provider isn't as important to them as seeing a provider when they want to.

And, I'm just curious how that -- how this measure will impact that as long, I think to your point, Tara, as the communication is good, if they can use telehealth, if they can come in an evening or see the provider when they want, it might not be as important to them that they see the same provider.

MS. ARTHUR: That's an interesting question and I can't say that the evidence that we were looking at looked at generations. So, is there anything else I can address for you?

MEMBER FATTORI: No, I was just curious if that came up. It is one of the things that we're seeing more and more with our younger parents.

MS. ARTHUR: Yes.

MEMBER FATTORI: And, needing to meet the needs of those parents in our primary care settings to -- as far as when they want to be seen and have their children seen.

CO-CHAIR BROOKEY: Yes, so coming back to this particular measure, we're looking at the evidence to support this approach to continuity of care with better outcomes to refocus our discussion for future vote.

So, David?

MEMBER KELLER: I'm sorry, the other question that I had which I forgot because I was so obsessed with that other thing --

(LAUGHTER)

MEMBER KELLER: -- I was sure it was -- I was reading it wrong.

Is the question of primary care visits, and how do you -- how to identify primary care visits? At least in our state Medicaid agency, we are -- state Medicaid database, we are unable to differentiate a subspecialty visit from a primary care visit within our system because we bill a single provider code.

So, I think that would lead to some challenges if you were trying to apply it in our state.

MS. ARTHUR: Sure. And, it's an excellent point. We had the good fortune to have the data programming and analysis people that work with Jeff sitting next to you in development of the specifications as well as the Washington State Medicaid data and analysis people.

And, in working with them, we tried to come up with this comprehensive list of codes as we could to capture these visits appropriately.

And, by doing the combination of the NPI, the provider identifier and the place of service, you can be pretty confident that it's primary care because providers are, by their NPI, there are ways to get at whether they're subspecialty or primary care.

So, by limiting the type of provider by NPI that we allowed in, we were hoping that we would avoid some of the problem that you're alluding to in terms of if you were just to use E and M codes for the type of visit.

MEMBER KELLER: Right, I mean, and I would have thought that would have been easy, too. But, I've been told --

MS. ARTHUR: It's not easy.

MEMBER KELLER: -- because I was going to say --

MS. ARTHUR: And, we would never purport that it's easy.

MEMBER KELLER: Okay.

MS. ARTHUS: But, it is --

MEMBER KELLER: Good, because, yes.

MS. ARTHUR: We found that it is doable.

MEMBER KELLER: It's doable?

MS. ARTHUR: Yes.

MEMBER KELLER: Our state Medicaid agencies struggles with that and we struggle with them around that.

MEMBER AGORATUS: This is Lauren, just to tag team on that, I'm not sure if there's a way to capture if the primary care physician is a specialist, for example, for a transplant patient. Your PCP might be the nephrologist.

DR. MANGIONE-SMITH: Right. So, this measure would not look at subspecialty medical homes, it would only look at primary care medical homes.

And, so, if it were a child who is getting most of their care including their primary are from a subspecialist, which certainly does happen, they would very likely not be included in this measure because they wouldn't meet the threshold of having had four primary

care visits in a measurement year.

MEMBER AGORATUS: Thank you for clarifying.

DR. MANGIONE-SMITH: Sure.

CO-CHAIR BROOKEY: Okay. So, we have a number of questions and comments and, remember, we're talking about evidence for the measure. So, I'm going to --

Marlene, did you put your flag down or did you not have a comment?

MEMBER EDIGER: I found --

MEMBER MILLER: Oh, I didn't know it was up.

MEMBER EDIGER: Oh, sorry, it's Maureen.

CO-CHAIR BROOKEY: I'm sorry, Maureen.

MEMBER EDIGER: Similar names.

MEMBER MILLER: Sorry.

MEMBER EDIGER: That's all right.

My question was about emergency room visits and I think I found my answer in the report.

CO-CHAIR BROOKEY: Okay.

MEMBER MODAK: Just briefly to address the concern about team based care and about an RN care coordinator.

We actively use those for our medically complex patients. And, that has actually, I think it aligns well with this because we've found it increases continuity with the same provider or we have a dyad of providers who take care of patients in case one is not there.

CO-CHAIR BROOKEY: So, you don't have any concerns if they're not included?

MEMBER MODAK: I don't think so, I think it actually -- it's totally aligned with this, yes.

CO-CHAIR BROOKEY: Yes, okay.

Carol?

MEMBER STANLEY: Yes, I was wondering if you could speak to not having separate stratification for age categories since you're putting adolescents with the younger population

and the recommendation for primary care visits through AAP is vastly different for adolescents versus younger and how that bodes with the requirement of four or more visits?

DR. MANGIONE-SMITH: Right. So, there's not a requirement that you have four visits, we only put you in the denominator if you've had four visits.

So, if you're an adolescent who's not requiring much primary care, even though you're medically complex, you wouldn't even make it into the denominator for this measure.

That's why the majority of our population was in that younger age group because those are the kids who are coming in over and over again to primary care and those are the kids we want to be sure are having good continuity.

MS. ARTHUR: Right. I just was going to add, if you are an adolescent patient coming in at least four times, then we'd want to make sure you have continuity just like a younger patient is going to.

CO-CHAIR BROOKEY: Okay. I'm just trying to think through that just a little bit. So, you're trying to give better outcomes and you're probably looking at complex children, whether they're young or adolescents that are likely coming in more often.

So making them and having them come in four is probably reasonable to look at improved outcomes of decreased ED utilization and so forth. Is that right?

MS. ARTHUR: That's correct.

CO-CHAIR BROOKEY: Okay. Is everyone clear about what we're trying to measure? And, again, we're looking at the evidence for this measure and its relationship to outcome.

Are we ready to vote for evidence you think? Anybody on the phone have any other concerns before we move on?

DR. MANGIONE-SMITH: We did want to point out one thing in the review was that we did actually evaluate the quality of the evidence using the Oxford Center for Evidence-Based

Medicine criteria. It was one sentence in the evidence review and I think it just got missed.

CO-CHAIR BROOKEY: Yes.

DR. MANGIONE-SMITH: They were all Level II cohort studies.

CO-CHAIR BROOKEY: I think we're good. Okay, you want to vote?

MEMBER MILLER: Can you remind the scales?

MS. JUNG: Oops, sorry, hold on just one moment. No, we didn't start yet.

Voting for Measure 3153 is now open for evidence. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

Voting is now closed. The results are 14 percent for high with 3 votes, 77 percent for moderate with 17 votes, 9 percent for low with 2 votes and 0 percent for insufficient with 0 votes.

The evidence has passed -- or the measure has met the criteria for evidence.

CO-CHAIR BROOKEY: Great, so now we're going to move on to gap. And, what page am I on here? I don't see a discussion on gap. Here it is, okay.

So, David, do you want to start out?

MEMBER EINZIG: Gap exists, I don't think that there was much else to add from other committee members comments.

CO-CHAIR BROOKEY: Okay.

Marlene?

MEMBER MILLER: Yes, I agree, gap exists. I think one question I had is, there was a substantially large sample size, I think it was like over 11,000 cases.

And, so, there was a lot of significance, but I wonder if the developers can comment about the clinical significance?

So, for example, there was, you know, a different pass rates of 67 percent continuity of care versus 62 percent. And, is that sufficient? Is that clinically meaningful is what I was lost on because it's such a large

sample size, I was worried if we lost sight of clinical then there's just statistical significance?

DR. MANGIONE-SMITH: Hey, Marlene. Yes, so, I think looking at the pass rates by age range or some of the other criteria, the P values are certainly being influenced by the large sample size.

I think what was striking to us was, when you look at the range of scores across the states, the 23 percent to 96 percent indicating that there's huge variation in performance across state Medicaid plans that we looked at in the 17 different states.

That's where I think we're really seeing what would probably be considered more clinically significant differences in scores.

CO-CHAIR BROOKEY: Let me ask Tara to go next. Tara, do you have any comments?

MEMBER BRISTOL-ROUSE: No, nothing additional.

CO-CHAIR BROOKEY: Okay.

Jeff?

CO-CHAIR SUSMAN: It stretches my credulity that there's a 96 percent at a statewide level? I mean, I just -- I have almost -- is that right?

DR. MANGIONE-SMITH: Yes, it's right.

CO-CHAIR SUSMAN: Might just miss it on me. Because some of you work on the plan level, state level all the time. It just -- wow.

CO-CHAIR BROOKEY: Is that correlated to density and population or density of providers or --

DR. MANGIONE-SMITH: So, interestingly, the variation available sample size across the states was there was a big variance, but this was actually one of our states with a bigger sample size to draw from.

So, they had 450 eligible children. So, it was not that it as a tiny sample size and they just did really well on a few kids.

CO-CHAIR SUSMAN: Yes, no, just wow.

DR. MANGIONE-SMITH: You know, I guess

I could out them but I'm not going to.

CO-CHAIR SUSMAN: So, that's your secret sauce?

DR. MANGIONE-SMITH: Right.

CO-CHAIR BROOKEY: Let's see, Carol?

MEMBER KELLER: Actually, I do know that sample because there was a -- Massachusetts has a specific program for its sickest kids that's called CCM and my wife was the medical director for it during that period of time.

And, if that's the sample that they're drawing from, those kids have incredible continuity of care. So, I do believe.

CO-CHAIR BROOKEY: Okay, Carol?

MEMBER KELLER: She wasn't the one seeing them, but she supervised the program.

CO-CHAIR BROOKEY: Okay, let's see, Ricardo?

MEMBER QUINONEZ: Did you -- was there a correlation between that gap and primary care and the outcomes you're trying to prevent so that the states that have the lowest follow up have

the highest rates of emergency visits, for example?

DR. MANGIONE-SMITH: So, the analysis controlled for state. We did not look at ED use by state. But, in the overall analysis, the states that passed the measure that had higher pass rates had lower ED utilization.

MEMBER QUINONEZ: Okay.

CO-CHAIR BROOKEY: Jon?

MEMBER FINKELSTEIN: So, I, too, am the breadth of the variability by state makes me a little skeptical. So, because these are events that are happening at the provider group level being rolled up to a state and it's not -- it seemed implausible to me that one state has only wonderful provider groups, you would expect some regression to the mean as you combine different groups.

So, I just wonder, is there -- did you look for other kind of data quality explanations for this? So, people billing under a single NPI number in one clinic or any other things that

could have introduced funny things to make a state look better than it is or worse than it is?

DR. MANGIONE-SMITH: So, we deliberately included both states that had strictly fee-for-service, states that had a combination of managed care and fee-for-service. And, we know from working over the past five years with our Medicaid state partners in Washington and Minnesota that the managed care data is questionable at best many times and not complete.

We also know that the 2008 MACs data suffers terribly from incomplete managed care data. So, I would not be at all surprised if we looked at the states with lower performance, they probably have higher managed care penetration in terms of this population of children.

And, the ones that are mainly covering their kids with fee-for-service are probably the high performers because their data is much more complete.

So, I would say that that's probably

the biggest culprit here. We did try to use the states with Jeff Silber's help, we got a lot of information about which states have the higher quality data in the MACs data and those are the ones we tried to target. But, I still think we ran into this issue.

CO-CHAIR BROOKEY: So, at this point, we're just trying to decide if there's a gap. So, if there's other --

(LAUGHTER)

CO-CHAIR BROOKEY: -- important questions about whether we have a gap, then I invite them.

So, David?

(NO RESPONSE)

CO-CHAIR BROOKEY: Oh, okay.

So, is everyone decided whether we have a gap or not? Can we vote for gap?

(NO RESPONSE)

CO-CHAIR BROOKEY: Okay, all right.

MS. JUNG: The voting for Measure 3153 for gap is now open. Option 1, high; option 2,

moderate; option 3, low; and, option 4, insufficient.

The voting is now closed. The results are 41 percent for high with 9 votes, 59 percent for moderate with 13 votes, 0 percent for low with 0 votes and 0 percent for insufficient with 0 votes.

The measure has passed for performance gap.

CO-CHAIR BROOKEY: Okay, we can move on to reliability. And, can I get a volunteer from Marlene, David or Tara to go first?

(NO RESPONSE)

CO-CHAIR BROOKEY: Hearing none, I'll go with Marlene.

(LAUGHTER)

MEMBER MILLER: Great, I get the short straw, right?

I had no concerns about the reliability. They're pretty much all based on algorithms and they're well executed out the code.

What I wasn't clear on, and I think it's in there, but I just wanted to ask specifically was how much of this was done exclusively on ICD-10 which, obviously, we all know greatly expanded the coding.

I saw some wording there that there was ICD-10, but I just wanted to be sure that this was tested with ICD-10.

DR. MANGIONE-SMITH: So, the current testing that was done was done on the 2008 medical extract file which is all ICD-9. So, it has not been testing in ICD-10, although the algorithm to get the denominator, the pediatric medical complexity algorithm will be released, the ICD-9, ICD-10 combined version will be released at the end of this month.

So, and that has been tested using both ICD-9 and ICD-10 coded data and validated with both.

So, we're confident that, at least in terms of identifying the denominator population, that should not be an issue.

MEMBER MILLER: Okay, so I thought I was because I was -- that's why I was confused. I know is saw ICD-10 in some places but the bigger bulk of it has not been tested yet in 10?

DR. MANGIONE-SMITH: Yes, the testing in this document was all done with ICD-9, but we wanted to make clear that the ICD-10 version is on the brink of being available.

MEMBER MILLER: Okay.

CO-CHAIR BROOKEY: David or Tara, any comments?

(NO RESPONSE)

CO-CHAIR BROOKEY: I should note, I think it's six eligible individuals per state Medicaid agency for necessary, that's not very many.

DR. MANGIONE-SMITH: We're just .7.

CO-CHAIR BROOKEY: For --

DR. MANGIONE-SMITH: Yes, if you have up to 20 -- we recommend up to 25 because that pushes your reliability up to .9.

CO-CHAIR BROOKEY: Right, right.

DR. MANGIONE-SMITH: Yes.

CO-CHAIR BROOKEY: So, any concerns about reliability? I think this is clearly described here.

Jim?

MEMBER BOST: Maybe it's not a concern but from what I -- I see that it was only done on Medicaid, and yet, your unit is health plan. So, I guess just a little concerned that reliability wasn't assessed at commercials plans.

DR. MANGIONE-SMITH: Right. So, we only looked from state to state in terms of the Medicaid health plans, that's correct.

Many of the states did have managed care products, so there was some commercial health plan in there.

But, you're right, we did not look at commercial health plan -- compared to commercial health plan.

CO-CHAIR BROOKEY: So, just for my benefit, because I'm very California-centric, when you say Medicaid health plan, you're

including both fee-for-service and managed care?

DR. MANGIONE-SMITH: Yes.

CO-CHAIR BROOKEY: Okay. Okay. Any other questions or concerns?

MEMBER AGORATUS: Yes, this is one from Lauren.

In terms of the continuous coverage, what happens if they switch plans?

MS. ARTHUR: We --

MEMBER AGORATUS: Is that in the report?

MS. ARTHUR: Sure, thanks for the questions.

So, to clarify, we were looking at continuous enrollment in order to be eligible. And, continuous enrollment in a single plan and they could have no more than a 30-day gap.

And, so, if a child switched plans, they wouldn't be eligible.

MEMBER AGORATUS: Thank you.

CO-CHAIR BROOKEY: If they went from fee-for-service to managed care, they would not

be eligible.

MS. ARTHUR: No, that would be a plan switch.

CO-CHAIR BROOKEY: Okay.

Carol?

MEMBER STANLEY: So, a little while ago, you mentioned that disparity between the numbers, one of your hypothesis is that in managed care, the data's not being captured as well possibly as it is in fee-for-service.

And, so, with health plans being the level of analysis, why would there be -- why would you use claims and not use a hybrid methodology for this measure if you're finding that with health plans there may be data integrity issues?

DR. MANGIONE-SMITH: Right. So, let me clarify.

There are issues with the managed care data in the MACs file, which is what we used. Our understanding, at least from our Medicaid partners in Washington State, now that almost all

children are on Medicaid managed care, there's almost nobody on classical fee-for-service anymore, their data that they're getting from the health plans is far more complete than it used to be.

So, and the testing we did in our two states, the 1,500 that we did for part of our measure development, we saw, you know, very similar performance to what we're seeing here, but did not have -- we did not have the same concerns with data being absent in that run that we did in the MACs file.

So, I don't -- I do think these source health plans have the data that's needed to do this. But, the states, at least back in 2008 in the, you know MACs file, we're not getting complete data from them.

CO-CHAIR BROOKEY: Any other comments about reliability or should we vote?

(NO RESPONSE)

CO-CHAIR BROOKEY: Vote? Okay.

MS. JUNG: The voting for Measure 3153

for reliability is now open. Option 1, high; option 2, moderate; option 3, low; and, option 4, insufficient.

The voting is now closed. The results are 23 percent for high with 5 votes, 64 percent for moderate with 14 votes, 9 percent for low with 2 votes and 5 percent for insufficient with 1 vote.

The measure has passed for reliability.

CO-CHAIR BROOKEY: Okay, so let's move on to validity.

David, do you want to lead this one?

MEMBER EINZIG: If there are any other members who are more of a numbers person, that'd be great. Otherwise, it appears valid and I don't have much else to add.

MEMBER MILLER: This is Marlene, I can chime in, if that's all right.

Can you hear me?

CO-CHAIR BROOKEY: Yes, go ahead, Marlene.

MEMBER MILLER: Okay, great.

So, I think there's a lot of evidence presented both for empiric and face validity. I think the main concern I had with validity was the fact that the continuity of care with this scale all the ways you can measure and all the variability we just talked about and reliability, between states, for examples, they put down to a simple dichotomization of pass/fail.

And, I really worried about that having real validity in terms of -- because we know continuity of care is much more complex than that. It's either not a pass/fail, black or white.

So, I was wondering if the developers could comment on that in the methodology.

DR. MANGIONE-SMITH: Right, so, this score is pass/fail based on whether your COC index ranges from .5 or higher.

So, COC index can go from 0 to 1, all of the studies that we reviewed, you know, when they were looking at relationship to outcomes,

they were looking at different levels of the Bice-Boxerman Continuity of Care Index and how that related to outcomes.

The vast majority of them, in fact, I think all of them found that the better outcomes started happening right around .5.

So, we wanted to give credit to any case that was included in a state sample, in the Medicaid sample where that child's Bice-Boxerman Continuity of Care Index was .5 or higher rather than saying, you know, we want to know the continuous value or the mean value of Bice-Boxerman COC for your state.

We could have done it either way, but we felt that the evidence was really suggestive that there were these clear cut points where outcomes were better.

And, that's -- and it also, I think, makes it a more straightforward measure to interpret and potentially to improve on.

So, if you know that the bar you're trying to make is I just I need to get my

population at least up to .5 if I want to see improved outcomes, it's just a little bit more straightforward, we felt.

MS. ARTHUR: Also, I can add that the mean COC in our study in this analysis was .65 actually. And, in those previous studies that looked at improvements stating at the .4 or .5 five level, the mean COC was actually around .4 or .5.

DR. MANGIONE-SMITH: So, they're pretty similar.

CO-CHAIR BROOKEY: Carol? I meant Jon.

MEMBER FINKELSTEIN: We get confused all the time.

(LAUGHTER)

MEMBER FINKELSTEIN: So, in looking at your data on ER visits, the outs ratio was in the .9 range which is a 10 percent lower rate of people hitting the ED who have good continuity compared.

Were you happy with that? Was that the

magnitude you were -- as a guy who kind of philosophically believes in continuity, I was a little disappointed, but maybe that's the magnitude of effect you'd expect. I was just wondering about that.

DR. MANGIONE-SMITH: Yes, actually, I think it is the magnitude we would expect. There are very few interventions that can drag down ED rates by 10 percent.

So, if we could even get that far down, I think we'd be accomplishing something.

MS. ARTHUR: Also, there was a study that we hadn't mentioned today that looked at costs avoided with ED visits. And, they actually found, this was a study by McBurney and colleagues, and they found that increasing continuity of care by ten percent points yielded a decline in expected ED visits from around 1,300 -- 1,362 to 1,290. And, there was a cost savings of almost $20,000.

So, I mean, you'd love to see more, but it's meaningful.

CO-CHAIR BROOKEY: Well, there is one intervention that's effective, it's called co-payments. I don't think any of these states have co-payments, do they?

MS. ARTHUR: The grand health experiment.

CO-CHAIR BROOKEY: Yes.

Jill?

MEMBER MORROW-GORTON: So, they might have co-payments, depending on the state.

But, anyway, my question --

CO-CHAIR BROOKEY: This is true, there are some states that have Medicaid co-payments.

MEMBER MORROW-GORTON: Yes, there are a few that have --

CO-CHAIR BROOKEY: Yes.

MEMBER MORROW-GORTON: -- co-payments for various things, yes. It's like a dollar or $3 or something like that.

But, for a poor family, that could be meaningful.

(OFF MICROPHONE COMMENTS)

MEMBER MORROW-GORTON: Yes, so my question, as you looked at ED visits, and it could be that they are pushing not going to the ED, going to their primary care doc and then getting direct admitted from the primary care doc, which you would not capture in ED visits. Did you look at hospitalizations?

So, you've got ED visits savings, but you might, you know, you might not really have gotten any savings because we wouldn't have paid for that ED visit if they got admitted. We would have just paid for the admission.

DR. MANGIONE-SMITH: In this particular analysis, we did not look at hospitalizations. We felt we would have to look at ambulatory care sensitive hospitalizations. And, in pediatrics, at least, that is not -- they are not well defined.

There is only one study that we were able to find that had any outlining of codes that you might use for it.

And, the quality of the MACs data was

such that we just felt we couldn't adequately look at that outcome in that data. So, it's two incomes late.

CO-CHAIR BROOKEY: Any other questions about validity?

(NO RESPONSE)

CO-CHAIR BROOKEY: On the phone?

(NO RESPONSE)

CO-CHAIR BROOKEY: Okay, should we vote?

MS. JUNG: Voting for Measure 3153 for validity is now open. Option 1, high; option 2, moderate; option 3, low; and, option 4, insufficient.

Voting is now closed. The results are 5 percent for high with 1 vote, 77 percent for moderate with 17 votes, 18 percent for low with 4 votes and 0 percent for insufficient with 0 votes.

The measure has passed for validity.

CO-CHAIR BROOKEY: Okay, so for feasibility, let me just, before I turn it over

to the discussants, just to ask you to clarify a comment you made about getting data for managed care, and so, in terms of how it relates to feasibility. Is that going to be an issue from state to state?

DR. MANGIONE-SMITH: So, I think for state Medicaid agencies, as they have moved away from fee-for-service and towards Medicaid managed care, my understanding, at least from our partners, and I can only talk about my NF2 which is Minnesota and Washington, that it is getting better.

They are -- their data is more complete than it was, say, five years ago and they have much more confidence that it's complete.

Having said that, there will probably be states where there is not complete data. And, I, you know, I don't, unfortunately, have like, you know, fact-base to go on because I don't have direct contacts at all 50 states, so I don't know.

CO-CHAIR BROOKEY: Okay, David or Tara or Marlene, any comments about feasibility?

MEMBER MILLER: No, I mean, I had some concerns, I don't know whether -- where to kind of go with them that this would require anybody who wanted to do this to reach out to the developer's website and get the SAS code.

You know, it's not something that they could run independent of their own shop and then they always worry about accessibility and will that always be there kind of thing.

MEMBER EINZIG: It's electronic claims data, so it's feasible.

CO-CHAIR BROOKEY: Yes, it's feasible.

Tara, any comments?

MEMBER BRISTOL-ROUSE: No, I mean, obviously, one of the benefits is, you know, as opposed to parent report data is that it comes from claims data.

CO-CHAIR BROOKEY: Any other comments about feasibility? Go ahead.

MEMBER FEI: My question's probably

rather general, but any thought to the feasibility, but also being able to implement it within commercial claims?

I mean, I know, in a commercial PPO world, perhaps the volume is less, but we would be interested to know if we could actually take that SAS code and like apply it to our warehouse?

DR. MANGIONE-SMITH: Absolutely. So, one, actually, two of the studies were done in commercial data.

MEMBER FEI: Okay.

DR. MANGIONE-SMITH: One was done by Jeff Tom who actually used the exact same code that we used for this study in the Hawaii, I think it's Blue Cross, it's the -- they cover like 80 percent of the privately insured lives in Hawaii.

So, he ran his analysis and actually, interestingly, back -- the question about ambulatory care is sensitive hospitalizations, he was able to look at that because we have this very complete data and his was the one study that

did show that, as your COC went up, your ambulatory care sensitive hospitalizations went down and children with at least one chronic condition.

CO-CHAIR BROOKEY: Okay, another other comments or should we vote on feasibility?

(NO RESPONSE)

CO-CHAIR BROOKEY: Okay, we'll vote.

MS. JUNG: Voting for Measure 3153 for feasibility is now open. Option 1, high; option 2, moderate; option 3, low; and, option 4, insufficient.

Voting is now closed. The results are 36 percent for high with 8 votes, 64 percent for moderate with 14 votes, 0 percent for low with 0 votes and 0 percent for insufficient with 0 votes.

The measure has passed on feasibility.

CO-CHAIR BROOKEY: Okay, so then we'll move to usability and use. Any of the discussants want to lead?

David?

MEMBER EINZIG: So, just a few comments on usability and use. So, it says MAP imbedded, I'm curious, this might be in here, but a small group of providers, does that define what is the number for a small number of providers?

MS. ARTHUR: What we really were intending to say with that was that the Bice-Boxerman Index, as we mentioned before, is sensitive to detect continuity with more than one provider.

So, perhaps it wasn't clear that we define -- you know, we define small group. But, it really was just that concept that it's more than one and it can still detect continuity.

MEMBER EINZIG: And, then, just final comment, you know, I think parent perspective is really valuable here, too, that if we're using this -- planning to use this measure, I think it would be useful to get more parent and family feedback in terms of from the family perspective if it's ongoing -- of ongoing value in current day terms.

MS. ARTHUR: So, I'm actually wearing two hats here today. I am the proud mother of medically complex twins who are three, so I can't resist putting on that hat for just a minute to say that my -- in my experience, I've been doing pretty much all the care coordination. That's been my personal experience with my end of one hat on.

But, what I also will say is that I have very rarely seen the physician who is supposed to be my primary care provider because every time my girls are sick, we see whoever can see them right away or whoever can see them to follow up from the ED.

And, so, to sit here and see this measure and think, wow, if there was -- if it was a priority for me to actually see that person who really, truly does know our history, although now, it's been over a year since I've seen him, it would have made a difference.

So, I can just -- if you're asking for a parent and family opinion, I can't keep quiet.

DR. MANGIONE-SMITH: And, the other thing I'll add, the process that our center used, Carolyn Allshouse, who's the person who directs Family Voices of Minnesota, as part of our center.

And, she brought in a whole group of parents into our measure development process. So, every measure that's come out of our center was, the literature reviews are read by the parents, the measures were commented on by the parents.

And, we really took their feedback and quite, quite seriously. They were involved from the beginning to the end.

In fact, Carolyn's an author on many of our papers that have come out of the center work.

So, we are very cognizant of the importance of parent input on this particular population. And, really tried very hard to have serious incorporation of that view.

CO-CHAIR BROOKEY: Comments about

usability?

MEMBER BRISTOL-ROUSE: Yes, I think, you know, I would just reiterate my comments about communication being so key because I can think very specifically about one provider who we had never seen before from our practice who was as good if maybe even not better than our primary care provider.

So, I guess what I am struggling with this is that I think this measure is necessary, but it just may not be sufficient.

CO-CHAIR BROOKEY: Yes.

So, Jill?

MEMBER MORROW-GORTON: And, I just want to piggyback on that. You're measuring both sort of well visits and continuity visits and sick visits and if you only have sick visits, the likelihood you're going to be able to see the same practitioner is going to be small because they can't work 24 hours a day 7 days a week 365 days a year.

But, I guess my question is, where is

the quality? Is the quality in any visit? Is the quality in those regular, not your sick, we have a problem, but those regular how are things going, what do we need to fix, those sort of, you know, the kids with ADHD that you see four times a year because that's how you keep track of what's going on versus, you know, they have a sore throat and they get it checked because they happen to see the same doc.

And then, my second comment is from the vantage point of a payer, what this says to me is we should just pay docs to see kids more when I'm -- I mean, not more money, but we should just pay them for more visits.

When, you know, David's comment about the CCM program, well the reason the CCM kids get to the doc as frequently as they do is because they have a nurse case manager who they're talking with on a probably weekly, if not multiple times a week. And, it's that person that's driving things.

So, my concern is, you know, is this

capturing what is driving or is this sort of a proxy for that?

DR. MANGIONE-SMITH: So, a lot of things in there.

(LAUGHTER)

DR. MANGIONE-SMITH: Where can I start?

So, I think your point that the, you know, sort of the biannual, quarterly, whatever, continuity visit, right, where you're trying to see that same person who's doing the care coordination and thinking about how do we tune up your care plan? How do we do all the coordination things that are needed?

You know, do I want you to see that same person or no more than two people who are focusing on that?

Yes, I think it's -- those visits, it's probably more important than the strep throat visit.

That's part of the reason the measure takes you down to .5 and still says you pass

because we know you're going to have to see other people when you're sick a lot of the time. You're not, I don't want to work 365/7, you know, 7 days a week, you know, whatever. And nobody can do that.

I don't think this measure incentivized seeing more visits. I want to be clear or pushing people to have more visits.

This measure is looking at those people who are naturally already doing that and saying we want to make sure you're having decent continuity if you're coming in a lot and accessing care a lot.

Because that's a signal that things are not under good control. And, one of the tenants, at least in our conceptual framework, the way you get things under control, is to have it just a couple of people who are trying to figure out why are things breaking down? Why aren't things going well? What do we need to change so that your child does better?

CO-CHAIR BROOKEY: So, just sort of

backing into the evidence, I think that we agree that the evidence was strong for this measure. But, speaking as a pediatrician that only sees patients one day a week, every time a patient sees me, it's a failure of the system because I'm seeing somebody else's patient because I don't have a panel. Okay?

So, what I try to do is communicate as well as I can, to Tara's point, because I know that I may never see this patient again.

But, we do and if you have an organized system, and not everyone does for Medicaid, then you have to incentivize your medical groups and so forth to make a priority for PCP bonding and having the patients only primarily see one particular provider or a small group of providers.

So, I think there is an action that can come out of this, but it's not an easy one, especially when the primary goal of most Medicaid providers is not bonding, it's access in general.

MS. ARTHUR: I think it's important to

keep in mind, too, though, we are talking about like six percent of the U.S. child population and that we've focused this on children with medical complexity.

And, I think that that's an important distinction to make.

CO-CHAIR BROOKEY: I'm just saying that it's still not easy to do, but I think there has to be some energy and leadership around it to make it happen.

So, David?

MEMBER KELLER: No, I share Jon's love of continuity and I think it's important and I share your problem as a one day a week person who actually does have a panel because we all have to have a panel and it's very challenging.

I spend a lot of time talking with my colleagues who see patients the days I'm not there.

Actually, I had a difference question, though, which is, you've proposed this as a plan level or a sort of system level. Has there given

much thought to making this a healthcare system level measure in addition to being a plan measure?

I can -- I think there is one of the problems in our large multispecialty organizations that we have, the Kaisers of the world, the children's hospitals of the world, is that we don't value continuity within our own organization and I can see value in pushing out a measure like this to be used at an organizational level as well as a plan level.

DR. MANGIONE-SMITH: Not something, obviously, that we tested, definitely something of interest. I think with accountable care organizations, this sort of thing should be at this, you know, should be very central to what they're trying to do.

CO-CHAIR BROOKEY: Okay, so, I'm going to take a comment from Maureen. We do have lunch here if you want to try to wrap up a little earlier, we can get there a little earlier.

So, Maureen?

MEMBER EDIGER: That's a lot of pressure to be concise.

(LAUGHTER)

MEMBER EDIGER: So, you asked for the parent opinion, so just really quick. When I think about the times when it was most important that I had the continuity of our pediatrician who knows our family so well that she's had our twin's birthday party at her house.

Unfortunately, when you have a medically complex kid, you end up making these relationships that are outside of the sort of normal.

And, so, I feel really lucky to have those and I know I have a couple of nurses that have just given me their cell phone number and said, just text me, don't bother coming into to the ER.

So, this doesn't capture all of the work arounds that parents have done who have very complex kiddos. But, I think it's important to acknowledge that not all parents have the ability

to make those relationships, that they're not going to be assertive enough, they're not going to speak the same language, they're not going to be able to forge those kind of relationships.

So, I just feel like it's important to acknowledge that you get into a little bit of -- there's not all families are playing the same system. And, the goal is that, you know, my goal is to stay out of the office and the hospital as much as possible and I wish 2017 Maureen could go and talk to 2009 Maureen and tell her that that would be goal because it took me a while to get there.

(LAUGHTER)

MEMBER EDIGER: So, that's why I think there's something really important about parents being able to mentor and help other parents find those skills, not just about asking questions, but being assertive and finding a way to kind of work around some of the systems.

That's all. Lunch.

CO-CHAIR BROOKEY: Not quite, but any

other comments before we vote?

(LAUGHTER)

CO-CHAIR BROOKEY: All right, good discussion. Let's go ahead and vote.

MS. JUNG: The voting for Measure 3153 for use and usability is now open. Option 1, high; option 2, moderate; options 3, low; and, option 4, insufficient.

The voting is now closed. The results are 5 percent for high with 1 vote, 73 percent for moderate with 16 votes, 23 percent for low with 5 votes and 0 percent for insufficient with 0 votes.

The measure has passed for usability and use.

CO-CHAIR BROOKEY: Okay, and before we vote for overall, any additional comments?

(NO RESPONSE)

CO-CHAIR BROOKEY: Okay.

MS. JUNG: The measure is now open for overall suitability for endorsement. Option 1, yes; option 2, no.

CO-CHAIR BROOKEY: People are pointing all over the room.

(LAUGHTER)

MS. JUNG: Voting is now closed. The results are 77 percent for yes with 17 votes, 23 percent for no with 5 votes.

The measure is recommended for endorsement.

CO-CHAIR BROOKEY: Thank you.

All right, so, before we break, do we have any NQF member of public comments?

OPERATOR: At this time, if you would like to make a comment, please press star then the number one.

And there are no public comments from the phone lines.

CO-CHAIR BROOKEY: No comments, okay.

All right, are there any comments from you before we break?

(NO RESPONSE)

CO-CHAIR BROOKEY: Okay. So, we have -- we're a little ahead of schedule and I'd like

to keep us ahead of schedule if that's okay. So, we're supposed to get how much time for lunch? Twenty minutes, so, let's say at ten after, even if you have food in front of you, we'll get started again, is that all right?

Okay.

(Whereupon, the above-entitled matter went off the record at 11:49 a.m. and resumed at 12:15 p.m.)

CO-CHAIR SUSMAN: Okay, folks, I'd ask you to gather around and we're going to restart. Please feel free to get your dessert or fruit, we have some healthy options here today.

And, we're going to start with rate of emergency department visit use for children managed for identifiable asthma, visits per 100 child years. And, our measure developers are calling in, so let's see if they're on the phone.

Oh, one person's here. Okay. I'm glad you're here, welcome.

DR. KLEINMAN: I'm here, too.

CO-CHAIR SUSMAN: Okay, Suzanne, it

is?

MS. LO: Yes.

CO-CHAIR SUSMAN: Okay, Suzanne, you can go ahead or your colleague and provide us an overview.

MS. LO: We have Dr. Kleinman on the phone so he's --

CO-CHAIR BROOKEY: You'll need to use your mic.

DR. KLEINMAN: Hi, thank you, Suzanne and thank you to the committee at NQF.

This is Larry Kleinman, y'all can hear me okay?

CO-CHAIR SUSMAN: Sounds good.

DR. KLEINMAN: Okay, great.

So, I'm very pleased to present this measure which is a result of work from the Collaboration for Advancing Pediatric Quality Measures, one of the CHIPRA Centers of Excellence.

And, we were asked to measure on emergency department asthma and overuse was the

whole suite of things that we did.

We felt first, we needed a good measure of emergency department use, that's what this is an attempt to do. We feel very comfortable with it.

We -- there was a question in the -- that we received, what is the overall rate in the state? So, I can tell you that it was 20.65 visits per 100 child years and in the various age groups, it ranged from a low of 15.08 in teens to a high of 29.7 in the younger children.

And, we think that there are biological and behavioral differences in age that it's better to report it as a stratified measure than as a single measure. It's a more accurate reflection.

But, we're open to whatever the committee would direct us towards.

We were not able to come up with a statistical summary of reliability appropriate for this distribution. We tried, we looked and we consulted with statisticians who thought our

best idea was to do either a Poisson or a zero-inflated Poisson or a hurdle model.

We've done all of them, they all come up with very similar results which suggests that we get good confidence intervals with many plans and many counties being different from one another.

I have some specific data I'd be happy to share. We shared it with staff this morning. I'm sorry we weren't able to get it in advance. We had IRB issues that prevented me from doing the analysis until over the weekend and I'm home with a newborn baby, so it's been a little bit of an administrative challenge and a personal challenge to get some of the responses all done.

But, we do have data that's now been shared with the committee and I can discuss it with you.

There was a question about some codes. We omitted some codes in terms of our denominator. We didn't add any, so I can explain that if there's questions about it.

And, I know there's going to be some interest in risk adjustment and I'm happy to discuss that at some length.

But, let me just say, I think this is a really good measure. It is a true epidemiological rate. It was developed by a systematic process that involved stakeholders and an expert, a RAND expert panel.

I know you've had discussion about the RAND panels earlier in the day and I'd be happy to answer questions and respond to what the committee would like to know.

CO-CHAIR SUSMAN: Thank you.

Suzanne, any further comments?

Okay, so, we will go to our lead discussants. Karen?

MEMBER DORSEY: Do you want to talk a little bit of overview or do you want to focus on evidence?

CO-CHAIR SUSMAN: Why don't we, if you have key concerns that you can foreshadow those, but let's stick with the evidence since that's

where we'll have to start.

MEMBER DORSEY: So, the developer already brought up on of the concerns which is really, you know, how to interpret and think about the analysis that they provided for reliability of the measure. So, I think we'll get to that when we get to that component.

You know, the other, you know, this measure looks at ED utilization for children who have persistent asthma. And, the developers provide a lot of explanation about how they came up with the definition of persistent asthma and I sort of recognize the challenge of coming up with a sort of robust and well accepted definition.

And, I think that they, you know, sort of did jump through all the hurdles to make sure that they had a decent one.

But, you know, the measure sort of rests on your belief in that to a great extent. And, so, it may be worthwhile hearing directly from them a little bit about that process of developing the definition and sort of their --

the evidence that supports it.

And, other than that, the evidence linking, you know, ED visits as a poor outcome and the fact that hospitals can -- and health plans and providers can take interventions to reduce ED visits. I feel they are on very solid ground there as an outcome measure.

So, I don't think there's any issue with the evidence for this measure.

CO-CHAIR SUSMAN: And, Jonathan?

MEMBER FINKELSTEIN: So, I agree, sticking to the evidence frame, I have no problems seeing the ED visit as an outcome for childhood asthma and the link to evidence that we can do something about that is broad, deep and well summarized by the developers.

I think there will be issues about the specifications that I'd like to have clarified, but on the evidence, I'm good.

CO-CHAIR SUSMAN: So, what I'll do is I'll go through our discussants and then I'll ask the developers to talk a little bit about the

specifications so we all know what we're talking about to begin with.

So, Carol?

(NO RESPONSE)

CO-CHAIR SUSMAN: Carol Stanley?

MEMBER STANLEY: Yes?

CO-CHAIR SUSMAN: Do you have any comments?

MEMBER STANLEY: No, I don't have any comments actually.

CO-CHAIR SUSMAN: Okay, well, yes.

(LAUGHTER)

CO-CHAIR SUSMAN: Much ado about nothing.

And, Ricardo?

MEMBER QUINONEZ: I agree that the evidence for what they are saying which is that the evidence that high utilization of ERs is a sign of poor quality and asthma is there.

My only concern, and it does go to evidence, is our ability to diagnose asthma down to two years of age.

And, so, I, you know, I know that they stratify by age, but the -- I don't know of any convincing evidence that we can diagnose effectively asthma at that young of an age.

So, there may be some consideration as to the actual age groups that are included. I think three or above or four and above would be way more precise than including children with two years of age in which you're probably including kids who have diagnosed with asthma but they actually don't have it.

That's probably recurrent wheezing due to viral infections and the evidence for efficacy of treatments there is just not as robust.

CO-CHAIR SUSMAN: Good and, Carol, immediately decided she had to talk.

(LAUGHTER)

MEMBER STANLEY: I did decide that.

And backtracking through some notes, I did want to ask about the evidence of how you're factoring in environmental factors given urbanization and quality of air and quality of

home environment, which is really out of the realm of the health plan and the provider.

So, how would you explain that these visits to the ED are solely attributed as an outcome to the provider or the health plan?

CO-CHAIR SUSMAN: Okay. So, I've heard that comment, a comment about the definition, how persistence was defined and the issue about including the youngest aged children.

So, perhaps the developers could help us with those before we talk about voting?

DR. KLEINMAN: Sure, happy to, thank you.

So, the first thing I would say is that what we did and the way we defined the age groups, both how we grouped ages and how we decided who was in and who was out was through a RAND modified Delphi process of national experts that included pediatricians, family physicians, ER docs and pediatric ER docs, all of whom through an open, transparent process.

The group felt that two was the lowest

age at which one might think about asthma. I would say if the committee, in its judgment, said three, I would disagree, but not argue and be happy with that.

What they tried to do, and this now relates to the definition, the idea was which children, keep in mind that in order to get an enumerator, you have to have an asthma as first or second diagnosis in the emergency room.

So, with that as the requirement of having some form of signal, the question that the attempt to identify persistence or identifiable asthma as it's evolved to, was which children should the practice and/or the health plan be aware of sufficiently? They should be managing their child for the wheezing illness in an effort to try to prevent them from coming to the emergency room.

So, I think it was -- I think there was a bit of agnosticism on whether the diagnosis itself was asthma. As you can see, recurrent bronchitis is a part of the specification. If

there's enough of them, depending on the age group.

And, there need to be more illness in younger children under five to consider that asthma was identifiable than there did in children over five because of this diagnostic uncertainty.

Now, with regard to -- does that answer the question about age, first of all? And, then I'll take identifiable asthma with some numbers that may help you understand it a little bit.

CO-CHAIR SUSMAN: Ricardo?

DR. KLEINMAN: What was that? I'm sorry, I didn't understand.

CO-CHAIR SUSMAN: Okay, I think the group -- it is what it is.

DR. KLEINMAN: It is what it is, that's a fair point. Okay.

So, with regard to this, we wanted something that would be less stringent in identifying asthma than the HEDIS persistent

asthma definition which was very restrictive and, in New York State, only identified 3.1 percent of children as having persistent asthma which we know, based on clinical and other information in the state health plan and Medicaid in New York, it is low.

We had about 15 or 16 percent of New York Medicaid children who would answer -- a parent would answer a survey that they've never been told they had asthma. That number was too high for what we were looking for because we did want a more select group.

And, if you look at anybody who had an asthma claim, it was somewhere around 11 or 12 percent in a given year.

And, this measure, when we applied it, identified 8.6 percent.

So, we think that there is some construct validity by the virtue of the fact that it is a targeted population. It's less than if you look at anybody who has an asthma definition or asthma diagnosis in the course of the year.

But it is less restrictive that the HEDIS definition which was looking at hospitalizations for children with persistent asthma.

We also have some other evidence that doctors actually managed to this. Because when we, as part of our other measures, something that's under consideration now in your coordination of care, the NQF coordination of care call, we have a measure as to whether children after their ED visit get inhaled corticosteroids within two months following the visit.

And, in that, using that metric, we find that children who are seen in the emergency room without this identifiable asthma classification received an ICS prescription about 13 and a half percent of the time.

While those with the identifiable asthma classification receive it 34 and a half percent of the times.

That's almost a threefold increase which suggests that this construct has meaning in

the real world and there's just not an artifice of an expert panel to put it together.

With regard to environmental factors, this measure is not able to account for them. We are using readily available data in order to be feasible, so we're not going into the chart. We can't look at things like asthma action plans or environmental triggers.

But, we can say that one of the points of the National Heart, Lung and Blood Guideline is that children in more challenging circumstances need to be managed more aggressive and that asthma control, which implies no ED visits for because of exacerbations is the goal of treatment for all children. So, we think it's consistent with that.

And, we also would say that, like any measure, it's not perfect, and some judgment may need to be applied because these are all proxies for the real world.

But, we think it's a really good measure in that regard.

CO-CHAIR SUSMAN: Okay, so, are there any other questions that are directly around evidence?

Yes, Jill?

MEMBER MORROW-GORTON: Just a quick question: did you look at matching ED diagnoses of asthma to, say, primary care or -- or pharmacy or some other measure that might have validated that?

DR. KLEINMAN: I -- we looked at visits to primary care docs before and after, but I am not sure -- I am not seeing the connection that you're asking, and maybe we did look at it, if you could clarify, as a validating construct in this regard. So what were you thinking of?

MEMBER MORROW-GORTON: I have reviewed lots of ED records, and the diagnoses don't always make sense.

(Laughter.)

MEMBER MORROW-GORTON: Just -- just the thought, if the ED doc, maybe not even a pediatrician, saw a child who was wheezing, said

asthma, was that borne out after? I mean, was that something that was known before? Did the primary care know that? Was that -- or was that something that was sort of proposed at the time of the ED visit, and -- and the child went on to never have another wheezing episode?

DR. KLEINMAN: These are claims data, so that -- that is actually beyond the scope of what we could have done, but keep in mind that in order to qualify for the denominator, they had to have had a number of indicators prior to the ED visit that this was asthma. That is where this definition -- that is why restricting to only those with identifiable asthma comes into play.

As part of the -- the validation for the appropriateness measure which is coming up, we did look in charts, and -- but we didn't -- we did not cross-reference claims and charts. We did not have those for the same children, so we couldn't do that.

CO-CHAIR SUSMAN: Jeff?

MEMBER SCHIFF: This is probably maybe

more of a comment than a question, but some of the ED visits are preventable only with social service interventions, you know, adequate housing, less social chaos, those kind of things, and I don't -- I don't think you -- you didn't do any literature -- I guess my question is there is no real literature -- literature review about that, on that being -- I just want to -- I -- I think that the challenge of this measure is how much to hold health plans or providers accountable for things that are outside of that scope.

MEMBER DORSEY: And can I just -- just I think a point of clarification: the developers do present data on disparities that I think we're going to get to, which I think speaks to this issue, but I don't believe that they stratify the results by -- they do by race and ethnicity, but not -- I don't believe, and I would like some clarification there, about whether they stratify looking at any of these types of SES indicators or, you know, how they thought about that.

DR. KLEINMAN: Sure. I mean, we -- we provide an opportunity but don't require stratification by both level of urbanicity, using the Department of Agriculture's Urban Influence Codes, and by the degree of poverty in the county of residence of the mother, so -- or of the caregiver.

We do have something that looks, but one I would suggest -- and the appropriate measure is an attempt to capture this, and keep in mind, this is a suite of measures that sometimes it's not because they are sick, but because --

CO-CHAIR SUSMAN: You are really breaking up, and we can hear about every third word.

DR. KLEINMAN: Okay. I am going to take you off the speaker, then.

What I was saying is the -- we have created a suite of measures. This is -- this is one of the core measures, as it quantifies the amount of ED use, but the appropriateness measure

looks at whether the children themselves were actually sick or in the emergency room, perhaps for other reasons, and the -- and -- and there -- there is going to be a background of failure. There is going to be a failure rate that every plan has. While in theory it should be zero, in practice, it is never going to be.

And again, we would hope that these are used by people who are thoughtful and -- and able to -- to identify that there is a tremendous amount of signal here, but it is -- it is not without some noise, and -- and we hope that we have demonstrated that through the -- through our analysis. I would think if -- if the idea was everything had to be attributable to the health plan, we couldn't even use immunizations as a measure because really that is there as well.

CO-CHAIR SUSMAN: Okay. Ah yes, please.

MEMBER FINKELSTEIN: So I just think this suite of measures is complicated, and we should stay focused. My understanding is if we

say this is an outcome measure, all we have to decide on is that the actions of the healthcare setting can move that needle, and I am -- I think that is -- that is really clear from the literature. If I know a kid has asthma, I own whether or not he shows up into the emergency -- I can't prevent everyone, but I own it, and if there are cockroaches in the home, I own that, to the extent I can do something about it.

CO-CHAIR SUSMAN: We will make sure you get cockroaches on every visit.

(Laughter.)

CO-CHAIR SUSMAN: I think that is well stated, and unless there are others who would like to comment or have pressing questions, why don't we move on to a vote?

(Pause.)

CO-CHAIR SUSMAN: And this is evidence.

MS. JUNG: The voting for Measure 3189 for evidence is now open. Option 1, pass; option 2, not pass.

CO-CHAIR SUSMAN: This is an easy one.

(Pause.)

MS. JUNG: Still waiting for one more.

CO-CHAIR SUSMAN: Do you have your people from on the line?

MS. JUNG: Okay.

CO-CHAIR SUSMAN: Okay.

MS. JUNG: Oh. The voting is now closed, but --

CO-CHAIR SUSMAN: Voting is closed, but -- .

MS. JUNG: Let's try that again. Yes. Just --

CO-CHAIR SUSMAN: Okay. We're going to have to re-vote, so if you will get your clickers?

MS. JUNG: Try that one more --

CO-CHAIR SUSMAN: I thought it would be --

MS. JUNG: -- time please.

CO-CHAIR SUSMAN: -- easy, I mean -- yes, no.

(Laughter.)

CO-CHAIR SUSMAN: Are we ready yet?

MS. JUNG: Yes. It is open now.

CO-CHAIR SUSMAN: Okay, open, so please.

MS. JUNG: It is not working.

CO-CHAIR SUSMAN: Vote early, vote often.

DR. KLEINMAN: I am from Jersey. That works for me.

(Laughter.)

MS. JUNG: Let's see. It is still not working.

CO-CHAIR SUSMAN: Yes. I don't know if it is acceptable to do a hand-vote, or --

PARTICIPANT: Yes, for this one right now.

CO-CHAIR SUSMAN: Okay. Well, while we try to sort out the mechanical, why don't we do a hand --

PARTICIPANT: Close your eyes.

CO-CHAIR SUSMAN: Yes, close your

eyes, raise your hands. We will close our eyes too, and we'll just go with the Force.

(Laughter.)

CO-CHAIR SUSMAN: Okay. So --

DR. KLEINMAN: Do you want me to count from here?

CO-CHAIR SUSMAN: Pardon me?

DR. KLEINMAN: I can count from the phone if everybody's eyes are closed.

(Laughter.)

CO-CHAIR SUSMAN: Yes, right, yes, thank you.

Okay. So all those who vote 1 for pass? And I will rely on our --

PARTICIPANT: 1, 2, 3.

CO-CHAIR SUSMAN: -- what is that? Okay.

PARTICIPANT: 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21.

CO-CHAIR SUSMAN: Okay. And all those who are voting 2 for not pass? Okay. We've got one, so we have one not pass, the rest pass. It

passes.

Okay. Let's talk about gap.

MEMBER DORSEY: So although it was -- it wasn't in the place that I am used to finding the performance distribution, the -- the developer did provide that, and there is a distribution across measured entities, and they also provided some analysis looking at rates among different racial and ethnic groups and in the different age ranges, and so there is evidence of both gap and disparities presented in the materials.

CO-CHAIR SUSMAN: Jonathan?

MEMBER FINKELSTEIN: Yes, I agree. Nothing to add.

CO-CHAIR SUSMAN: Okay. Carol, anything to add?

(No audible response.)

CO-CHAIR SUSMAN: Okay. And Ricardo?

MEMBER QUINONEZ: Nothing to add.

CO-CHAIR SUSMAN: Okay. So other comments or questions? I think we heard a few

questions about other dimensions that might influence this measure, degree of urbanness, for example. All potentially important to consider.

(No audible response.)

CO-CHAIR SUSMAN: Seeing no hands, I think we can go on to vote. Now, whether we will be able to vote electronically will --

MS. JUNG: I think we will for this one. So for Measure 3189, voting for gap is now open. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

CO-CHAIR SUSMAN: I see the numbers running. It looks like it is working.

(Pause.)

PARTICIPANT: This is the most important part.

CO-CHAIR SUSMAN: Yes.

MS. JUNG: Voting is now closed. The results are 39 percent for high, with nine votes; 57 percent moderate, for -- with 13 votes; 4 percent for low, with one vote; and 0 percent for insufficient, with zero votes. The measure has

passed for gap.

CO-CHAIR SUSMAN: Okay. Moving right along, we have the issues of reliability, and maybe it could be clarified by the NQF staff what additional materials were sent and the sufficiency of such analysis, because it sounded to me a little bit like there was a Poisson distribution assessment, and I wasn't clear exactly what the -- we had.

DR. NISHIMI: The original analyses that came in spoke more to the issue of meaningful differences, so we did follow up with the developer, who did provide some, a couple hours, just, so we haven't had time, as we sat here, to then look at those, so we have asked the developer -- and we told the developer that, and so we have asked them to summarize what they found.

CO-CHAIR SUSMAN: Okay. Maybe in a minute, or very very briefly, the developer could remind us what you did and found with that so we have a basis --

DR. KLEINMAN: Sure.

CO-CHAIR SUSMAN: -- that we --

DR. KLEINMAN: So we -- happy to do that, thank you. We have done the model three different ways, all of which are theoretically coherent, and the results of which are coherent Poisson, zero-inflated Poisson, and hurdle.

In the Poisson model, which is what we shared just because it is easier to -- to present, we ran it with a number of different plans as the index plan, which you can -- you can control in SAS with a little bit of data manipulation, and this is a -- this is a typical finding, so I have got one, two, three, four, five, six, seven, eight, nine, ten of the -- of 18 plans have a -- a different p-value of less than 0.001. Using chi-squared, we have one that is 0.873 from the index plan, one at 0.516, so you're -- one at 0.58, so you are seeing that some plans are similar, some plans are different, which is what you would want to see.

The standard errors are typically

small relative to the -- to the estimates, so a coefficient of 0.34, with a standard error of 0.05, 0.65, and 0.05, so you are seeing what you would like to see to be able to show that there are tight confidence intervals, but that not every plan is different from every other plan. And that is what we were able to do. This -- this model does not produce an s-statistic, which means we can't calculate in the standard way the -- the variability -- oh, not the variability, the reliability, but -- and this model is adjusted for age group because that is how we are recommending that it be -- it be done.

CO-CHAIR SUSMAN: I know that we are perhaps at some slight disadvantage, not having the data, but as a non-statistician, I am not sure that would help a whole lot. But let me invite the lead reviewers to make comments, and then the group to weigh in.

MEMBER DORSEY: Right. So I mean, it's a little bit difficult for me. I am also a non-statistician, but it's a little difficult for

me to really conceptually grasp what -- what you all are providing. It still sounds more in the realm of -- of sort of model validity for constructing the measure score and not a precision estimate, which is really making sure that if you repeat a measurement, that at the level of the attributed entity, you are getting the same value.

And -- and so it -- that makes it difficult for me to make recommendations or -- or think about how to vote on this, and I don't know if, you know, there is a role for digesting the information that was sent a little bit more here at NQF and then revisiting this -- the vote on reliability, but it is really hard for me to figure out how to go forward without better understanding whether this -- this testing really represents a test for reliability.

MEMBER FINKELSTEIN: So --

CO-CHAIR SUSMAN: Jonathan.

MEMBER FINKELSTEIN: Yes, so let me just say, and maybe there is guidance from NQF:

the whole notion of -- this is a claims-based measure. The whole notion of reliability, inter-rater reliability, test/retest reliability, does not really hold in the same way for a claims-based measure.

If you run it, if you run a SAS code on claims, and two days later, you run it again, you will get the same result. That is test/retest. It is not like -- it is not like asking somebody and then asking them three days later. It is -- also inter-rater reliability has no meaning, so the best you can do are these quirky mathematical things that most of us who are not PhDs in statistics don't really understand the nuances of.

So to me, reliability and validity for these claims-based measures blur somewhat, and I think there are some issues that we should -- should talk about with regard to whether you are measuring with precision the thing you want to measure, that is, validity. Whether you are measuring it with precision is the reliability,

and for me, they boil down to several things.

Number one is the issue -- I would love the developer to explain the issue of the denominator time. My understanding is that if I had an ED visit in -- in February, the time for me to be qualified as having asthma is 13 months. It is the prior year plus January of that year. If I have an ED visit in December that is in the numerator, I have actually 12 plus -- through November, I have like over 20 months in order to be qualified as -- as having asthma, and I am worried about what bias that introduces given the seasonality of asthma, so that is one.

A second one is the developers made some interesting choices based on their expert panel to exclude a class of medications that is usually included as asthma medications, specifically --

CO-CHAIR SUSMAN: Jonathan, I am going to interrupt just a moment --

MEMBER FINKELSTEIN: Yes.

CO-CHAIR SUSMAN: -- because it sounds

like a lot of these are more validity issues --

MEMBER FINKELSTEIN: So do you want me to --

CO-CHAIR SUSMAN: -- as opposed --

MEMBER FINKELSTEIN: -- hold them to --

CO-CHAIR SUSMAN: -- to --

MEMBER FINKELSTEIN: -- validity?

CO-CHAIR SUSMAN: -- to the --

MEMBER FINKELSTEIN: That is fine.

CO-CHAIR SUSMAN: -- precision, the issues of reliability --

MEMBER FINKELSTEIN: Okay.

CO-CHAIR SUSMAN: -- at this point, yes.

MEMBER FINKELSTEIN: Yes.

CO-CHAIR SUSMAN: We have a hard enough time doing reliability without mixing it with validity, but good comments, for sure.

Now just to presage what we can vote on, one option that will come up is to say that it is insufficient because we just have not had

the opportunity to digest the material, or the material ultimately was not provided, so, you know, frankly, I find it difficult to be able to vote on something with any validity when it is not present. But that is my own bias.

So I am looking for further comments here about reliability, if any.

DR. KLEINMAN: May I make one comment, two comments, quickly in response to things that have been said that are related to reliability, not the other component?

CO-CHAIR SUSMAN: Sure.

DR. KLEINMAN: Okay. Thank you. One thing, in response to what I believe it was Karen was saying, we actually have provided a precision analysis. So when I say that the estimate is 0.34 and the standard error is 0.05, that tells us the 95 confidence interval is between 0.25 and 0.45, and we have that for each plan, and we also have that for every county. So the -- and the confidence intervals are narrow enough that plans differ and are outside of one another's

confidence interval, so that is the precision analysis.

The other, again, I apologize that we didn't get this to you sooner. It was actually an impossibility because I left -- I moved from Mount Sinai to Case Western, and when our IRB -- and I remained the PI because I am an adjunct at Mount Sinai, our IRB did not know how to figure this out. It took them four months to grant our renewal, or our continuation, of the IRB, so I was literally not able to touch the data --

CO-CHAIR SUSMAN: Yes. You know, that is --

DR. KLEINMAN: -- so --

CO-CHAIR SUSMAN: -- certainly understandable, and we certainly appreciate your explanation. We do have an option of asking for the data, and then being able to reconsider this during our phone call, so I think, as I am looking around the room, that there is a fair amount of head-nodding around trying to be able to look at all the expounded information that you

are discussing, but let me get Jim.

MEMBER BOST: So what you added with your Poisson analysis gets more -- gets as well to the meaningfulness by providing additional data, but it didn't sound like it actually gets at what you would call a typical definition of reliability. I mean, I could see you potentially doing some chart reviews to see if your claims algorithm matched up with the chart reviews, but I didn't see that, or if -- if you have a -- a program that runs this, you could have tested that program at a variety of different entities to see if it runs, so I guess I was looking more for those kind of things for reliability here that I did not see.

DR. KLEINMAN: I -- I think we did provide a literature review that took the claims to the constructs under the validity section, so that part, we do have, and we did -- we looked at various plans, and we looked at counties and found that it worked in both all within a single data set, which is what we had available. Thank

you.

CO-CHAIR SUSMAN: I am wondering if a vote on reliability, and if we indeed vote that it is insufficient, there will be the opportunity for the developer to provide all the information that has been asked for, some of it provided, and we can take a look at that at our leisure. So unless there are objections, I would suggest that we would vote on reliability.

CO-CHAIR BROOKEY: Which would mean we would discuss it in the phone call meeting, is that correct?

CO-CHAIR SUSMAN: Which would mean we would discuss it on the phone call meeting.

DR. NISHIMI: Post-comment call.

CO-CHAIR SUSMAN: The post-comment call, thank you.

Okay. So let us move to voting on reliability.

MS. JUNG: The voting for Measure 3189 for reliability is now open. Option 1, moderate; option 2, low; and option 3, insufficient.

(Pause.)

MS. JUNG: Still waiting on one more vote.

(Pause.)

CO-CHAIR SUSMAN: Everybody might want to -- quick, again, we have our offsite person.

MS. JUNG: Oh, there we go. Voting is now closed. The results are 4 percent for moderate, with one vote; 17 percent for low, with four votes; and 78 percent for insufficient, with 18 votes. The measure did not pass for reliability.

CO-CHAIR SUSMAN: And with it not passing reliability, do we continue on, or do we stop there?

Okay. So we will stop there. I think my sense from the group here is that this is a very viable measure, but having to conform to the NQF procedure, we just need a little bit more data, and you might likewise consider the information that was provided for validity testing, making sure that all our potential

answers at the post-comment call will be provided. So thanks very much. Yes?

DR. NISHIMI: It might be useful for the developer to hear some of the concerns about validity that Jeff wanted to raise just so that, although the committee won't vote, it will be useful feedback in anticipation of the post-comment call.

CO-CHAIR SUSMAN: Great, great point. So let's take this opportunity then to focus on validity, or any other really core questions, but validity is also a must-pass element, so Jim, did you have a comment still, or -- okay. Jon.

MEMBER FINKELSTEIN: So -- so I will try to be -- I will try to be really succinct. The three issues -- I think this is really useful because then the developer can bring us back information. The three issues that stood out to me are the variable ascertainment period for who is an identifiable asthmatic and how that varies, especially seasonally throughout the year, that is number one.

Number two was the exclusion of Beta agonists, short-acting Beta agonists as an asthma medication, which I understand is a -- is just a decision that the expert panel made, but I did not see any analysis of how that does or does not affect the measure performance. And the third one, again, a decision the expert panel made to include the diagnosis of bronchitis, and in the asthma research I know, that does not affect things very much, but -- but if that is just a decision the expert panel made, I am wondering if they did any analysis of whether including or not including those diagnosis codes makes any difference at all. So those were my three validity issues.

CO-CHAIR SUSMAN: Thank you so much. Karen, do you have any additional comments there?

MEMBER DORSEY: No. I agree that the denominator definition is the one that I am -- that I focused most on, so a little bit of more clarity there I think is really helpful.

CO-CHAIR SUSMAN: And again, you may

well have provided this, but if you can make it absolutely crystal so at our post-comment call we can address this, it would be very useful.

The rest of the group, other comments that our measure developer should be considering for a resubmission at the post-comment, or the ongoing submission, if you will? Okay, yes, Dave. Dave?

MEMBER KELLER: Well actually, the only -- only -- just a question for the Chairs: the next measure we're about to consider sort of builds on this measure, and the fact that we deferred decision on this measure, does that impact our ability to consider the next measure?

DR. NISHIMI: No. You should -- you should judge the next measure with what you have in front of you. If you judge that it is similarly insufficient, then we would stop and provide the developer feedback, but we wouldn't just set it aside.

MEMBER KELLER: Got you.

CO-CHAIR SUSMAN: God has spoken.

(Laughter.)

CO-CHAIR SUSMAN: All right. We appreciate the hard work and the very interesting and useful concepts here and look forward to reconsidering that at our post-comment call, and we will move forward then to another very related measure and allow our developers to discuss the broad overview of that one.

DR. KLEINMAN: Sure.

CO-CHAIR SUSMAN: This is a --

DR. KLEINMAN: Thank you very much. I am sorry. Go ahead?

CO-CHAIR SUSMAN: I was just saying this is the appropriateness measure.

DR. KLEINMAN: So thank you, and I appreciate your feedback on the other, and this is a part of our suite of asthma measures. As I indicated, some are being considered by Care Coordination, some by Pediatric.

This measure was an attempt to recognize that there are various reasons for use of the emergency room, one of which is a clinical

outcome failure which is generally discussed and which that first measure was designed to identify. Another may be failures of the primary care system or other aspects of the -- of the way that care is delivered, as has been demonstrated dating back to the Three City Study in the 1980s, I believe, early 90s.

And so we wanted to know, if a child was in the emergency room, was that the appropriate level of care for that child? And our expert panel gave us several -- several key reasons for them to be there, some of which would require patient-centered information, which we can't collect, and -- in this measure, which is a chart review measure, but most of which are in the chart.

And so we set this up as either appropriate or of questionable appropriateness. It should be clear we are not using the word "inappropriate," in part because of that. We -- we think this is an important measure. It is a measure both of overuse and a measure which

allows for interpretation of other data, such as the -- the frequency of ED use, whether you count it with our measure or some other, and it also can help to drive improvement as well as being an accountability measure.

Anyway, I am happy to take questions and to address the -- the concerns and interests of the committee.

CO-CHAIR SUSMAN: Okay. Thank you very much. Anything further, Suzanne?

(No audible response.)

CO-CHAIR SUSMAN: All right. Well, let's go down to our leads with Ricardo.

MEMBER QUINONEZ: Well so I understand we will be discussing the evidence --

DR. KLEINMAN: I can't --

MEMBER QUINONEZ: -- first --

DR. KLEINMAN: -- hear.

MEMBER QUINONEZ: We will -- we will be discussing the evidence, correct? And since this is an outcome measure, we would consider whether there is any modifiable process that can

influence the outcome, and so based on that very narrow definition, I would say there are processes of care that can influence the outcome. There's -- there's other concerns that I have, but not related to this narrow definition of evidence for outcome measures.

CO-CHAIR SUSMAN: Okay. Let's go down the list: Marlene?

MEMBER MILLER: Yes, I mean, while I agree with that, I think this does bring up some of the same issues as the other measure. Obviously, for example, this measure predisposes that everything is modifiable by the health plan, but there is no ability to account for a simple issue such as compliance, some of the -- you know, medication compliance in the home setting, the social stressors, et cetera, so it gets very blurry for me when we try to say the evidence. The evidence is that we can impact those rates, but this measure to me the way it is constructed suggests that you would impact all of it, and because there is no accounting for these other

factors which are, you know, likely beyond full control of the health plan, i.e. for example compliance of medication use at home.

CO-CHAIR SUSMAN: Okay. Thank you. My issue probably revolves around calling this "appropriateness" because you don't know who doesn't go to the ER, and while it may well be a very small number that don't appropriately end up in the ER when they should, all we know are the people who go.

So it is in some ways I guess more of an overuse measure, potentially, but it is just as important to think about underuse, so when I think about outcomes, I mean, maybe it would be a far worse outcome to have a child not sent appropriately to the ER, so even though it might be very infrequent, differences that are meaningful might be very important to capture.

I see a lot of now hands up, if you will. Let me go to James to give final comments from the primary reviewers, and then we will get to the further comments and questions.

MEMBER BOST: Sure. I think there was certainly evidence that -- of -- of appropriate/inappropriate overuse in the ED for asthma. When I was looking at the measure and its definitions, so if -- if the rate is low, that is telling me that there are a lot of kids coming who could be taken care of by a PCP or specialist that are not. So, you know, I guess at a health plan level, you could be working to make sure those kids get assigned to a specialist. Not sure at an individual hospital whether you could do that, but maybe you can.

If -- if the rate is at like 100 percent, let's say, then, you know, in essence, that is great in terms of that being assigned to a PCP, but if over time, that stays really high, but your denominator is going up, then that means that there are more kids coming where the PCP is not doing a good job of taking care of them, so they are having more crisis events. So it got a little confusing in my head how you -- how that rate, standing alone, could be used to gauge

whether improvement is done.

CO-CHAIR SUSMAN: Yes, and it -- at the developer, it seems to be in some ways that you meant this to be a sort of paired measure, even though we do have to look at this on its own as submitted?

DR. KLEINMAN: Yeah, I would say it should -- it would best be paired with some estimation of how commonly ED is used. It improves interpretation. But I do think -- I think the point that was just added, I am not sure who that third reviewer was, I am sorry, I didn't hear your name, actually hits the bull on the head.

This is -- it is a challenging measure because if it is high, it means one thing, if it is low, it means something else, but they both have important meaning potentially for improvement, and context matters. So this is an attempt to actually bring a little bit of nuance into the measurement we have done, and in part, it is definitely an overuse measure.

**And it was also done because we did not want the assumption that emergency room use meant overuse of the emergency room at the clinical level, because we were concerned about the potential policy implications of that. We don't want those services cut when it's a safety net for children who need it. So this is -- this is really -- it is a rich and nuanced measure that comes out of a -- a rich tradition in terms of appropriateness, but also access and coordination of care. It sort of brings them all together.**

**CO-CHAIR SUSMAN: Okay. We have a number of questions and comments. I will start with Jon.**

**MEMBER FINKELSTEIN: So my -- and this is maybe definitional, but I think important. So when I was going through this measure, the other measure I totally saw as an outcome measure. I did not see this as an outcome measure. Getting to the ER is an outcome, but this to me is -- is -- the question is is an overuse measure --**

because I agree that is exactly what it is -- a process measure or an outcome measure? I generally think of -- of overuse -- that is just definitional.

DR. KLEINMAN: So Jon, when I originally submitted this measure, and it was submitted initially to the pulmonary group, I submitted it as a process measure, and the committee and staff told us it would be better considered as an outcome measure. So I think in terms of a formal academic framework, it is probably more process, but in the way that operationally these things are used, it would be -- it is an outcome measure. I have become convinced of that, but I share your -- your initial reaction to it.

DR. NISHIMI: I just want to reinforce what Larry said. The committee did come back and recommend to him that he submit it. It is splitting hairs, you know, ten ways, but they came down on it being an outcome measure, so when you apply our criteria, it becomes a pass/no

pass.

CO-CHAIR SUSMAN: Okay. Let's go down the row. David?

MEMBER KELLER: Jeff actually had his up first, but mine -- mine just is an observation that as was said, this -- what this reminds me of as an outcome measure is looking at readmissions, where you can change the rate by altering either the numerator or the denominator, and in this case, the way -- which of those combination of things changes matters. We found in a number of our readmissions work that as we reduced the number of unnecessary admissions, we had some hospitals that actually would increase their readmission rate because all that was getting admitted was the complicated patients, and that ended up hurting them in some of the schemes.

So I think that is -- the question to the developer is had -- was there consideration of looking at the rate as you did in the previous measure of unnecessary ER use per 1000 child years of children with asthma rather than per --

than using emergency room visits as the denominator?

I think the other piece to that is that this -- the way it is constructed, I could see this as a measure that would be perceived as a problem for the emergency room to solve, and this is not a problem for the emergency room to solve. They -- they have to take anyone who shows up, and so the real question is work that goes on outside the emergency room, and this is clearly a measure of system function rather than of emergency room function, but -- but I could imagine this measure being used, and, you know, on the dashboard of our emergency room at the hospital --

DR. KLEINMAN: Yes.

MEMBER KELLER: -- and then -- but the emergency room docs would have very little to be able to do about that other -- except in retrospect.

CO-CHAIR SUSMAN: Okay. And then finally, Jeff?

MEMBER SCHIFF: I -- just to echo one thing David said, I would like to hear from the developer why they wouldn't use this numerator in their previous denominator, because that would be a measure of the severity of -- that would give you a measure, a better measure of severity of the folks who show up in the ER.

To me, the -- I have a comment and a question. The comment is that in this measure, a high percent may be a measure of -- of insufficient care prior to the emergency room, because somebody would not -- you would say that they -- they should have been treated so they didn't have a low PO2 or S saturation. A low percent may be a measure of bad triage, which is I think what the developer was getting at.

My question though, just, it is probably more the evidence question, is in the criteria for explicit criteria, there's a lot of variation in that, and especially the last one, there -- it says there is clear documentation that prior to arrival in the ED, any of the

following occurred, and I just want to -- I wanted to -- was curious about how that final component was developed, you know, because it seems like that is -- there are some of these that are pretty hard, you know, 2 saturation below 90 percent, a blood gas was obtained, even though labored breathing I think is maybe subjective, I just want to know how they -- how they -- where is the validity or the evidence for that?

DR. KLEINMAN: Should I respond? Is that -- is there --

CO-CHAIR SUSMAN: Yes, please --

DR. KLEINMAN: -- a question?

CO-CHAIR SUSMAN: -- if you will.

DR. KLEINMAN: Sure, of course. So I think in response to the last -- not the last comment, but the last commenter, and also to David's comment about sort of creating a population measure that integrates appropriateness and the count measures, this is something we have thought about. We just were

not there. We did not have the resources to do it. That was -- we thought that was a -- potentially a later stage of development.

We only had the resources to assess this in one hospital. I agree it's a system measure, but chart reviews are difficult and expensive and time-consuming, and given the -- the constraints that existed from AHRQ and CMS regarding this measure, this is what we were able to do, but I do think that is a future direction that is -- is worth doing and worth thinking about.

In this metric, some of the criteria that the committee suggested such as previous history of rapid deterioration were not operationalized into this measure, and that is why I say some of those are things that require a -- talking to the parent. And since we were restricting ourselves here to chart review data, that is one of the reasons we don't ever call it inappropriate. We say "questionable appropriateness," and that is one of the ways

that that question could be answered.

We do have -- I have one of my former fellows in New York who is currently seeking funding to expand work on this and also to develop a patient-centered survey in the ED that would help to complement these data, but we are not --

CO-CHAIR SUSMAN: Okay.

DR. KLEINMAN: -- implementing those here.

CO-CHAIR SUSMAN: Okay. I am going to move things along. John?

CO-CHAIR BROOKEY: Yes, I just want to make a couple comments in terms of being a health plan measure. We do know there is a pretty good correlation with asthma medication ratio and ED utilization. It is probably the one thing that you can do from a health plan perspective to really change ED utilization is to improve the compliance with medication, as well as to provide access in primary care, so those are the two sort of actionable things a health plan could actually

do.

But having looked at this pretty extensively in Kaiser in California, we know a couple of things. We are -- our Medicaid utilization is double commercial for ED utilization, but if you look at hospitalization, it is not so different, and as I mentioned earlier, our access in the ER is very very good with no copay, and so that kind of confounds your analysis of the data, and I -- I don't look at ED utilization as being an index of severity of illness, but rather access.

And I would look at hospitalization differently, and I think that is kind of what this measure is getting at, is trying to figure that out, and I agree that maybe the denominators need to be switched or something, but I like where it is going. I am not sure the group will say this is the right measure at the right time, but -- but trying to get at severity illness I think is a very important thing.

CO-CHAIR SUSMAN: Well, this has been

a very robust discussion. Does anybody have anything new about evidence?

(No audible response.)

CO-CHAIR SUSMAN: And if not, let's make the vote.

MS. JUNG: Okay. Voting for Measure 2816 for evidence is now open. The option is option 1, pass; option 2, not pass.

PARTICIPANT: What happens if you vote 3?

CO-CHAIR SUSMAN: If you vote 3, we will make you serve on three more NQF panels.

(Laughter.)

MS. JUNG: Voting is now closed. The results are 55 percent pass, with 12 votes; 45 percent not pass, with -- I am sorry, with 10 votes, and then the pass -- 55 percent pass, with 12 votes; 45 percent not pass, with 10 votes.

PARTICIPANT: And what does it have to be to --

MS. JUNG: It has to be at 60 percent, so this does not pass -- I am sorry --

CO-CHAIR SUSMAN: This is a --

MS. JUNG: -- I keep saying that. Consensus is not reached for this.

CO-CHAIR SUSMAN: And we continue to go on then? Yes, okay.

So thank you. We now get to look at gap, and let's have our primary reviewers comment briefly about gap.

MEMBER QUINONEZ: I think the reviewers definitely showed evidence that there is -- there is a gap, particularly I think similar to the prior measure, that there is also a lot of disparity in ED utilization with asthma, so it -- I mean, it -- the evidence that ED utilization varies for patients with asthma I think is definitely there.

CO-CHAIR SUSMAN: Okay. Marlene?

MEMBER MILLER: I do think there is obviously evidence for inappropriate use. I still have trouble that we -- there is not evidence into the parsing of the reason behind it in terms of access, non-compliance, et cetera,

which obviously influence the ability to move that needle, if you will, on appropriate use of the ER.

CO-CHAIR SUSMAN: I don't have anything new to contribute. And Jim?

(No audible response.)

CO-CHAIR SUSMAN: Nothing? Questions or comments?

(No audible response.)

CO-CHAIR SUSMAN: And if not, let's vote on gap.

MS. JUNG: The voting for gap for Measure 2816 is now open. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

(Pause.)

MS. JUNG: Lauren, we -- oh, there it is.

(Pause.)

MS. JUNG: One more vote.

CO-CHAIR SUSMAN: Folks, please vote again just to make sure we captured them all.

Oh, one has stepped out.

MS. JUNG: Snuck out.

CO-CHAIR SUSMAN: Uh-oh.

MS. JUNG: Voting is now closed. The results are 9 percent for high, with two votes; 82 percent for moderate, with 18 votes; 9 percent for low, with two votes; and 0 percent for insufficient, with zero votes. The measure does pass for gap.

CO-CHAIR SUSMAN: Okay. Let's move on and talk about our validity issues -- or reliability issues. I am trying to really move us along.

MEMBER QUINONEZ: So following the -- the script that we were given for discussing reliability, here we considered the numerator statement, the denominator statement, and I think this is -- this is where a good discussion of the developer's criteria for appropriateness may be -- may be in order. I have a lot of difficulty with that because I think there is -- you know, since we're looking at appropriate ER visits, and

I realize that the -- they used a RAND/UCLA Appropriateness criteria method to get at these criteria, but I think there's a lot of opportunities here to game the system.

You know, I think for example, all of the reliability criteria are "or," and -- and so passing any of those would -- would give you -- would tell you that -- that the -- the visit is appropriate, so, you know, for -- one example is recorded oxygen saturation below 90 percent. Is that on presentation, or is that at any time during the ER visit? When you get three back-to-back doses of Albuterol, many children drop below 90 percent. It does not really mean much. And so -- but by these criteria, you would meet that -- that appropriateness criteria.

And so I -- you know, and I think there's -- there's several things in this. The subjectivity of -- of the physical exam in respiratory distress, I think there's a criteria when an ABG was obtained, so I would basically -- you know, I could get an ABG on every single kid

that shows up to the ER because I want to meet this criteria, and so I could game the system that way. So I have a lot of concerns about the appropriateness criteria that worried me in this specification.

CO-CHAIR SUSMAN: And it sounds like at least some of this is about the validity -- face validity of the measure.

MEMBER QUINONEZ: Yes, except the script really wants to talk about that.

DR. NISHIMI: Right, but it does --

CO-CHAIR SUSMAN: It also -- yes.

MEMBER QUINONEZ: Yes.

DR. NISHIMI: It goes to the specifications, which is the first element.

MEMBER QUINONEZ: Right.

DR. NISHIMI: So it's both.

MEMBER QUINONEZ: Yes.

DR. NISHIMI: Yes, no, it's both.

CO-CHAIR SUSMAN: Okay. Did you want to make a comment, Robyn, about how we're going to handle the reliability and actually not vote

on that, as I understand it.

DR. NISHIMI: So one element of the reliability is what Ricardo spoke to, which are the specifications. So you need to keep that in the back of your mind. But the developer has claimed data element level validity, unlike the measure you just discussed where we were looking at actual scores.

So in this case the developer has relied on the literature for the identification of the denominator and then has done some empiric testing, which is described. He had ten charts at the facility against -- the gold standard was the transposition and the other two abstractors. So that would be the empirical testing of validity against a gold standard at the data element level. And that would be appropriate for reliability as well. So you won't vote on reliability testing because we're going to rely on the validity testing.

CO-CHAIR SUSMAN: So as you're thinking about that -- I think your comments are

very appropriate for at least both of those.

We have a bunch of questions here.

MEMBER DORSEY: I just had a question about that guidance specifically. You want me to hold it?

CO-CHAIR SUSMAN: Yes. No, no.

MEMBER DORSEY: So I was just curious whether there's any requirements when we think about data element validity testing about a facility. So it's just one facility, but is that -- that's up to us? There's no specific guidance from the staff about that?

DR. NISHIMI: No, whether you feel that is sufficient. The validity calculation also asks you questions about meaningful differences. Can't do that with only one. But again, that's just one element and the Committee members have to judge where to weight all these different issues.

CO-CHAIR SUSMAN: Jim?

MEMBER BOST: So when I was reading the guidance, it says, was empirical -- and this

is on the reliability. It says, was empirical validity testing of patient-level data conducted? Well, this is item level, isn't it? So I don't see patient level. So I would have said that then goes to insufficient.

DR. KLEINMAN: May I respond to that?

CO-CHAIR SUSMAN: Yes.

DR. KLEINMAN: The overall appropriateness score would be the patient level. The items are sort of supporting data -- information for that, but overall appropriateness is at the patient level.

CO-CHAIR SUSMAN: But aren't there the assessment and an element level?

DR. KLEINMAN: It's element level building up to patient level, and overall appropriateness is at the patient level, not at the element level.

CO-CHAIR SUSMAN: So there --

MEMBER BOST: So how would we assess that if it's only one institution and --

CO-CHAIR SUSMAN: Well, I think that's

certainly an important point. I mean, if we're supposed to be looking at differences and the testing is only at one place, even if we give them a pass on validity, it's hard for me to get too excited about that being an adequate demonstration.

Okay. We are sort of continuing to talk about both reliability and validity because of the interrelationship here. Are there other comments or key issues around this that will ultimately influence our voting?

Yes, I see some hands, so I'm going to get Jim first and then we'll go down.

MEMBER BOST: So it should also be mentioned that in the item testing all items could not be tested because a couple of them were not actually collected at that institution. Now the institution did say that it really wasn't needed, that the other items in the list were sufficient to do that. And they did talk to nine other institutions and showed them the list of items. And they thought it was okay to get what

they're needing. But again, we didn't get an item-level analysis on every item.

CO-CHAIR SUSMAN: And just to add to that, the pharmacy data was variably accessible.

So, okay. From our --

DR. KLEINMAN: And --

CO-CHAIR SUSMAN: Just a moment. Let --

DR. KLEINMAN: Oh, sorry.

CO-CHAIR SUSMAN: Yes.

DR. KLEINMAN: I was just asking if it was time. Okay. Great.

MEMBER EDIGER: I'm just confused. Since it didn't pass on evidence, I just think --

(Simultaneous speaking.)

MS. JUNG: Consensus was not reached.

CO-CHAIR SUSMAN: Consensus was not reached and per the NQF protocol we continue on.

MEMBER EDIGER: Okay.

CO-CHAIR SUSMAN: Yes.

MEMBER EDIGER: Okay.

CO-CHAIR SUSMAN: So, sorry for the

confusion.

MEMBER EDIGER: That's all right.

CO-CHAIR SUSMAN: Okay. Other comments? Jim and Jon.

MEMBER BOST: To its credit, it was designed like a HEDIS measure so that specification stuff I thought it was well designed in terms of assessing whether it could be reliably collected. And the kappas that were collected on the item level for most were pretty good, although they really needed a lot of practice because the kappas were much better after they were doing it for a while.

CO-CHAIR SUSMAN: Jon?

MEMBER FINKELSTEIN: So the criteria for the appropriateness of the ED visit -- some of them are incredibly objective whether a blood gas was ordered, although I'm guessing that happens a very small percentage of the time anywhere. Others are so subjective and much more frequent, and I'm concerned would drive the measure. So I'm talking about referral by a PCC.

I understand why they thought about that, but I'm not sure that makes it appropriate. It just could make it an inappropriate primary care provider. And I'm guessing that that's going to be a much more common occurrence and could drive the measure.

CO-CHAIR SUSMAN: Or if you've got a busy office, just go to the ED.

(Laughter.)

CO-CHAIR SUSMAN: Kerri?

MEMBER FEI: Thanks. Kind of along the lines of what Jon said, a lot of the things in this list you -- would be tough to test because you don't expect to find all of them all of the time. So getting that type of testing, of testing all of the data elements that are listed here I think would be tough. And again, there's going to be some you're going to find a lot and there's going to be some that you aren't going to find very often. And a lot of them are pretty subjective, which would be tough.

CO-CHAIR SUSMAN: Okay. So are there

other -- did you have another comment, too?

(No audible response.)

CO-CHAIR SUSMAN: No? Okay.

Are there any other new questions, concerns or comments from the panel?

(No audible response.)

CO-CHAIR SUSMAN: Okay. So we're actually going to be voting on validity.

DR. KLEINMAN: May I respond to some of the comments because I think I can help clarify some of the things.

CO-CHAIR SUSMAN: Okay. If you could keep your comments very brief.

DR. KLEINMAN: I will do it. There are nine criteria the panel endorsed, only six of which are specified here. All the specified criteria were tested. There were 1,000 or so charts that were looked at, so while I was at one institution it was fairly robust. O2 saturations rarely fell below 90. That was not a major reason for appropriateness.

ABGs I think it's unlikely to game

because if you remember what was talked about earlier, when appropriateness is very high, it suggests you're doing a worse job of asthma management. Low is a worse job of access. So there is a balance intrinsic to the measure that ought to work against gaming. I'll leave it at that.

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

Okay. Let us go ahead then and vote for validity.

MS. JUNG: Okay. The voting for validity for Measure 2816 is now open. Option 1, moderate, option 2, low, and option 3, insufficient.

(Voting.)

MS. JUNG: One more. Is everyone in the room?

(Voting.)

MS. JUNG: Voting is now closed. We have the results. Four percent for moderate with one vote, 74 percent for low with 17 votes, and

22 percent insufficient with five votes.

CO-CHAIR SUSMAN: So this is a must-pass criteria and we did not reach the threshold, so we will stop there.

Thank you very much again. This is a really important work and I think a very important concept. And I see everybody around the table -- which you can't see -- shaking their heads around that, and trying to address some of the, if you will, methodologic challenges here I think will be very useful for the field. So thank you.

DR. KLEINMAN: Thank you. Appreciate your time and your feedback.

CO-CHAIR SUSMAN: Okay. We have another measure, and we're going to be talking about the informed coverage from Children's Hospital of Pennsylvania, CHOP. And do we have the measure development -- walking up as we speak. Thank you.

So if you haven't figured it out, you can provide us your brief overview, introduce

yourself. And we appreciate your attendance.

DR. SILBER: Thank you. It's nice to be here. I am Jeffrey Silber. I direct the Center for Outcomes Research at CHOP and I'm going to be talking about informed coverage.

The measure is constructed to help states get a better sense of their participation rates of their children in Medicaid, participation rates meaning the denominator, the number of kids that are eligible and how many months they should be eligible during the study period, the observation period, and the numerator being of those that are eligible what's the -- or how many actually are enrolled. So we're talking about a fraction of that time period where the kids are enrolled when they're eligible.

The measure fills a gap. Right now there's two kinds of measures that the states can use easily with their claims data. They can use a measure called duration, which is measured as after a gap in enrollment. You follow how long a patient -- once they're back on insurance -- how

long they stay on insurance.

That measure has a problem. That duration measure has a problem because if you don't have gaps, you don't show up in the measure. And we've talked to stakeholders and policymakers at the states who understand that that is a problem. If they're doing very well with a child who's always on during the study period, they don't get counted in the duration metric.

If you -- on the other hand, the other measure is the continuity ratio. It's in the literature. It hasn't been as formally developed as the duration metric. Continuity ratio looks at anyone during the study period who ever had insurance and then it says, well, that's -- then they should have been covered the whole time or participated with being enrolled the whole time. Then what fraction of their months were enrolled. So that's the continuity ratio.

So that has the opposite problem, and that problem -- the problem with the continuity

ratio is that there are kids that never are in the system at all. Those are the worst players, that they were eligible but you didn't know it and those don't show up in the continuity ratio.

So on the one hand you're penalizing states when they do a great job and then on the other hand with the second measure you're ignoring the very worst cases.

So we developed informed coverage to try to get at a better measure of this participation ratio. We developed it using a -- what we wanted was an event that would be random that could tell us about participation. It wouldn't be related to your primary care. It wouldn't be related to any of the care that you've gotten.

And we thought, well, okay, appendectomies are random events. There's some literature about slight changes though with sex and with season, but appendectomies are very random. And then we can then check to see if you actually were enrolled before you came down with

your appendicitis and had your appendectomy.

And we know that you wouldn't show up in the MACs claims unless you were eligible. So we have a measure of eligible kids and then we can check going back in time to see if they were enrolled or not. And that's our measure of participation.

Now we bound this measure with two extremes. One extreme is -- we call the presumed eligible. These are looking at all the claims -- not just an appendectomy, looking at all the claims and saying let's make an assumption that if we see you, you should have been eligible for the whole period. We make some modifications. That's similar to CO, but with some modifications for aging out.

And then there's presumed ineligible, which says, well, we presumed that if we didn't see you, you weren't eligible and therefore that gives you a higher participation rate. So we create two bounds, but we basically -- with this measure -- are using the appendectomy rate and

the percentage that had insurance prior to -- and were enrolled prior to getting their appendicitis event as our metric of participation in the states.

And then what we do -- and I'll finish up -- what we do is we compare our rates to the other two rates, to continuity and to duration. And we also go to what some consider a gold standard, which is the -- a survey method. So the American Community Survey, that is -- I'm saying it's often considered a good standard to look at. It surveys patients and families and asks them whether they're insured or -- whether they were on insurance or not. And so we present that data in our application as well. And we show that our metric is best -- better correlated with the ACS than is continuity or than is duration, and it also has less error than the continuity ratio compared to the ACS.

States can't get a hold of the ACS. In part, they can't link their data to the ACS and the ACS is -- varies over time and there's

even threats about whether that will be available in the same way. So for states that want to use their data to understand their participation rates, we think we have a good solution.

CO-CHAIR SUSMAN: Thank you very much for your succinct and clear explanation.

And we'll move on to Amy.

MEMBER HOUTROW: Yes, thank you. So as we heard, this is a measure of coverage over an 18-month period. And what we need to do here -- because this is an outcome measure -- is take the top part of our guidance for evidence, which is looking at whether this outcome measure has a relationship to at least one healthcare action. And they have provided evidence here that both for the individual who is insured, having health insurance improves their access to health services and outcomes.

And then in addition to the health plans perspective, the Medicaid plan, the disenrolled children who need to be re-enrolled is a very costly process. And so there's both

kind of the patient level related outcomes that they point to and then the systems level outcomes that they point to. And I would take those two as kind of the basic for how we would move forward in a decision about this outcome measure and the evidence for it.

CO-CHAIR SUSMAN: Okay. Kerri?

MEMBER FEI: I totally agree with what Amy said and I think there's plenty of evidence here to support that this is a measure that is needed.

CO-CHAIR SUSMAN: And what does the man from Ohio have to say?

MEMBER KNUDSEN: Agreed.

(Laughter.)

CO-CHAIR SUSMAN: And, David?

MEMBER KELLER: Yes, I actually not only agree, but I applaud them for their creativity. I think using appendicitis as a tracer into the system was a really interesting idea and I think will give some new -- a new way of looking at an old problem in the system.

CO-CHAIR SUSMAN: And the senator from Minnesota?

MEMBER SCHIFF: I didn't announce yet.

(Laughter.)

MEMBER SCHIFF: I think the approach is really novel. I -- my only comment was I just -- the presumed eligible and presumed ineligible, I had trouble with those terms and I was wondering if it could be something like coverage presumed maximally eligible and coverage presumed minimally eligible, because you really are cutting this differently and you're just making an assumption about that. And when I looked up presumed ineligible, I thought of all the reasons why people would be -- think about that as that folks somehow could not get coverage at that time when it's not that. It's just that they -- they're just not on at that moment.

DR. SILBER: I think you're right. We could have better terminology. And you're right that they're supposed to be created to create a bound so that -- what's the worst case scenario

and the best case scenario for what your participation rate is. And we'll -- we could talk about the word maximum in there. I think you're right.

CO-CHAIR SUSMAN: So I'm not hearing a lot of controversy among the five primary reviewers. Are there any other questions or comments? Jon?

MEMBER FINKELSTEIN: So I guess --

CO-CHAIR SUSMAN: Wow.

MEMBER FINKELSTEIN: And this will be my last comment for the day.

(Laughter.)

CO-CHAIR SUSMAN: Okay.

MEMBER FINKELSTEIN: So I'm maybe being dense, but I, too -- I think the -- oh, thank you.

(Laughter.)

MEMBER FINKELSTEIN: Even better. I totally get the appendectomy thing. I agree it's -- that's really good. Are these guard rails that you've -- I don't quite understand the guard

rails, and are these guard rails that you've set up on either side to bound it just kind of theoretical and they're so extreme on either side that states would rarely hit them and what this really is is an appendectomy-focused measure, or are these guard rails guard rails that come into play a lot and people would be maxxed or minned all the -- frequently?

DR. SILBER: So the guard rails are in place for two reasons. So the first is that they create a bound and occasionally the appendectomy rate will be under or over the guard rails. So we didn't want -- we felt uncomfortable with that, so we bound it. It very seldom kicks in. It has a very slight effect, but occasionally you'll see that they're -- that it hits the bound.

The second reason that we have the PI and the PE is that when we look at strata -- let's say we want to stratify by a region or by race. What we do is we work out for every state what the mixture of PE and PI would be to

replicate the appendectomy rate. And then we use for the strata -- we don't go getting appendectomy rates on a minority group or on a subset. We look at the actual PI and PE with all the thousands and thousands of patients in the state. And then we weight them according to what we learned from the overall state rate.

So we still see differences by race and other subgroups based on PE and PI and based on the mixture that was achieved when we first made the overall metric for participation based on appendicitis. So we didn't want to use appendicitis on subgroups where the ends might be too small.

So we use all the patients. We estimate a PI and a PE on everybody, but we -- and so subgroups can be looked at, but how do we put PI and PE together? We use the overall tendency in the state to have PI and PE reflect what we believe to be the true rate, which is the appendectomy rate.

CO-CHAIR SUSMAN: Okay. Jill and

then Jeff.

DR. SILBER: That's the purpose of why we have --

CO-CHAIR SUSMAN: Thank you.

MEMBER MORROW-GORTON: And I may have missed this, but a lot of the disenrollment/enrollment in Medicaid is kids that are eligible one month, not eligible the next month because of parental income. I mean, does this take into account what happens over time for a child versus, you know, your appendectomy rate is at one place in time.

DR. SILBER: Okay, so the nice thing about this metric is it's actually a point in time. So if that child, when it -- when they came down with appendicitis and had their appendectomy, if at that point in time they were eligible, right, then we say that that patient was eligible and then we see if they did or didn't have insurance.

Now how do we do this? We actually -- and it's a little complicated in the description.

We go back four months because we know that kids that are eligible but not enrolled, they can be backdated and so that they can then get insurance. So we say -- we go back exactly four months from when the patient had their appendectomy, and when you ask -- we ask then were you insured or not? Did you have -- you were eligible.

We knew you were eligible four months later, so four -- exactly four months earlier -- because no one thought about the kid coming with appendix when their parents did or didn't have insurance. We then ask, well, did you -- were you enrolled four months before? You might have dropped out the next month and the next month. That's not the case because we'll know that you could be backdated or not backdated.

MEMBER MORROW-GORTON: But you may not have been eligible four months before.

DR. SILBER: Right, you might --

MEMBER MORROW-GORTON: It's July. My parents do seasonal work. I don't qualify for

Medicaid in the summer.

DR. SILBER: Right.

MEMBER MORROW-GORTON: I don't qualify for Medicaid until November when their business closes for the winter. You go back your four months and that kid's going to be not eligible.

DR. SILBER: Right. The beauty is that it's random. So, in other words our estimate isn't going to be biased because we go back four months. It's a random event.

MEMBER HOUTROW: Well, I think this discussion is something we might need to look at a little more in a subsequent part of the discussion, because I think for the evidence part we have a very clear mandate on what we're voting on. And then there are some additional questions I think that will stem exactly from where you're coming from later on.

CO-CHAIR SUSMAN: That's a great point. Jeff?

MEMBER SCHIFF: Just a quick thing about presumptive eligibility and retrospective

coverage. I don't know the rules in every state. Do you know where that's at, because that would -- that -- this measure obviously hinges on that.

DR. SILBER: Right, so when we did our search and we talked to our stakeholders and did our literature search and talked to the authorities, we thought that four months should do it, that we don't think we have a problem going back four months. Three, two and one, states can vary, but no one goes back farther than four that we know of.

MEMBER SCHIFF: And do states -- do all states allow some amount of retroactive eligibility for kids?

DR. SILBER: My sense is they do.

MEMBER SCHIFF: Okay.

CO-CHAIR SUSMAN: Okay. So very interesting comments and questions, some to defer. Anything new before we take a vote on evidence?

(No audible response.)

CO-CHAIR SUSMAN: Seeing and hearing

none, let's move on to a vote.

MS. JUNG: The voting for Measure 3154 for evidence is open. Option 1 is pass, option 2, not pass.

CO-CHAIR SUSMAN: As a clinician this seems very arcane, but important.

(Voting.)

MS. JUNG: Voting is now closed. The results are 100 percent pass with 23 votes. The measure passes for evidence.

CO-CHAIR SUSMAN: Okay. Very good. So, let's move on. Amy, you want to carry forward? You did such a good time the first time.

MEMBER HOUTROW: So we're in the reliability section. And as you guys heard earlier, the numerator is the number of covered months in the denominator.

CO-CHAIR SUSMAN: Gaps.

MEMBER HOUTROW: Oh, gap. I'm sorry. I jumped ahead.

(Laughter.)

MEMBER HOUTROW: Okay. So --

CO-CHAIR SUSMAN: You tried to sneak that by us.

MEMBER HOUTROW: -- I'll save that thought.

So we have heard a little bit about the other methods by which there are ways to measure coverage and in the comparison that they did with those other measures. So they chose the ACS survey as the kind of gold standard to compare themselves to. And then two other measures of continuing coverage which they performed better than in their analysis.

And so I think this does help address a gap that they're seeing in the ability of states to actually demonstrate coverage in a more helpful way because we're getting at an issue regarding around whether children were appropriately covered when they could have been covered.

I do have a question. Very early in your -- you make a comment about good reasons

that people fall out of coverage and bad reasons that people fall out of coverage. So a good reason is that you're no longer eligible, but when you live on the very edge of eligibility -- as you were just saying -- you can fall in and out of coverage and it is actually probably negative. And in fact your coverage levels when you did your analysis were much lower amongst the children who were in the higher parts of the income group than the kids in the lowest end.

And I'd like to hear about how we're going to get better at this issue regarding good reasons versus bad reasons for disenrollment.

DR. SILBER: So good and bad refer to what you would really want as a meaningful statistic when you're evaluating a state. So you don't want to punish a state for a parent winning the lottery and then no longer needing to be on Medicaid. And so that's what we consider a bad reason -- that's a good reason to leave. It would be a bad reason to punish the state.

So I think that's what I -- that's the

terminology that we were referring to. And a good reason to punish the state would be that the patient really was eligible and they were not enrolled and the state needs to be punished for that. So that the family forgot to -- or didn't re-enroll when they should have and then they're kicked off the rolls, etcetera. So, okay. That defines the good and bad as what we meant.

Now again, you're going back to this issue of, well, on the margin you're on and you're off and you're on or you're off. And so we are saying that we're taking a random event, which is the event of appendicitis, which has nothing to do with your healthcare or whether you were enrolled or not enrolled. Then we're just arbitrarily going back four months.

Now the reason we do is because we don't want any back enrollment to affect the fact that you -- we didn't want people to say, well, you were enrolled. Look, they were enrolled when they had their appendicitis, but that's because you back enrolled them and that wouldn't be fair.

You're giving the states credit to something they shouldn't.

So we look at a random point in time four months before the random event and at that event you're either enrolled or you're not enrolled. So to us that is a measure of whether you had insurance or you didn't have insurance when you were eligible. So it might not be perfect, but at least it's random and there shouldn't be bias involved with that.

MEMBER HOUTROW: If we're thinking of this in terms of the percentage of children who are covered or however the rates -- when we're thinking about what the quality mark is for that, we do have this issue of knowing where a threshold belongs, because you don't want to punish the states for families who make too much money, but you don't -- what you want is for kids who need to be on the service to be on the service.

DR. SILBER: Well, I agree. I just don't think that this is punishing the states for

families that all of a sudden win the lottery in any way that's biased. Let's put it that way.

CO-CHAIR SUSMAN: So let's try to focus on gap. And, Kerri, do you have any further comments?

MEMBER FEI: No, I don't think I have -- I don't have anything else to add. I think what was presented to support the gap is sufficient.

CO-CHAIR SUSMAN: Kraig?

MEMBER KNUDSEN: Sufficient.

CO-CHAIR SUSMAN: David?

(No audible response.)

CO-CHAIR SUSMAN: Jeff?

(No audible response.)

CO-CHAIR SUSMAN: All right. Anyone else?

(No audible response.)

CO-CHAIR SUSMAN: So we're going to vote on gap.

MS. JUNG: The voting for Measure 3154 for performance gap is now open. Option 1, high,

option 2, moderate, option 3, low, and option 4, insufficient.

CO-CHAIR SUSMAN: And it looks like Jon is absent, so --

(Voting.)

CO-CHAIR SUSMAN: He has a lapse in coverage here.

MS. JUNG: Voting is now closed. The results are 45 percent for high with ten votes, 50 percent for moderate with 11 votes, five percent for low with one vote, and zero percent for insufficient with zero votes. The measure does pass for gap.

CO-CHAIR SUSMAN: Okay. And did you want to carry on since you were so enthusiastic with the reliability?

MEMBER HOUTROW: I'm excited about reliability specifications, so --

(Laughter.)

MEMBER HOUTROW: Okay.

CO-CHAIR SUSMAN: If you can get excited about it, it's all yours.

**MEMBER HOUTROW: Right. So produce consistent and then credible valid results. So this is all coming from administrative claims data, and you can look at the level of analysis at the state or potentially more regionally, and their interpretation as a higher score means better quality.**

**And as we know, there are children who age out, and so when they did their testing they made sure that the children who were in the aged-out age groups were not included because they would -- anyone who is 16 could be aging out. Everyone who is 17 to 18 would age out. And then of course as they used their appendectomy, they were thinking about specifically very rare events in the children under the age of two.**

**And we have heard a lot about how they have come to that, and this does require this related hospitalization. So we need to talk about whether we think this kind of use of the appendectomy is an appropriate surrogate. And I believe that it does help us get there.**

And in addition, we need to talk about the relationship to the different things that they tested. So the PE, the PI -- or the CI and the Community Survey as their gold standard.

And if there are any questions about how the data elements are defined and if the codes are appropriate and is the algorithm clear -- I think we had a good presentation of that and I'm feeling quite comfortable with this. And again, my only issue is relating around the children who are in the highest income group who will change eligibility, and they will have lower scores. And I don't know that that means necessarily lower quality for them. And so that's where I have an issue as the population of very poor are always going to be income-eligible, but those on the cusp are not, and I don't know that a higher score means better quality for those kids who are on the cusp.

CO-CHAIR SUSMAN: Okay. Let's go around to our primary reviewers first and then I'll get everybody else.

MEMBER FEI: Sure. So keeping in mind that this measure would more than likely be used for accountability, I'm not sure about the ability to really tell -- compare states, tell differences just because of the breadth of the confidence intervals and the overlap that you see.

I don't know that you can really discern state A is really doing better than state B, than state C, than -- you know. I think that would be a tough -- I think it might be tough to do. I'm not saying you can't do it. I think it would be tough to say somebody's really, really doing better than someone else. So maybe --

CO-CHAIR SUSMAN: It does seem like that's an important issue as defined by the algorithm, so if you have a comment --

DR. SILBER: Sure. I'm not quite sure why you're saying that it doesn't look like you can see that states are different. If you look on page 8 where we show the graph --

MEMBER FEI: Not different, but

comparable.

DR. SILBER: Well, we're talking about confidence intervals that often for many of the states -- or half of the states aren't overlapping with the other half. So we lined them up in order of their informed coverage, and you can see that they're -- if you look at -- we gave you the confidence intervals. We thought that was very strong evidence that there's going to be states that are very different than other states. So I'm not quite sure.

I mean, these are -- there are some states that are very small and that the confidence intervals are wider. So Delaware and Hawaii. And some states that are somewhat funny like Nevada, which is the very last one. But the great majority of the states have tight confidence intervals and they are -- and if you look at the graphs, they're not overlapping with each other. In many, many -- at least half don't overlap with the other half.

But furthermore, you don't -- it's not

just that a confidence interval has to not touch each other. You can easily show significance when they're still touching each other or even overlapping to some extent. So we thought this was just obvious that it was reliable and that it showed that the hospitals were different.

We did have this -- I mean, you can interpret it that way. I don't quite understand the comment.

MEMBER HOUTROW: Well, I guess I want to follow up on that. So I'm thinking of a state like Alabama where people have very low incomes. And then I'm thinking of a state like -- I'm just going to say New Hampshire. And I think they probably have not as low incomes. And if the rates of income are different in those states and the rates of chronic health conditions among children are different in those states -- so two things that we demonstrated are -- you've showed in your data.

So let's say Alabama has a bunch of very, very poor kids and a bunch of very sick

kids, and New Hampshire has less poor kids and weller kids -- which actually I think is true. And so you could then make a statement that because Alabama's continuous coverage is higher -- which relates to the poverty status and their health status -- that they have better quality than New Hampshire?

DR. SILBER: Okay. So first of all, what I would say first of all is that they might show better informed coverage. That's the first thing, right? So there is a difference. When you talk about adjusting for income or adjusting for race or adjusting for other characteristics, you get into a philosophical question regarding how do you evaluate the states. We looked at it. From our perspective the reason why we didn't put in these adjustments, from our perspective it was because we thought that states have to do a good job for whatever population they have. And so, that's why we didn't do those statistical adjustments.

So we can get into a philosophical

debate about whether we should have adjusted or not or whether we should add adjustment, but I didn't think that it was appropriate for what we thought the goal was as -- when we were asked to help develop this measure from the PQMP, but maybe I'm not understanding your question.

MEMBER HOUTROW: Yes, so you would -- you corrected me, and it's true -- and this is about informed coverage, right, so higher informed coverage. But the statement in this is that higher informed coverage is equal to better quality. And I think that's where we're struggling because as a quality measure, then not risk-adjusting for factors that we know influence it then makes it such that you can't successfully compare on quality between New Hampshire and Alabama. You can still assess them on their informed coverage, but not on the quality of that.

DR. SILBER: Well, put it this way. What I would do is say if I was a state that had a bad statistic on my informed coverage, I'd at

least want to study it and understand it. And if my excuse was that everyone was winning the lottery and somehow the statistic was off -- which I don't believe is going to be the case -- then okay. But at the very least it's a strong indicator that you might have a problem. And so, in our sense if I was a state, I'd rather have a high number than a low number.

And Illinois is an interesting example. We've looked at Illinois over time and then when they instituted a program about ten years ago that was strongly trying to increase their participation rates, they moved up by our statistic radically.

And so to me it is a measure of quality. Put it this way. Any time you have -- you don't look good on a quality metric, you always have to examine it and ask, well, is it real, is it not real? But at least I think it lets the federal government and the states say what's going on there? At least we want to understand it.

CO-CHAIR SUSMAN: Okay. I'm going to try to move this on. First of all, let me ask for any of the primary reviewers who had further comments or questions.

MEMBER KELLER: Yes, only a comment -- and I think this relates to the previous comment, I actually applaud you for including error bars because I always remember that every measurement that we take -- whether it is with one of our tools or with a ruler -- has error in it. And we forget that a lot. We tend to think these things are absolute numbers. They all have error, and I applaud you for making it explicit.

So I think that actually strengthens the measure's reliability, the fact that we can see what the error is depending on the size of the sample and various other things that we're working with. And I don't think that reduces its reliability at all. I actually thought that was great.

CO-CHAIR SUSMAN: Okay. Jeff?

MEMBER SCHIFF: Just a couple things.

I think that when I looked at the variation in states, I thought is this really a measure of the hassle to maintain coverage for individual kids and -- because every state has its way of doing eligibility, and I think that -- so in some ways my concerns are -- which was with the first term is about, presumed ineligible and -- it's almost like this is like the measure of the hassle of how much work it is to stay on coverage for families, because some states that may be -- we may think of as not having great coverage actually may make it easier for people to stay on or get enrolled, where other states -- and I'm looking at my state, which is in the lower end of this, aren't so -- maybe they aren't so good at that. So I think that's important.

It also may -- this is an editorial. It may be important if some of the proposals to reform the ACA about sticking to -- about if you lose coverage, you get penalized coming into effect.

So, and then the last thing to say is

it'll be really interesting -- and I'm not sure if you did this or not -- to look at this by race and ethnicity.

DR. SILBER: We did. That's in the appendix. And there are differences. In general what we're seeing is that blacks have a higher rate of coverage -- of informed coverage than whites in a lot of studies.

CO-CHAIR SUSMAN: Okay. Let's get the other cards, which seem to be proliferating. Jill?

MEMBER MORROW-GORTON: So I'm not at all surprised that blacks have a higher -- because they have lower incomes in general. I mean, I think -- I still think I have concerns about the potential for missing a systematic error in the higher income families who are bouncing in and out, tends to be seasonal work, tends to be -- and for a state who might -- that might -- and I don't know even -- I can't see where Massachusetts is in there, and I don't really care in the sense that if you're a state

that has higher incomes in general or you have a bigger population in those higher incomes that would be bouncing in and out, that's never good.

But the reality is doing Child Find for those kids for the months that they're eligible versus the months that they're not is a lot of effort and may not be that -- so I think you need to think about should you be -- it's -- this is not about winning the lottery. This is in fact about the guy that has a landscape business who makes $50 too much this month and doesn't -- his kids don't qualify for Medicaid. That's what we're talking about.

DR. SILBER: All right --

MEMBER MORROW-GORTON: But I think that, one, you need to think about that because to tell a state that they're doing a really bad job at Child Find when it's that kind of an issue is a disservice in the sense that that doesn't help the state find kids that are not on Medicaid that should be on Medicaid.

CO-CHAIR SUSMAN: I'm going to ask

that we --

DR. SILBER: May I respond?

CO-CHAIR SUSMAN: Just a second. I'm going to ask would you really try to focus on the reliability issue here.

Briefly please respond, but I want to try to bring this to a closure.

DR. SILBER: I probably didn't make it clear enough, but bouncing in and out is not going to adversely bias our metric. It's a point in time. We don't bounce in and out for a point in time. The point in time is exactly four months before you got the random event of appendicitis. We don't look to see whether you did or didn't have enrollment on the day that you had appendicitis. A random point for that patient who we knew was eligible at some point. That random point is what we use and we simply ask were you enrolled.

So I don't think that we're biasing the metric by using the point in time prior to the appendicitis. We could be adding some

variability, and what you're saying is maybe there's more variance, but I don't think we're biasing the metric.

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

MEMBER DORSEY: So, yes, I just wanted some guidance on the reliability vote in particular. I noticed reviewing this that the sort of preliminary guidance about reliability voting on the sheets was low. And my question was about the test that was used to assess reliability and whether that was the issue, right? That this is about the tests had shown as point estimates for the states and the error bars around them and whether or not we're -- our discussion should be is that an adequate test of reliability? That's a question to NQF.

DR. NISHIMI: So the PA focused on the ability -- because it was at the score level -- to distinguish among the states, and the issue that Kerri or someone pointed out in terms of the overlap. So for accountability purposes the

assessment was that the neighboring states that are clustered in the middle there -- the score may not be a reliable indication when you started making comparisons. And so that's what led to the low rating. But again, that's the staff's best assessment at the time and the Committee members need to make their own.

CO-CHAIR SUSMAN: Okay. With that explanation and further comment, anything else standing between us and voting?

(No audible response.)

CO-CHAIR SUSMAN: All right. Then let's go ahead on reliability.

MS. JUNG: The voting for Measure 3154 for reliability is now open. Option 1, high, option 2, moderate, option 3, low, and option 4, insufficient.

(Voting.)

MS. JUNG: Voting is now closed. The results are four percent for high with one vote, 48 percent for moderate with 11 votes, 39 percent for low with nine votes, and nine percent for

insufficient with two votes. So with that we did not reach the 60 percent threshold, so consensus has not been reached for reliability and we continue to vote.

CO-CHAIR SUSMAN: We continue on, right? So we're in that intermediate range, if you will.

Okay. So let's move on to validity.

MEMBER HOUTROW: So in validity, as we had heard before, they used the ACS gold standard and that was highly correlated showing good numbers. And if you look on -- what page is this? There's a table. Table 2, which is their Pearson correlations of the continuity rate and the duration, as well as their own in comparison with the gold standard in each of the two. And you will see there informed coverage has -- with the continuity rate, appears in correlation of 0.77. With duration it's lower, 52. And with ACS it's the highest, 46. And similarly they look good in comparison to the other measures against the gold standard of the ACA.

We need to talk briefly about exclusions. We've talked about the 18-month period, which means you do need to do some exclusion for some 16-year-olds and all 17-year-olds. And in the appendectomy analysis that they did, they actually dropped 15 percent of them. And I would assume that most of those are the older kids, not the under age two kids. But that -- more than 15 percent is higher than a ten percent rate, which would be something I would feel a little bit more comfortable with.

We need to talk about if they're -- these exclusions are consistent with the evidence. I think from the perspective of you need to not include people who are going to age out and the under the age of two for appendectomy that makes sense, and if there are any issues regarding the impact of that exclusion. And I think we don't need to do potentially any additional empirical testing, although I would ask is it mostly the older kids, that appendectomy for that 15.6 that were excluded?

DR. SILBER: Yes.

MEMBER HOUTROW: Okay. We have already looked at the confidence intervals in the graph of informed coverage by state and the comparison there. And then the issue regarding missing data, which is 12 percent of the states weren't able to provide data.

So I think it's very straightforward to get appendectomy codes both in ICD-9 and ICD-10. That's good. It seems to be relatively clear. There are some issues with the states, but I think in general this looks quite good.

And then we've talked a little bit already about the variability in scores and the concern.

CO-CHAIR SUSMAN: Kerri?

(No audible response.)

CO-CHAIR SUSMAN: Okay. Any of the other reviewers have further comments on validity?

(No audible response.)

CO-CHAIR SUSMAN: Seeing none, anyone

else?

(No audible response.)

CO-CHAIR SUSMAN: Let us vote.

MS. JUNG: Voting for Measure 3154 for validity is now open. Option 1, high, option 2, moderate, option 3, low, and option 4, insufficient.

(Voting.)

MS. JUNG: Okay. Voting is now closed. The results are zero percent with zero votes for high, 74 percent for moderate with 17 votes, 26 percent for low with six votes, zero percent for insufficient with zero votes. And with that, the measure does pass the validity threshold.

CO-CHAIR SUSMAN: Maybe I'll ask Kerri to start us off just to give you a break.

MEMBER FEI: I don't want to steal all of Amy's thunder, you know?

CO-CHAIR SUSMAN: Yes, I know.

(Laughter.)

MEMBER HOUTROW: I'm not so excited

anymore.

MEMBER FEI: Did you lose your excitement? You used it all up.

CO-CHAIR SUSMAN: Feasibility.

MEMBER FEI: Feasibility. I don't really see any real problems here. I think the problem isn't with the measure. It's with perhaps the data submitted to the MACs database, which is out of the control of the measure developer, so I don't want to make a big deal about that. I think if you were going to use this on a national level, efforts to improve the database would help this measure immensely, I would think. So, that's an aside.

But other than that, that's all I have.

CO-CHAIR SUSMAN: David?

MEMBER KELLER: Yes, I mean, I agree that the MACs database -- data set had some troubles that they clearly identified in their development of the measure. I think that it is in the interest of both the payers and the

providers to make sure that appendectomies are appropriately billed and paid for. So I think that there is an incentive to get that data in place that you should be able to get from within a state Medicaid agency if you were trying to report on it.

CO-CHAIR SUSMAN: Kraig?

MEMBER KNUDSEN: I was just surprised by the missing data, actually.

CO-CHAIR SUSMAN: Yes. Jeff, anything further?

MEMBER SCHIFF: Nothing specific, except I guess one of the questions I had is whether this will be a measure that the state officials will want their folks to do, which would require some technical assistance for when it's done nationally through the MACs or some other national set.

DR. SILBER: Well, a couple things. We actually have done the confidence intervals -- which would have been the technical side of this -- two ways. We did bootstrapping, which we

didn't think the states could do. And then we did approximate formulas, which are in the appendix, which the states could just apply in a SAS program. And they are almost exactly the same. They give -- to the last second or third decimal point they give the same results. So we made this as user-friendly for the states as we could. That's why the appendix has the confidence interval formulas in there.

CO-CHAIR SUSMAN: Amy?

MEMBER HOUTROW: No, I'm good.

CO-CHAIR SUSMAN: Nothing to add?

MEMBER HOUTROW: One time ever.

(Laughter.)

CO-CHAIR SUSMAN: We know you're good, but we're -- okay. Jill?

MEMBER MORROW-GORTON: Okay. So a quick question. Knowing that the MACs data is often not great and you lost six states just based on this, is this something that could be done in a state MMIS system or a -- or so that it could be done at the state level maybe not using

the MACs data, but the state using their own data?

DR. SILBER: So to us the big issue was whether the managed care patients -- when they get appendectomies, do we see bills. And so, if the states know that they can see a bill from that, then -- and often they do have bills. So when they -- if we know that the bills will show up, then great. Then we can -- then a state can -- and if a state knows that, they can use that data. We have in the appendix the test that we use to see if the state had adequate managed care bills that were ending up with appendectomies.

So the way we did that -- and it's explained in the appendix -- is we looked at all the appendectomy cases and then we know whether they were managed care or not. And then we matched every appendectomy case to ten kids in the state, same age, same sex, who didn't have appendectomy. They just had a data set. They were just in the data set. And then we looked at

their managed care rate.

If we found that the managed care rate in the kids who weren't hospitalized but just were in the system was too much higher than the appendectomy rate kids, then we said the state had a problem. And that's how we -- so we had a filter to get rid of states that had inadequate -- that looked like they had problematic data.

So a state can try to figure that out. They can ask themselves whether their bills are ending up or they can run our filter which we put in the appendix.

CO-CHAIR SUSMAN: Thank you. Okay. Any other comments about feasibility?

(No audible response.)

CO-CHAIR SUSMAN: So if it's feasible, let's vote.

MS. JUNG: The voting for Measure 3154 for feasibility is now open. Option 1, high, option 2, moderate, option 3, low, and option 4, insufficient.

(Voting.)

MS. JUNG: Voting is now closed. The results, four high with one vote, 96 percent for moderate with 22 votes, zero percent for low with zero votes, and zero percent for insufficient with zero votes. And this measure does pass for feasibility.

CO-CHAIR SUSMAN: Usability. Kerri, why don't -- you did such a good job.

MEMBER FEI: Oh, sure. She gave me the easy part.

No, as far as usability I guess from a -- and my -- and it's addressed in the documentation using it for accountability I think is always the big discussion. Knowing that that hasn't been determined yet, I would like to see kind of an evaluation of how that could perhaps take place only because that is kind of the end result of where people are going to want to take this. And then I question if it really should be used in a traditional accountability fashion.

I don't -- I'm not sure this is a measure we should be either rewarding or

penalizing based on performance, but I think it's a necessary thing to know the results in order to improve the enrollment for kids on that for Medicaid. So I guess that's my struggle is whether -- it's an awesome -- it's a great measure. I think it's really innovative and novel. I don't know that using an accountability program would be best use of its results is my concern.

CO-CHAIR SUSMAN: Fair enough. Other comments? David?

MEMBER KELLER: Yes, I mean, it certainly would have -- I struggle with this with actually all of our measures is how they will be used, and I think that any accountability measure would require a state to develop a feeling for how this measure works and some time and energy to go into it.

That having been said, I actually think it should be an accountability measure eventually once they do get used to it, because the issue of continuous coverage and churn is

just -- is so important at the state level. So I think it's -- but I think it is -- in its current form the developer has made it as usable as possible and the question is, like all great power, will they use it for good or for evil?

(Laughter.)

CO-CHAIR SUSMAN: Wow. How could you possibly follow that? But I'll call --

MEMBER KELLER: I have to quote Spiderman whenever possible.

CO-CHAIR SUSMAN: Yes.

MEMBER FEI: I mean, and that kind of echoes my concern as well. Yes, I think it needs to be used -- used the right way.

CO-CHAIR SUSMAN: Kraig?

MEMBER KNUDSEN: Working for the state government, comparing states, it's kind of like an art of sorts. People look at -- when we're looking at ourself and we go, oh, well, California is doing this or Hawaii is doing this. And then somebody always in the room is always saying, yes, yes, that's not us. So we're not

comparing ourselves to California or Hawaii. How about Michigan, or something like that? So it's useful in terms of a state -- your own state comparison from year to year, but not necessarily across all states.

CO-CHAIR SUSMAN: Jeff?

MEMBER SCHIFF: I think this would be an interesting measure to present to the National Association of Medicaid Directors to say what's going on?

I do think though that there's a marketing issue. I think it should be -- almost be called the Coverage During Random Appendicitis Measure, because I think that's where it really -- that's the strength of it and it gives people the sense then of whether or not they're -- what they're -- because it's really -- I don't want to go back to my hassle-of-coverage kind of issue, because I think that's where the accountable entity is the enrollment -- or the goal of doing enrollment in Medicaid.

CO-CHAIR SUSMAN: Amy and then I've

got a couple of other comments here about usability.

MEMBER HOUTROW: I think they've spoken to the things that I was concerned about.

CO-CHAIR SUSMAN: Okay. Jill?

MEMBER MORROW-GORTON: I just have a great -- this would be a great state self-assessment tool.

CO-CHAIR SUSMAN: Yes. Well said. Carol?

MEMBER STANLEY: Yes, so if I were to put my hat on in my previous role with the state Medicaid, I would see this as a valuable structural measure, especially measuring how good of coordination we have -- the Medicaid agency and Department of Social Services has in maintaining enrollment. And I could see the outcome of this measure being something to leverage getting more resources and coordination going.

But one of the questions I have relating to that is in Virginia if a child is

enrolled in Medicaid or CHIP, they have coverage for a full year regardless of any churning. And I don't know how that -- how it is in other states, if that's a federal law or if -- so -- and then the -- another thing, now that I work for -- in a provider setting sometimes and have actually observed children coming in where their coverage has lapsed and the dilemma that the provider faces, this would be a valuable tool to pitch to the state. Look at how you're doing and how it's affecting us.

So one of the questions I have as far as usability are as far as churning between Medicaid and CHIP -- and we saw that a lot during the recession -- so when you factor in the calculations if a child moves between Medicaid and CHIP, is that counted as -- how does that come to fruition in your measurements?

DR. SILBER: So if we have evidence that they're eligible and they're enrolled, then they're in and the data sets -- to the extent that they pick up both of those, then we have

that information. So we're not -- it's -- again, it's a point in time --

MEMBER STANLEY: Right.

DR. SILBER: -- so we're not thrown off by a change as long as they have something.

MEMBER STANLEY: So if they're in Medicaid in your snapshot today, but they were in CHIP six months ago, they're still going to count.

DR. SILBER: That wouldn't make a difference, no. It would be a point in time.

CO-CHAIR SUSMAN: Jon?

MEMBER FINKELSTEIN: So what I'm going to say I'm not suggesting should be determinative, but I think it's important. So under use and usability for me a perfect quality measure is easily understandable by all stakeholders. So we read it ahead of time. I really delved into it. We've had 40 minutes with the developer. And if each of us was forced to take a whiteboard and explain it to five state senators, none of us would get it exactly right.

I'm not sure many of us would get it mostly right. I'm talking about myself. So that doesn't make it not technically great and worthwhile. And other things may outweigh this, but it is not an easily understandable measure.

CO-CHAIR SUSMAN: I mean, on the other hand, what I'm hearing from some of the folks who deal with these issues regularly on a policy or state level is that it makes sense to them. And I think that's important, because there are going to be different audiences for these sort of measures. And I would concur. I couldn't do a very good justice to this, but for someone who's running a Medicaid program it may be very valuable.

So, unless I see any other hands up or cards up, let's go ahead and vote on usability and use.

MS. JUNG: The voting for Measure 3154 for usability and use is now open. Option 1, high, option 2, moderate, option 3, low, and option 4, insufficient.

(Voting.)

MS. JUNG: Voting is now closed. The results are four percent for high with one vote, 74 percent for moderate with 17 votes, 22 percent for low with five votes, zero percent for insufficient with zero votes. And with that, the measure passes for usability.

CO-CHAIR SUSMAN: Okay. So any final comments before voting?

(No audible response.)

DR. NISHIMI: We don't vote on overall.

CO-CHAIR SUSMAN: Oh, we don't?

DR. NISHIMI: Because it didn't get --

CO-CHAIR SUSMAN: Oh, it didn't pass -- the one was indeterminate? Okay.

So what do we do? We pray.

DR. NISHIMI: So the developer will have the opportunity to bring additional information. You'll also receive comments from whoever chooses to comment on the measure. Those will come back to the post-comment call. You'll

consider the reliability issue, which is the criterion for which you did not achieve consensus. If you achieve consensus and it passes, then we will take an overall suitability for endorsement.

CO-CHAIR SUSMAN: If it doesn't?

DR. NISHIMI: If you don't achieve consensus or it fails, then it just stops.

CO-CHAIR SUSMAN: Okay.

DR. SILBER: Just a general question just for the process, because when we were talking to NQF, they instructed us that this was what we needed to do. And to us it showed clearly that some states were very different than others. So I would request as close guidance as possible.

I'll do whatever this Committee wants, but I just need to know exactly what it is that they want. And right now it's not clear to me because it still in my mind looks clear that some states look different than others, but maybe I'm missing the point. So I'd just request some help

if possible so that I can answer your questions.

DR. NISHIMI: I don't think it's a matter of analyses. I think it's a matter of how to interpret your graph. You -- we asked you for the graph and you produced exactly what -- and I think what we heard today -- and I don't want to put words in the Committee's mouth -- is that for purposes of accountability the fact that the states that were clustered around, you wouldn't be able to tell one better than the other. And I think that's --

DR. SILBER: For any distribution there's always going to be a cluster in -- you know, somewhere, and that's going to be ones that are different than others. So again --

DR. NISHIMI: Yes.

DR. SILBER: -- I don't understand it, but I will --

DR. NISHIMI: I think about --

DR. SILBER: -- work on --

DR. NISHIMI: -- half of them weren't and not.

CO-CHAIR SUSMAN: Yes, I mean, does the Committee want to weigh in on this question? Amy, you have -- your hand is --

MEMBER HOUTROW: I don't know if it's appropriate here, but I would have loved to see some potential risk adjustment for things that you knew influenced the rates.

CO-CHAIR SUSMAN: Yes, Jeff?

MEMBER SCHIFF: I don't know if it would -- I don't know if it's exactly the same thing, but it would be -- there seems to be such variability that I can't explain and why the cluster is where it is, but I wonder if there's some work that could be done to understand what the -- what explains the variation. I think that's -- that's not about the fact that it's not reliable. It's about the fact that there's some sort of special cause variation that I can't get at.

CO-CHAIR SUSMAN: Other comments?

DR. SILBER: Pardon me?

CO-CHAIR SUSMAN: Yes, sure. Please.

DR. SILBER: A clarifying question. When you say the cluster, what cluster -- or what figure are you referring to?

CO-CHAIR SUSMAN: It's on -- yes, page 8.

DR. SILBER: So the cluster -- we've lined them up so that they'll look close next to each other, because they're --

(Simultaneous speaking.)

MEMBER SCHIFF: And I think what -- I guess what I would say is where the -- it's probably not a good -- cluster is not the right word. Where the confidence intervals maximally overlap is what I guess I would say is the cluster, but then there's the -- like the Nevada or whatever one is very low. So it's really about -- I guess it's really about trying to understand why there is that variation, maybe. But that's maybe beyond the scope of this.

DR. SILBER: And that's part of reliability, I guess. I'll try to answer that.

CO-CHAIR SUSMAN: I think --

DR. SILBER: That's part of validity I guess.

CO-CHAIR SUSMAN: -- there is clearly some use internally for a state to look at this sort of data. The question is if this is for accountability, can we reliably say that the state that is on the right is really better than the one next to it and the one next to it and -- not necessarily that it's ordered or 100 percent, but you have to go pretty far along before -- and there's pretty wide confidence intervals. And of course that's going to happen in any distribution to a certain extent, but there are also a number of states that have very wide confidence intervals and it may well be just the number of incident cases that are there to analyze. But I think those are the sorts of issues that anything you can do to make the Committee more comfortable would be helpful.

And I'll leave the last comment to Jill.

MEMBER MORROW-GORTON: I just had a

couple of thoughts. One would be think about stratifying income levels and looking at what those numbers look like, because states look very different.

And the other thing you might think about is looking at eligibility at both of those two points in time to get a sense of how many of them are the same -- so those kids would have been eligible that whole time period -- and how many of them are different?

DR. SILBER: That question -- I'll have to explain it better in the next try. The question you're asking is -- I've already addressed and I thought I had. So it has to do with a random point in time and it's -- so I'll try to explain it better.

CO-CHAIR SUSMAN: Carol, did you have any final word or just -- okay. Well, I want to thank everybody, and particularly the measure developer who I think did a tremendous job of trying to make a somewhat difficult series of concepts well -- easily understood.

And we have regressed to the mean. We're right on time. And we have scheduled a break. Then we'll reconvene at a quarter to.

Do you want to give any guidance, John, about our next task or --

CO-CHAIR BROOKEY: No, but we have five measures to do between now and going home, so just think about that. So see you back in 15 minutes.

CO-CHAIR SUSMAN: All right. Fifteen minutes. Make it sharp.

(Whereupon, the above-entitled matter went off the record at 2:32 p.m. and resumed at 2:45 p.m.)

CO-CHAIR BROOKEY: So I first want to start out by getting a head count of those that are going to be able to stay the long haul. I know some of you have 7:00, 7:30 flights. We had asked people to stay until 5:30, but I know that's not possible for everyone, so let's just start out by asking how many can stay until 5:30, if necessary.

So 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

Okay. And how many -- let's just say if I said 5:00, how many can stay until 5:00? How many more -- how many who said no could stay until 5:00? Does that help anybody? That helps one.

So why don't we just see how it goes. And these five measures have a lot of overlap, and I think if we can get through the first one or two, the remaining measures should be pretty easy to do. And if -- worst case scenario we'll do it on the follow-up call. We've already committed to a phone call, so I don't think that we should beat ourselves up if we have to leave without getting them all completed, although last year we got them all done before we left, and that was a nice feeling.

The developer is on the phone or in person?

MS. THEBERGE: On the phone.

CO-CHAIR BROOKEY: On the phone?

DR. SOLLOWAY: Hi, this is Michelle Solloway, and I am on the phone, as is my colleague Naraa Gombojav.

CO-CHAIR BROOKEY: Okay. So let me just ask you before you get started, do we -- first of all, do we have everybody in the room? Are we all here? I think we are.

So I would ask you to help us a little bit. We're not going to vote on these as a bundle of measures. We're going to vote on them individually. On the other hand, there are obviously a lot of commonality between these. And so, if you would give us a brief overview of sort of the set of measures and just very briefly touch on the points that are -- distinguish these measures in terms of why they're different one to the next. Just five minutes. But if you could touch on those topics, it would be really helpful.

DR. SOLLOWAY: Sure. Thank you so much, and thank you for inviting us on the call. And I first of all want to just express my thanks

to Suzanne and her staff for really helping us through this process. Christina Bethell -- who actually was the original developer of this -- was not able to be on the call today, so I just wanted to first of all say thank you to everyone for all your help with this, because this is the first time I've gone through this process. So please bear with me if I don't understand all your procedures.

So I want to say that all of these measures -- 3219 to 3223 -- are all part of one tool, the Promoting Healthy Development Survey, which was previously endorsed by NQF. So these are different sections of the same tool. And the way this tool is used is it is a survey that is sent by providers to the families after a well-child visit to find out if the providers actually provided the well-child care in accordance with national guidelines.

And these different components, different measures, as you will, are -- represent different pieces of standard national guidelines

for well-child care. So anticipatory guidance, asking about parental concerns, family-centered care and family assessment for psychosocial screening and behavioral health and home safety.

So all of those are different aspects of national guidelines that are recommended by the American Academy of Pediatrics. They're all based -- all these measures are based on those recommendations. All these measures are recognized by the National Quality Measures Clearinghouse as being measures that meet basic criteria for use as standard indicators for healthcare quality for children.

Nothing has changed about these measures since they were endorsed as a -- in total as a tool, as the PHDS tool previously by NQF. So that's why they all -- that's why the applications kind of all look very similar.

So the specific details on the validity and reliability might be different because of how the different measures are constructed, but they're really all meant to be

part of and taken together as an assessment of whether children are receiving well-child care in accordance with national guidelines.

Also, several of the items are actually in national surveys, so all of the items in the Anticipatory Guidance Measure, 3219, are in the National Survey of Early Childhood. Asking about parental concerns and family-centered care are also in the National Survey of Children's Health.

So -- and I also just want to say that the development of the PHDS tool and all these different measures were -- there was a very, very rigorous process a number of years ago that really involved experts in measurement development and research and child health and quality measures that really all came together from expert panels. It was followed up with testing with providers and testing with parents.

And so the ways in which these measures were initially developed went through a very, very rigorous process in which the

reliability, validity and feasibility were all very much flushed out and really done in detail. And also these are measures about which providers can do something to improve the quality of their services. So nothing is included here that is not actionable. So it's all very actionable.

And I think I'll stop there and see if you have any questions for me at this point.

CO-CHAIR BROOKEY: I think that's a good introduction. And so we'll launch into this discussion. We're first going to talk about the evidence. This is an outcome measure.

And I think, Sue, are you going to kick off?

MEMBER KONEK: Yes.

CO-CHAIR BROOKEY: Okay.

MEMBER KONEK: This is for 3219, Anticipatory Guidance and Parental Education. And I'll say a little bit more since she did a brief overview.

This is a -- this measure is used to assess the degree to which pediatric clinicians

discuss key recommended anticipatory guidance and parental education topics. Necessarily, anticipatory guidance questions vary by child age, and of course we understand that.

This is an outcome measure. The measure proposes a method to determine how well the providers are meeting the goal of providing anticipatory guidance and parental education.

Through a survey of parents post-well-visit, parents will indicate whether their anticipatory guidance and parental education are met. Providers reviewing results of the surveys grouped in ten or more surveys will learn how effective they are in meeting this goal. Then they can be trained in AAP Bright Futures training that's available addressing this topic and subsequently improve the AGPE to better meet patient and family needs.

Do you want me to go more into evidence?

(No audible response.)

MEMBER KONEK: Evidence was presented

that both parents and clinicians felt that anticipatory guidance was the most important point of discussion at well visits. Evidence was also presented to demonstrate that by using a survey tool followed the -- I'm sorry, the PHD Survey -- following that they could use a PDSA QI work, do that work, providers are able to achieve statistically significant improvement in providing this AGPE.

Parents are more likely to report their needs for anticipatory guidance were met in follow-up assessments and surveys and to note that it's an ongoing process, so improvement can be measured. And that's the evidence.

CO-CHAIR BROOKEY: Before Jim goes -- and maybe Jim can answer this question. I was looking for the actual survey instrument. I was --

MEMBER KONEK: Do you want me to -- oh, do you want us to read that? I have it.

CO-CHAIR BROOKEY: Well, the question I had is as I recall this is --

MEMBER KONEK: It's a patient --

CO-CHAIR BROOKEY: -- it's the parent's recall of the --

MEMBER KONEK: Yes, the patient reported outcomes.

CO-CHAIR BROOKEY: -- conversation between the --

MEMBER KONEK: Yes, it's the patient reported --

CO-CHAIR BROOKEY: -- pediatrician -- but was it a laundry list of anticipatory guidance items?

MEMBER BOST: So there's three questions. Well, it's -- there's three different age groups that have slightly different questions for this domain.

MEMBER KONEK: Different questions in each group.

MEMBER BOST: And there are three questions with sub-parts to them. So there can be a total of like 16 items.

And this particular domain -- in the

measurements and guidelines for it, they talk about a few different roll-up measures. What I believe we're talking about for this one is you look at all 16 items and you see whether or not for each of those items the parent said yes, I received some instruction or no, but I didn't need instructions. And then the measure for the physician is the percent of parents who said those two things for all the items.

DR. SOLLOWAY: Right, and I would just like to mention that there's really three sections, so there's five or six questions on physical health, there's six questions on behavior and language and learning, and then there's three questions on injury prevention. So you can actually just have scores by those three sections, if you want. But you are correct, the way you measure it is the percent of parents who said yes or I didn't need this information to all of the questions. So, yes, that's correct.

MEMBER BOST: So that's the way it's measured. And I don't really have anything more

to add to evidence, just some clarification on how the measure is actually done.

CO-CHAIR BROOKEY: So the evidence -- well, the evidence is whether this is actionable by the pediatrician who gets the results. And theoretically, if I get kind of a report card back, I'm going to modify my approach to well-child care. Is that how we look at the evidence for this particular one?

MEMBER KONEK: That's it.

MEMBER BOST: And they do mention in one of their studies that it was actually done and they did like a training program at one of the institutions and it showed improvement.

CO-CHAIR BROOKEY: Yes, okay. Jeff?

CO-CHAIR SUSMAN: So just to clarify, let's say there are 15 measures -- just to make it easy -- and I do 12 of them, I will not meet the criterion. And it's also then supposing that if you do 15 and I do 14, that there's some difference in quality that is supported by evidence since this is what we're talking about,

the evidence.

MEMBER BOST: But that's not this measure, like you said.

There's another way you could do this. You could look at what was the percent of those 15 items that the parent said yes and then take that average across all parents, but that's not what this is.

CO-CHAIR SUSMAN: Right. But I mean, it's suggesting that the outcome measure -- that this 15 out of 15 for argument's sake, is really a significantly better outcome than my doing 13 out of 15, if I'm following correctly.

MEMBER BOST: Right. Absolutely. Yes.

CO-CHAIR BROOKEY: Do the doctors get an itemized report or do they just get that sort of pass/fail for each section?

MEMBER QUINONEZ: I mean, it seems like --

DR. SOLLOWAY: Okay. I was on mute. Sorry about that.

MEMBER QUINONEZ: -- the numerator and denominator, it means you just -- you either did it or not for all 16 items.

CO-CHAIR BROOKEY: That's kind of what I thought as well, but go ahead, developer. Go ahead and clarify.

DR. SOLLOWAY: That's correct. It is really about all of the age-appropriate items because all of them are recommended to be given as anticipatory guidance. So that's the benchmark.

The way the providers get the report, they send this out and then they can generate a report. When they have at least ten surveys they can generate a report, but because we guarantee parent confidentiality, if for example there isn't enough data to maintain that confidentiality, it wouldn't show up as an item-by-item list. If there is enough, we do show that.

And actually, we -- what we do is we show in -- we have the three different domains,

physical health, behavior -- behavioral health and injury prevention. We show those three specifically, like what their percent in each of those domains, but then we also have a chart which is by age and by each of the questions.

And we put in red -- we show them in red which ones they didn't meet, but we can't give them the numbers unless there's really a lot of surveys. So they get the information to see what they're not doing well, but they don't necessarily get the metric on it if we can't guarantee parent confidentiality, if that makes sense. Does that make sense?

CO-CHAIR BROOKEY: So you're saying there's some volume threshold at which you would give them more specifics?

DR. SOLLOWAY: Right, because we're guaranteeing parent confidentiality in this report. So in other words, if they don't have enough surveys -- if they would know for example -- if they could remember what they did or go back in their records and look and they can

identify specific parents that they talked about specific things with, we have to make sure that we guard that confidentiality. So, yes.

CO-CHAIR BROOKEY: Yes, I get it, and I think it's a validity question, too.

DR. SOLLOWAY: Right.

CO-CHAIR BROOKEY: But it's also an evidence question as to whether you can actually act on the information, because that's really what we're looking for is an outcome --

DR. SOLLOWAY: Right.

CO-CHAIR BROOKEY: -- metric. And to say that --

(Simultaneous speaking.)

DR. SOLLOWAY: And we show them in red. Like, so we have a chart that has every item by age and by -- so that we highlight it in red if they didn't -- areas where they didn't make the cut, where they didn't ask those questions. So it gives them information about what they could do to improve.

Like for example, if they aren't

talking about car seat safety or burns or how to keep your child from getting burns, we can highlight that in red and then they could say, oh, we better get -- we better tell our people to ask about those questions.

CO-CHAIR BROOKEY: Yes.

DR. SOLLOWAY: Or make sure that they fill out some form where we've asked those questions, because it looks like we're not asking that enough.

CO-CHAIR BROOKEY: Yes.

DR. SOLLOWAY: So they --

(Simultaneous speaking.)

CO-CHAIR BROOKEY: And just to help you out, when you say they, you're really saying you, because I'm a pediatrician. So --

DR. SOLLOWAY: Oh, I'm sorry. Yes. Yes, you.

CO-CHAIR BROOKEY: So that's why I'm asking a lot of questions. But let's move on to -- let's see. Well, we have Rajiv and -- is Martha -- Martha's here, right?

Oh, there you are. We haven't heard from you. So do you want to comment?

MEMBER BERGREN: Yes, the only thing I was going to say was that there were several -- and I'm repeating a little bit that there were identified ways to improve the quality that had been demonstrated. So that's it.

CO-CHAIR BROOKEY: Okay. So you're saying that that supports the evidence part of this is what you're saying?

MEMBER BERGREN: Yes.

CO-CHAIR BROOKEY: Okay. Rajiv?

MEMBER MODAK: Nothing more on that.

CO-CHAIR BROOKEY: Okay. Any other comments from the group about the evidence? David?

MEMBER KELLER: So this is actually by way of a question because within -- I know Bright Futures pretty well and I know the folks who are on your Advisory Committee, and one of the challenges is that within Bright Futures, it recommends that providers focus on three to five

of the possible items at each visit, the notion being that over time you will end up covering all of these things. And I've been hunting through because I didn't do a close review of this to see what time -- when you're asking these questions, are these questions related to a particular visit or to a series of visits over time?

CO-CHAIR BROOKEY: And I wonder if that's a feasibility question, but go ahead and answer the question briefly. I think that may be more to feasibility and usability, but, developer, you want to comment?

DR. SOLLOWAY: Sure. So really it's about the survey is sent out and they're asking about the last visit.

MEMBER KELLER: The last visit? Okay. Thanks. That's all I needed to know.

DR. SOLLOWAY: Yes.

CO-CHAIR BROOKEY: Yes, right.

DR. SOLLOWAY: Yes.

CO-CHAIR BROOKEY: Any other -- Jim?

MEMBER BOST: So the fact that they

get that itemized list of what's red and not, they -- a provider doesn't really need this rate to know where they need to improve. So I would ask the developer what other uses will this rate have?

DR. SOLLOWAY: Well, you mean the overall scoring rate?

MEMBER BOST: Right. I mean, you can use it to compare or benchmark yourself, or --

DR. SOLLOWAY: Yes.

MEMBER BOST: -- do you guys use a cutoff that you think means you need to do something or --

DR. SOLLOWAY: Well, what we do is when we provide the information back to the providers or to you, what we would give you is here's the national standard, here's the national average and here's your score. And then the idea is that you can use that in a PDSA quality improvement cycle for which you can get Maintenance of Certification Part 4 credit to -- if you go through a process showing at two

subsequent time periods that you've implemented some activity to improve anticipatory guidance and that -- and then you have subsequent scores. So in other words you would use this survey at time one to get the baseline data, at time two and time three showing that you could improve over time.

CO-CHAIR BROOKEY: Okay, we don't want to talk about MOC though. That is a whole other discussion.

So, the question is about evidence, and it sounds like there is at least two or different things that can be done with this data. Does anyone have any other comments about the evidence for this particular measure or should we go to vote?

Jill.

MEMBER MORROW-GORTON: Just knowing that parents don't always hear what you say and they don't always remember, and depending on the timeframe between the visit and when they got the survey, do you have any evidence about or any

look at -- and maybe this is reliability, but look at what actually was talked about in the visit measured in some other manner and what parents remembered?

DR. SOLLOWAY: That is a really good question. And I have to say that I was not at CAHMRI at the time that this was developed. I would imagine and suspect that there may have been a little of that validity checking up on to see about that but I don't know. Naraa, do you know if that was done?

DR. GOMBOJAV: Yes, you know, I haven't seen at the time when the issue was up but I know from the previous literature using PHDS survey there are many literature searches being done to develop the survey. So we can -- I think we provided some of this literature found on these topics that --

CO-CHAIR BROOKEY: Okay, so it sounds like no. But I mean there are other things that are out there that depend on coding. So you can blow in a B code for dietary advice and

counseling about screen time and so forth, whether you did it or not, right, especially in Epic.

So this is getting to, really, hopefully what really happened or at least what the parent perceived happened. So you can decide yourself whether one is better than the other, I suppose.

Any other about evidence? David.

MEMBER KELLER: So just a quick question. How long until the families get the surveys? What's the timeframe?

DR. SOLLOWAY: The provider can send it out whenever they want. So, it's really up to the provider to make that decision. So they could do it at the end of every visit. You know, they could just have their staff send it out, or if it's available on a patient portal, because that is really up to the provider. We don't have any control over that.

CO-CHAIR BROOKEY: Okay, that may not affect the evidence discussion.

Any other questions or comments about evidence? Okay, I would recommend that we go to vote.

MEMBER EDIGER: I actually have a lot of thoughts about this.

CO-CHAIR BROOKEY: I'm sorry. I didn't see you down there.

MEMBER EDIGER: That's okay. So, I am one of the lead discussants on one that's further down so I am happy to save a lot of my thoughts but it seems --

CO-CHAIR BROOKEY: No, go ahead and go now.

MEMBER EDIGER: -- like they are relevant to all five.

CO-CHAIR BROOKEY: Yes, go ahead.

MEMBER EDIGER: So, I'm thinking about what you just asked about, like do parents hear what the checklist that the pediatrician's going through? And an example for me is, as I mentioned, I disclosed I have four kids. I had twins. I overshot. It's a long story.

(Laughter.)

MEMBER EDIGER: But so my pediatrician -- and we have gone to this same pediatrician. We love her more than oxygen. We've gone to her for over 16 years. And one of her standard questions is, do you have your own room? And every time, my kids are like, "no." Because I have four kids. And I always felt it always sat really uneasy with me. I always felt like it was a very judgmental question. Like, no, we don't have a house with five bedrooms but thanks for asking every single time I come in.

So, I finally asked her. I said can I just ask why you ask this question. And she said, "oh, well, I want to know if they feel like they have a place to go that is private. Like do they have space in the house to go to feel like they have their own space?" And I was like, well, then you should ask that. And I told her. I said I feel very judged when you ask my kids four times a year in front of me.

Anyway, so I think that is a really

good example of how good intentions -- she's going through her checklist, so I'm sure that question lines up in her mind with one of the questions on there, but with me it was salt every single time.

And I have other examples like that but just everything about reading about this survey, it didn't sit well with me because the parent gets nothing out of it except a report that says when you go back in 12 months for your next checkup, here's a list of things you should ask about: your car seat, if they have their own room -- yeah, 12 months later.

And I'll say, I'm specifically in the psychosocial screening so I'll save some of it specifically for that when we get there, but I have a lot of concerns about the intent of what this is supposed to do, and the people that do all the work actually filling out the 20-minute survey don't get anything out of it, in my opinion.

CO-CHAIR BROOKEY: You have to

disclose every time that you have guns in the house and things like that. So I appreciate your comments from a parent, because we do, we do survey -- we survey our patients all the time, whether they want to be surveyed or not. And there is very good reasons for doing that but I appreciate your perception of it.

And Tara?

MEMBER BRISTOL-ROUSE: So I feel a little bit differently, actually. So I work in quality improvement so there is an appeal to me that you can get the data back after ten responses so that you can make some semi-real-time changes in practice.

But I wanted to address the perception issue because I could see where that could be a sticky wicket for some providers, is that I think for parents or caregivers perception is reality. And so even if a parent is saying you never talked to me about breastfeeding, and this kind of gets to Maureen's point, like, you know, you have to look back and you go, as a provider you

might say, "I know I have talked about breastfeeding with everybody." So then it's like, "well, am I not talking about it in a way that is meaningful to the parent?"

And so I think regardless of how you look at it, the parent's perception is the weighty thing that should help you to improve your practice. So I just wanted to bring up that point about perception and what's meaningful.

CO-CHAIR BROOKEY: Good point.

Jeff.

CO-CHAIR SUSMAN: So, briefly, I am thinking about this in an accountability framework and trying to make a judgment about a provider's performance based on providing the whole haywagon-full of anticipatory guidance, 15 questions, whatever it is, rather than potentially meting this out and focusing on maybe half of the questions or two-thirds of the questions and then following up later.

So I am not convinced that there is an evidentiary link between doing all this stuff and

an improved outcome. I mean, I believe that each of these individual things are probably good things to do and there is probably some evidence we could find for each of them, but the all or none aspect, I am not clear about.

And in fact I know there is good evidence around changes in behavior, preventive services, that suggests that you focus on a meaningful few, as opposed to everything, or you do some negotiation process. So that's where my dis-ease and looking at it from an evidence perspective.

CO-CHAIR BROOKEY: So, of course I agree with that, but I also just want to go back and say we're looking to see whether or not the measure, as written, could potentially lead to, in this case, a change in behavior in the provider. Does it have to be broken down into every element or would -- sort of three different answers, you basically have three different pillars of this.

So, is it sufficient for evidence to

say that, even if you don't get the detailed report, that a provider could look at that report, compare themselves to others, potentially change their behavior? That's the question about evidence.

So, we have a couple cards up. I will start with you, Rajiv.

MEMBER MODAK: So I think that what is the challenge is, I think there's a potential to change behaviors when performance is very low and you can improve those behaviors. But when you get to a certain point, I don't know how much you really impact meaningful behaviors that will improve outcomes for patients and will even improve their satisfaction.

So, speaking to going over the top with the questions that you are asking where I think there is some support for the idea that you can move people who are low performers to the middle.

MEMBER FINKELSTEIN: So, I was listening to what you were saying, John, and I

would have said it very differently. If this is -- if we are considering this an outcome measure -- are we considering this an outcome measure? I think the question is not whether the measure would result -- having the measure would result in provider behavior change. That's not the question. The question is, if this is an outcome, does provider behavior -- is there something the provider can do or not do -- affect this as an outcome?

So, from my perspective, I'm going to have other issues as we get into validity, but I think, if this is an outcome, I could do something to affect this outcome.

CO-CHAIR BROOKEY: Yes, just to read, there needs to be a relationship between the measured health outcome and at least one health care action, which is structure, process, intervention.

MEMBER FINKELSTEIN: Right.

CO-CHAIR BROOKEY: So it could be -- so again, I mean, I think ultimately we want

better outcomes for the actual patient. But in this case, we are just looking for some change that could be -- some action that could result from this measure, the result of this measure.

So it's just the way that the algorithm is written. I would like there to be more evidence for actually improved patient outcome but that's not what we are looking at here.

Let's see. Carol, did you have your hand up?

Jim.

MEMBER BOST: I just want some clarification from staff. So it's my understanding that this survey was already approved. Is that true?

DR. NISHIMI: Well, not as a quality measure.

MEMBER BOST: Actually, what I'm trying to get at is, you have already signed off on the items in this instrument and the domains. So we're not really here to look at whether an

item is good or valid or reliable, just the roll-up measure itself. Or do I not get that?

DR. NISHIMI: There was an original survey instrument 0011. So the fact that it has number 11 means it was very old.

(Laughter.)

DR. NISHIMI: And it was done at a time when NQF considered instruments in their entirety. So the kind of review and the criteria have evolved and the types of testing required have evolved.

So in fact that's why these have new numbers because they have been split into specific performance measures based on those items. So previous endorsement doesn't really apply here. You're looking at this performance measure, which is the percent of, you know, blah, blah, blah.

MEMBER BOST: It sounded like the developers weren't ready to answer those kind of item analysis stuff that are previously published psychometric papers but were more here to focus

on the measure.

DR. NISHIMI: And you will see we require item-level testing and score-level testing. And so they have looked at and reported certain focus group-related stuff, but they've also looked at item-level testing. And so that is why that is in there, because our requirements now are different. We're actually looking at the performance measure, not the instrument.

CO-CHAIR SUSMAN: Just briefly, one of the other factors in this is benefits of -- benefits and harms. And I see what the sort of natural outcome of this is that the Epics of the world will have checklists, that people will dry-lab every one of those checklists because you want to get credit. Or if they are really honest, they will subject you to all 15 items every time you come in, or whatever the appropriate thing is.

So I think there are some potential harms here that we just need to be mindful of. I don't know whether that really shifts the

conversation fully one way or the other.

CO-CHAIR BROOKEY: David.

MEMBER EINZIG: So, nuances. So when we vote, are we voting -- I mean, I think it is easy to say that will change behavior, but if we don't think it will change behavior in a meaningful way, does that affect how we should vote?

CO-CHAIR BROOKEY: I think we're sort of blending our own personal experience with validity questions. We're questioning threats to validity. We're questioning unanticipated harm, unanticipated reactions from parents even perhaps. So there's a lot -- we're blending in a lot of this together. So I think we need to focus on the first question of evidence, which is, everybody, I refer you back to your algorithm, and just to say, whether you like it or not, the way the measure is created, whether you like the way it is reported back, it is not an item level unless you hit a certain volume threshold.

As it is written, do you believe that there is a relationship between the measured health outcome and at least one healthcare action, which could be a structure, process, intervention, or service, which means, in my mind, that the me getting this report will look at this and say, wow, I need to spend more time on this.

And so that's the question. That's what you have to wrap your head around to say whether or not you believe that is true or not. And that is what we are voting on. And kind of try to put aside all that other stuff for the moment because we are going to get to that.

So is everyone clear on what we are building on, I guess, is my question. So, it's not real satisfactory, is it, but it is what it is. So we're just going to have to go one step at a time.

So, let's go ahead and vote on evidence, pass or not pass.

MS. JUNG: The voting for Measure 3219

for evidence is now open. Option 1, pass; option 2, not pass.

Okay, voting is now closed. We have 68 percent for pass with 15 votes; 32 percent for not pass with 7 votes. And with that 68 percent, the measure passes for evidence.

CO-CHAIR BROOKEY: Okay, so let's talk about measurement or performance gap. Who's going to start this one?

MEMBER KONEK: Alright. So there appeared to be opportunity for improvement. The developers have provided three different well-designed research studies that show substantial variation in the measure across sociodemographic groups. And it showed there is room for improvement.

And we're going to hear more about those tests, those studies shortly. But there was clearly a demonstrated performance gap in providing AGPE, with rates varying from only 40 to 60 percent. The review of the Promoting Healthy Development Survey results of the top

five providers -- and you are going to hear about them through the whole thing; they looked at the people that had the most surveys that came back.

And so they looked at them and the range was 46.8 to 84.8 percent was seen by the proportion of parents who reported discussion of all topics or reported no need to discuss them, such as people with a couple of kids. And the cited KPNW study showed a range of 22.2 to 67.7 percent.

So, it looked like there was a big range in how well people -- the providers were providing this so that the family felt that they had their needs for the information met.

CO-CHAIR BROOKEY: So do you interpret that as a gap?

MEMBER KONEK: Well, that's how it was laid out. It seems they felt that there was a gap as far as performance for this. There was disparity shown with some groups, also, which were also discussed at length.

MEMBER BOST: Yeah, I mean, basically

the rates presented showed a lot of room for improvement. There were differences across socioeconomic groups, and, like you said, across physicians.

MEMBER KONEK: Right.

MEMBER MODAK: And I also just want to echo that that's the most important point, to me, was the disparity among ethnic backgrounds, socioeconomic backgrounds. And it's been reported that especially when there's language barriers there's a huge discrepancy. So I think there is a significant gap.

CO-CHAIR BROOKEY: Yeah, the use of interpreters and health literacy issues, all kinds of things can contribute to that, parents feeling like they weren't communicated to during the visit. Those are very important points.

Ricardo. Ricardo.

(Laughter.)

CO-CHAIR BROOKEY: Your flag is up.

MEMBER QUINONEZ: I have a virus. Sorry.

So, in the gap, in the evidence for gap, just as a question, does the developer have to show just that there's a gap or that gap means there is a difference in quality? I mean, that there's a gap in quality. So, meaning, yes, we know that there's decreased performance if you look across different providers, but I'm not sure that necessarily correlates with worse quality.

CO-CHAIR BROOKEY: Yeah, I think it's not about that. I think it's about whether or not this measure is going to -- if you measure it, you are going to see a gap.

MEMBER QUINONEZ: Big difference.

CO-CHAIR BROOKEY: Right.

MEMBER KELLER: I think that it's that the evidence supposedly establishes that there's something there and then the gap shows that there's room for improvement.

MEMBER QUINONEZ: No, but because we only voted on whether the evidence can be modified. We didn't actually vote on the merits of the evidence. So, that's why I am asking.

CO-CHAIR BROOKEY: We voted to say -- well, we did vote on evidence, which we defined in this situation as being something that could be acted on. And then the gap is that there is variation between providers and ethnic -- or there's socioeconomic or ethnic differences between the results.

So I think you have to kind of stretch it further to believe there is a quality gap but I don't think that is what we are voting on here.

Yes?

MEMBER DORSEY: So since this is a provider-level measure, the question that I want to see the answer to is either within do providers have differential quality for their higher-income/lower-income patients, or among providers who have a large proportion of certain vulnerable patients, do they tend to have lower scores? Rather than just the raw rate in the subgroups. That doesn't tell me much about the performance gap at the provider level.

CO-CHAIR BROOKEY: Right. So, John?

MEMBER FINKELSTEIN: So I have a question really for the NQF staff. And I'm okay continuing down this road but this was a problem that came up at our last meeting. I will just go on record. This is not an outcome measure. It is a patient report of a process. And when you treat a patient report of a process -- this isn't "can I get up in the morning and go upstairs." That is a patient-reported outcome. When you take a patient report of a process and start treating it like an outcome, we get into funny problems.

So, if that's the road you want us to go down because that's -- that's okay but I just want to call that out.

CO-CHAIR BROOKEY: Okay, duly noted.

Carol?

MEMBER STANLEY: Yeah, a couple of things about gaps. One was this instrument -- is this sent out in Spanish to patients whose primary language is Spanish? Because if that's not the case, then I'm not so sure you can say

there's a disparity or what kind there is with ethnicity.

The other thing along those lines is I think I read where the instrument is written at the eighth or ninth reading level. So, that's pretty high, considering Medicaid population. So I'm not really sure how accurate this disparity presentation is.

And I also agree with the comment that if this is a provider-level measure, then I think we would need to see numbers that showed disparities between the providers and not looking at these aggregate numbers.

CO-CHAIR BROOKEY: So for the developer, is it sent out in either Spanish or any other threshold languages?

DR. SOLLOWAY: I don't believe it is because I don't think we had the money to develop the Spanish version, but I could be wrong about that. Naraa, do you know?

DR. GOMBOJAV: Yes, no Spanish version for now.

DR. SOLLOWAY: I don't think so. I don't think we had the money to develop it.

CO-CHAIR BROOKEY: Thanks for bringing that up, Carol.

DR. SOLLOWAY: Spanish, yeah, it's a very good point.

CO-CHAIR BROOKEY: Jeff?

CO-CHAIR SUSMAN: It just seems like a lot of issues that we are talking about are really tending toward more validity issues and we should make sure we come back to them.

CO-CHAIR BROOKEY: Yeah, don't forget that one. That's a major concern.

So let's see if we can talk about, or at least vote on, gap, now that we -- do we all know what gap is now for this purpose?

So, let's go ahead and vote on it. Let's close the gap.

MS. JUNG: Okay, the voting for Measure 3219 for performance gap is now open. Option 1 is high; option 2, moderate; option 3, low; and option 4, insufficient.

Okay, voting is now closed. The results are 5 percent for high with one vote; 77 percent for moderate with 17 votes; 5 percent for low with one vote; and 14 percent for insufficient with 3 votes.

And with this, the measure does pass for performance gap.

CO-CHAIR BROOKEY: Okay, so reliability. Who's going to kick that one off?

MEMBER BOST: So reliability testing was done at three different studies, each that seemed to have adequate sample size and appropriate variability across socioeconomic status and age. The Cronbach's alphas were consistently in the 80 to 90 percent range for all items together and in the 70 to 80 percent range for the items within each of the three questions.

Though it's really mentioned in validity, the items also had high factor loadings on each and acceptable interclass correlation coefficients.

I think, if anything, I might have liked to see some test/retest to see if parents would respond similarly a week apart or something like that. But other than that, I thought there was good reliability assessment.

CO-CHAIR BROOKEY: So, Carol, go ahead. I mean Sue. I'm sorry.

(Off-microphone comment.)

CO-CHAIR BROOKEY: So getting back to the question about language and getting to the question about the provider sending it out themselves and they may send it out the same day or a week later or further down the road, does that affect potentially both reliability and validity or not?

MEMBER BOST: Yeah, that's why I would have liked to see some test/retest to see if it is different. Because then if you're doing something like comparing across clinicians, should you be doing that consistently or not? And that kind of testing might help us understand that.

MEMBER MODAK: So I think that that's a key point, especially with reliability, is when those surveys are getting sent out as well, because there is a huge difference in patient memory or parent memory from the day of the visit to three days later than after the visit. So, I don't think that has been vetted in this.

CO-CHAIR BROOKEY: So are you suggesting that is a threat to reliability as well as validity?

MEMBER MODAK: Yes.

CO-CHAIR BROOKEY: Okay. Now Carol.

MEMBER STANLEY: Yeah, and along those lines I think with needing to have six well-child visits by 15 months, if you are going by the HEDIS measure, and a parent gets the survey at month 16 and their child has had six well-child visits over the past 12 months with different providers, how do they know which provider they're evaluating if this is going to be a provider-specific report?

CO-CHAIR BROOKEY: I will ask the

developer to answer, that but I'm guessing that the questionnaire will say during your last visit. Is that correct?

DR. SOLLOWAY: Yes, actually -- well, some of the questions ask in the last 12 months. Actually, these questions ask, "since your child was born did your child's doctors or other healthcare providers ask you about the following." So it really includes any of the providers you had. So it's really kind of a comprehensive measure of taken as a whole.

MEMBER STANLEY: How can that be provider-specific, then? If the report is going to a provider, how do you discern those -- that it's encompassing the providers from the last 12 months?

DR. SOLLOWAY: Well, actually, I misspoke. These are not for the last 12 months. These are since your child was born did you get these questions.

And, you know, you're asking a very good question and I think there is probably some

assumption that they're staying with the same provider. But as I said -- oh, and actually in the injury prevention one it is in the last 12 months.

And I know that these questions were really designed to really coincide with the way the questions were asked because they were tested for their validity and reliability in the National Survey of Children's Health. So I know that these questions were meant to align with those questions, which is why the wording is the same. But I understand your point.

CO-CHAIR BROOKEY: So -- and I want David to go. I just want to back up a notch, having this new information. Going back to the question of evidence and saying that I can look at my report and do something about it, now I'm hearing that this report may not reflect the care that I gave. Is that an accurate statement?

PARTICIPANT: You could come to me and tell me what to do better.

CO-CHAIR BROOKEY: I'm asking the

developer if that's accurate, that the report is not strictly speaking just based on what I did. It's the perception of the parent over all the care they've received, and it could be from numerous providers.

DR. SOLLOWAY: I think that's true, yes.

MEMBER HOUTROW: To speak to that in particular, we did some research several years ago which shows that children with more encounters, including ill visits, which tend not to be with your regular primary care doctor, in fact increase the likelihood that you will get anticipatory guidance, which is a good thing, but it means that attribution is related to all the care not just -- I mean the care to one provider.

CO-CHAIR BROOKEY: Okay, David, I think, is next.

MEMBER BOST: I have the actual question in front of us. And the first one is "since your child was born, did your child's doctor or other health providers talk with you

about." The second question is also since your child was born. The third one is in the last 12 months.

CO-CHAIR BROOKEY: So this is not attributable to any individual unless the child only saw one individual provider.

MEMBER MORROW-GORTON: Or even a practice. Other healthcare provider could be the WIC clinic.

CO-CHAIR BROOKEY: I think we may need to go back and re-vote on evidence. But let's go ahead and go around the room here.

David.

MEMBER EINZIG: Yeah, so, I mean, I am having a lot of trouble with this, along with a lot of people, I think, but I just can't remember what happened an hour ago.

(Laughter.)

CO-CHAIR BROOKEY: That is a different survey.

MEMBER EINZIG: And so when you get these responses back, so that there is that

piece, and then there is the piece of this takes 20 minutes for a family and this time consuming. How do you feel when you get a survey in the mail like that?

And then the issue of, I mean, just taking into consideration there are so many variable, if it's a family with a bunch of kids and they are stressed and they are struggling to pay rent and then there's this. And you're comparing those surveys with the other families who have one kid and have all the time in the world. So there's a lot of variables that don't feel right to me here.

CO-CHAIR BROOKEY: Could I ask that -- could I entertain more comments about evidence? Would that be okay if we went back and talked a little bit about that, Helen?

DR. BURSTIN: Everybody, I'm Helen Burstin, Chief Scientific Officer. So apologies for not being here this morning.

I'm stuck on John's question because I think he's absolutely right. This is a process

measure. Just because it's a patient report doesn't mean it's an outcome. And in fact I was just pulling up, Lisa and I were just -- and some of you were actually here just for a behavioral health meeting just a couple of days ago -- and we in fact looked at measures from NCQA that were patient self-report of guidance about smoking cessation. You were here as well. That was a process measure. I just pulled it up.

So, to be consistent, unless it's actually an outcome or a patient experience, it would be a process measure.

And then just one other question. These measures, when they had come in before with the state level of analysis, and I wonder if when they've gone to the lower level of analysis, I'm not sure -- it would be helpful to hear from the developer of what additional work was done to see their applicability at this much lower level of aggregation.

CO-CHAIR BROOKEY: So, for the developer, did you hear the question?

DR. SOLLOWAY: Yes, Naraa, would you like to answer that?

DR. GOMBOJAV: Okay. So, based on the previous studies we have done, especially for the Kaiser study, we have tested the plan -- I'm sorry -- the provider, they will --

CO-CHAIR BROOKEY: Excuse me. I'm wondering if you could speak up a little bit. We are having a hard time hearing you.

DR. GOMBOJAV: I'm sorry. So we have done some low level analysis in the Kaiser study and also the peer review paper about PHDS testing being included. For plan levels, they were compared in different plans, so those --

MEMBER KONEK: Naraa, your voice is really muffled. We're having trouble hearing you in the room.

DR. GOMBOJAV: I'm sorry. Yes, I'm out of work today and I'm starting to cough. My voice is kind of hard. That's why and maybe you are hearing some difficulties. I'm trying to make it better.

So I am just saying we have done some analysis at the plan level, which was partly involved in the Kaiser study, the initial study we were testing the measures of survey measures. So those are in a peer reviewed paper already published.

CO-CHAIR BROOKEY: Okay.

DR. SOLLOWAY: And the other thing I would just like to say is, so, initially, this was designed to actually compare providers within a practice and then the practices within the plan. So initially, it was done, and it was also done at the state level. So we have state, plan, and practice and provider level. And then we kind of changed it to be just provider-specific and not to compare providers against each other.

CO-CHAIR BROOKEY: Okay.

DR. SOLLOWAY: So we did do individual provider levels as well.

CO-CHAIR BROOKEY: But these measures are specifically meant to be at provider level for individual provider feedback and for

individual provider action related to the feedback. Correct?

DR. SOLLOWAY: Yes, for individual providers. We have the capability on the back end to actually aggregate the providers within a practice because we know what practice they have.

CO-CHAIR BROOKEY: Well if you are in Kaiser Southern California, you can go all over Southern California.

DR. SOLLOWAY: That's exactly true.

CO-CHAIR BROOKEY: So I'm just clarifying the facts. It sounds like we started out by saying this was at an individual provider level, that we had concerns about it being in the sort of domains and not even individual items, but that this would be used for feedback, and, even further, it could even be used for financial incentives or all kinds of things that groups and health plans do with individual provider measures. But then we have discovered that in fact the questions do not necessarily relate to that individual provider.

So I just want to clarify that those are the facts.

DR. SOLLOWAY: I would say yes, that's true.

CO-CHAIR BROOKEY: So I am going to propose to the group that we revisit the evidence question and we'll vote on a revote. And if we vote to revote, we'll revote.

So can I open it up for questions or comments about the evidence? Okay, go ahead.

MEMBER FATTORI: I don't know if this is an evidence question but are we voting on this as a process measure or an outcome measure? Process?

CO-CHAIR SUSMAN: We're going to change it to process.

MEMBER FATTORI: Okay.

(Simultaneous speaking.)

CO-CHAIR BROOKEY: So, let's just sort of talk informally for just a little bit because if they completely rewrote this to make it a process measure, is it -- they are not going to

change the methodology of the measurement. So, is this going to be a measure that is going to do what it purports to do, which is to be an individual report card for an individual provider? Is it ever going to hit that mark, is my question?

MEMBER BOST: Well, it depends, also, on the instructions that you're given. Because if it comes up on the website and it says, please comment based on your visit on such and such a date, would you not think that they are talking about that provider?

CO-CHAIR BROOKEY: But they'd have to change the tool.

MEMBER BOST: I'm just asking.

(Simultaneous speaking.)

CO-CHAIR BROOKEY: I'm sorry. Jill, would you repeat what you said?

MEMBER MORROW-GORTON: So I think you could give those instructions, except the way the question is asked, at least the one that you gave us was, in the last 12 months, so that blows your

last visit, and your practitioner and all of your other healthcare practitioners. You know, people could have a PT that wasn't even related to the practice.

CO-CHAIR BROOKEY: So what is your conclusion?

MEMBER MORROW-GORTON: So my conclusion is that we could not use this to give feedback either to an individual clinician or to a physician practice because it's too open to other healthcare practitioner's input.

CO-CHAIR BROOKEY: Okay. John.

MEMBER FINKELSTEIN: So I just want to defend the measure for a minute and then raise a real concern. So I think when we flip to process it's a different pathway for evidence and a different level for evidence, so we are going to have to deal with that.

CO-CHAIR BROOKEY: Right.

MEMBER FINKELSTEIN: I think the most important question to me isn't the one we're talking about now. I get it, the attribution

isn't perfect. The evidence that this is based on comes from the evidence behind Bright Futures. And you can quibble with that evidence because the evidence for anticipatory guidance and well-child care in total is thin in some ways, but there was as much evidence to develop Bright Futures as those people could generate and we really have to look to that.

So here is the question, for me, for the developers, is I don't hear them talking about this as an accountability measure. I hear them talking about this as feedback to practitioners so that they can individually improve their care based on the last ten patients they saw and get MOC credit because that's a PDSA cycle. But do they ever think that this, at a practice level, that either individual practitioners or one practice to another to another, can be judged by these measures? I realize we are getting into validity, but do they even think about it as an accountability measure?

CO-CHAIR BROOKEY: I'm just going to

take all of your comments first. So go ahead, Carol.

MEMBER STANLEY: Along the lines of evidence, I'm not hearing anything that tells us that the scores wouldn't have improved anyway just at the mere fact, someone said it earlier, I think, that if you see a physician once and then twice and three, chances are you are going to see an increase and improvement in the scores because you are developing better rapport. So how has the developer sort of addressed that issue that is there evidence that using this tool is improving scores over time versus just the fact that the patient is seeing the same physician six times in 15 months?

CO-CHAIR BROOKEY: So I will let the developer answer that question. Basically, do you have data that shows actual improvement in individual provider scores?

DR. SOLLOWAY: Well, I would say yes to your question but no to the woman's question, because we haven't -- we didn't have research to

actually -- we didn't have the opportunity -- well, no, that's not right.

What I'm saying is asking the question, was it because they used the tool or because they developed a better relationship, would be in and of itself a separate research question and a separate research project.

So what we did know from our focus groups with parents is that they really -- they liked being able to report back to the providers. They liked being able to give this feedback to the providers. And they did see improvement, but whether it was because there was a better relationship developed or did the tool help them develop a better relationship, I don't know that we can tease that out from the data that we have.

CO-CHAIR BROOKEY: Okay, thank you.

So, Rajiv, then Jeff, and then Jill.

MEMBER MODAK: Just, again, on the evidence, we voted about the evidence based on the idea that it was a question asked in the last visit, were all of these questions asked in the

last visit. Now we are changing it to were all these questions asked since you were born. Even with or without a relationship, obviously, if you have three visits, you're going to ask more questions, the potential to ask more question than at one visit. So this is completely different than we initially had voted for.

CO-CHAIR BROOKEY: So what are you concluding?

MEMBER MODAK: Well, we need to take a revote. But the question is, I mean, if we are changing it to a process outcome, then -- the question would be -- my question is, if we have them completely rejigger this to be a process, is it going to change that point about it not being an individual visit questionnaire; it's based on the past 12 months, which could be multiple providers.

So I'm just trying to just ask whether this can ever be an accountability measure at the individual provider level, which is a premise behind all of this.

Jeff.

CO-CHAIR SUSMAN: I mean, to answer your question, I think it's invalid. I don't think there is any validity. I think we will vote it down on validity if we were ever to vote on it. I think, at a minimum, what we need to do now is ask for the developer to reprocess the application as a process measure, consider it -- pardon me?

DR. BURSTIN: Do a 511 analysis.

CO-CHAIR SUSMAN: Find the level of analysis, send it back to our comment meeting.

I still doubt very much whether it will pass because I think of the validity issue, but we'll see. We will have an opportunity.

CO-CHAIR BROOKEY: They have the other -- we can provide feedback to the developer, not only to redo the evidence review but to give them feedback that we don't think it is going to pass validity so you may want to think differently about how you are going to propose this measure.

CO-CHAIR SUSMAN: And moreover, it

applies to, I think, probably all the measures. If the unit of analysis is under question and we can't really determine that or need clarification, and if the process versus outcome measure issue, we're just going to be churning and not really accomplishing --

CO-CHAIR BROOKEY: It's not a good use of time.

Okay, Jill.

MEMBER MORROW-GORTON: I think the other thing we have to think about is, if this is in fact paralleled Bright Futures, Bright Futures does not say you should do 16 things at a visit. But the assumption in terms of the how do you improve your -- you know, it's an all or nothing; you had to have met the 16 and are we giving the wrong message to physicians -- if in fact we would say that you could measure an individual physician or a practice, are we giving them the wrong message that based on a single visit but using all the time that happened before, we should expect that all of those boxes would be

checked.

CO-CHAIR BROOKEY: So, we've been given the signal that the court will now recess for five minutes. So you can all stretch while we huddle in the corner. Okay?

(Whereupon, the above-entitled matter went off the record at 3:52 p.m. and resumed at 4:01 p.m.)

CO-CHAIR BROOKEY: Okay, the court is back in session.

CO-CHAIR SUSMAN: Hear ye, hear ye, the Honorable John Brookey.

CO-CHAIR BROOKEY: We have a plan.

CO-CHAIR SUSMAN: It may not be a good plan but there is a plan.

CO-CHAIR BROOKEY: Is everyone pretty clear on what the issues are? Okay.

So we need to go back and revote -- well, let me just ask for a show of hands. How many feel like it is appropriate to revote on evidence? Okay. Does anyone disagree with that? Okay.

We think, at this point, the way that it has been presented, it's not different than other measures that we've had that moving forward probably should turn into process measures. But in all fairness, they were asked to present it this way. And so we are going to vote on this measure as a self-reported outcome measure, which means that all the conversation we had earlier is still in play but we have new knowledge that this is not actionable at the individual practitioner level if the patient saw multiple providers.

So when we said this had an action as a result of the measure, we were all thinking that it was based on one visit, one provider and now we know that that is not true. So that may, in your mind, change the strength of the evidence for this measure.

Amy.

MEMBER HOUTROW: We are still in the green bar, which is that this measure -- there's a relationship between this measure and one actionable healthcare action: structure, process

intervention, or service, right?

CO-CHAIR BROOKEY: That's correct, but just to restate it, we were initially thinking that this would be -- I'm just going to paraphrase it -- this would be -- and I get these. If you're a practitioner, you get report cards. You get satisfaction report cards. We are used to this. We get feedback and we assume that feedback is attributable to me. And then we found out that it's not attributable to me, so I may not be able to use this data because I don't know if I'm the one who didn't give them the advice or not. So I think that's why we are revisiting this as an evidence measure.

MEMBER KELLER: I hear that, and I just wanted to make sure that we all understood that there's actually a number of outcome measures that are not rigged to look at the individual visit and look at the care provided by the system, even though they attribute that questionnaire to individual providers.

CO-CHAIR BROOKEY: That's correct and

that was mentioned earlier. It could just be at a state level. Could it be at a plan level? There's all different kinds of questions.

MEMBER KELLER: No, no, no. These are used at the -- so, within NCQA, one of the criteria for being a patient-centered medical home is that you regularly survey your patients and they give you extra points if you use the CAHPS survey.

The CAHPS PCMH survey, the questions are all about the care you have received in the last year, both the child and the adult, and they say by your regular doctor or another health provider. So the assumption there in that construct is that, as primary care provider who has a patient attributed to them, you are responsible for the totality of the care even if you don't directly deliver that care, and you are supposed to identify gaps that other people may have left out.

And we can think about whether that's right. I mean, that's actually, in that

construct of a patient-centered medical home, I don't think that that's wrong. I actually would argue that that's the right way to ask those questions.

CO-CHAIR BROOKEY: So I thought CAHPS had a plan level.

MEMBER KELLER: There is a CAHPS PCMH survey.

CO-CHAIR BROOKEY: Okay, I wasn't aware of that.

MEMBER KELLER: There's an individual provider level as well.

DR. BURSTIN: But PCMH is not individual clinician. It's to the medical home, which I think would be different.

MEMBER KELLER: It's to the medical home but in many constructs that is -- in Colorado, sometimes that's an individual physician because, you know, we still have solo practitioners.

CO-CHAIR BROOKEY: So that's a good distinction.

MEMBER STANLEY: But the CAHPS survey also requires 12 months of continuous enrollment.

MEMBER KELLER: Correct, yeah, they're looking at your care over the last 12 months.

CO-CHAIR BROOKEY: So this particular measure was presented to us as a measure at the individual provider level and it was presented to us as an outcome measure that would lead to some action, presumably by that individual provider believing this report is a reflection of the way they communicated with the parent.

So I propose that we revote, based on the new information, that we revote on the evidence as written.

Are there any other questions or concerns before we revote? On the phone?

Okay, we all know what we are voting for, right? Okay, so let's open it up.

MS. JUNG: Okay, voting for Measure 3219 for evidence is now open. Option 1, pass; option 2, not pass.

CO-CHAIR BROOKEY: I can see where

your thumbs are.

MS. JUNG: Okay, voting is closed. So we have the results are 36 percent for pass with 8 votes; 64 percent for not pass with 14 votes.

CO-CHAIR BROOKEY: Okay.

MS. JUNG: And with that, we do not pass.

CO-CHAIR BROOKEY: So the question is, I believe that all five of these measures have the same situation, is that not true?

Go ahead.

MEMBER FINKELSTEIN: So just a clarifying question, because I was assigned one of the other measures.

The next measure is, "were you asked if you had any concerns." And that is appropriately for the last visit. It's only about what happened in the last visit. It's a very narrow measure. It's at the last visit were you asked if you had concerns.

CO-CHAIR BROOKEY: So, let's just go to the next measure, then, and if there are

differences in that, then we will call them out and we can move forward.

Okay, so this measure did not pass so we are going now -- to is it 3220? Is that right?

DR. SOLLOWAY: You know, I could just say that I'm looking at this survey and asking about parental concerns, family assessment for behavioral health and safety, and family assessment for psychosocial screening, they all ask in the last 12 months did your child's doctors or other healthcare providers ask you. So it's the same issue for all of them, except, I think, family-centered care.

CO-CHAIR BROOKEY: Is that Lauren speaking? Who is that speaking?

DR. SOLLOWAY: Oh, I'm sorry. This is Michelle Solloway speaking, the developer.

CO-CHAIR BROOKEY: Oh, okay.

MEMBER FINKELSTEIN: So I just want to follow the logic here. Ask about parental concerns is something that should happen, I would

argue, at every well-child visit. This is going out after a well-child visit.

So, if I'm sending it out because I did a well-child visit and then it goes to a family, they should be -- if asked the question "has anyone asked if you had concerns in the last 12 months," if I don't get a yes on that question, that's a ding on me because I should be asking it -- because it should be asked at every well visit, it's different than "over the lifetime of your child has anyone asked you about blah, blah, blah."

CO-CHAIR BROOKEY: But if the answer is yes, it may not have been about you.

MEMBER FINKELSTEIN: Yes, I guess that's true.

CO-CHAIR BROOKEY: So is it actionable? So I don't see this as any different, personally.

So I'll just go around the room. Does anyone see this as different?

DR. SOLLOWAY: Can I just ask a

question about this? Because I think what the person who was just speaking is really asking that it sort of doesn't matter if it was, you know, someone else or not you; if you weren't asked about it, it is a problem.

CO-CHAIR BROOKEY: Correct.

DR. SOLLOWAY: And you should have been asking about it as a provider. So it still gives the information that you should have asked.

CO-CHAIR BROOKEY: Well, no, because if they answered yes and it wasn't about you, then it has nothing to do with your practice.

MEMBER FINKELSTEIN: I have to say, that's going to happen very rarely.

CO-CHAIR BROOKEY: Well, I know but --

MEMBER FINKELSTEIN: So there's some noise in the measure.

CO-CHAIR BROOKEY: You understand the logic.

MEMBER FINKELSTEIN: I understand the logic, but that's a rare event. So we're focusing on a problem now that that's only a

problem if the child was seeing someone else and recently came to my practice. And I agree, that's a problem, a serious problem. But it's a minority. If I did this on 30 kids, that would be a small minority of the kids.

CO-CHAIR BROOKEY: Okay, Ricardo.

MEMBER QUINONEZ: Yeah, I agree with you that the attribution problem is going to be the same for all. But isn't whether this is an outcome or a process measure going to be a problem for all, as well?

CO-CHAIR BROOKEY: Well, I think the question is -- so if we have them go back and rewrite this whole thing as a process measure, if the questions are the same, are we going to have questions with validity?

So I mean how far are we going to go down the road? We need to follow the process. So the process is we have to vote on evidence first. But as presented, and not just this measure, but other measures --

MEMBER QUINONEZ: But the evidence

level that we vote on is very different if it is an outcome or a process measure.

CO-CHAIR BROOKEY: That's correct.

MEMBER QUINONEZ: And we don't have the information. We are not going to --

CO-CHAIR BROOKEY: We don't have the information. So we had to make a decision as to whether -- because we are singling out this one but there have been others that we have gone ahead and voted on as an outcome measure, right?

DR. NISHIMI: Correct.

CO-CHAIR BROOKEY: So kind of changing the rules on the fly is probably not the best process for NQF. I think that the staff has gotten feedback and likely, there will be some changes. Probably not doing it in the middle of a meeting -- that is probably not appropriate.

So the proposal was to go ahead and vote for it as an outcome measure. And I am asking the question now is this problem any different with this particular measure or is it the same?

DR. NISHIMI: Can I have a show of hands of anyone who does not think it is the same?

MEMBER HOUTROW: Thinks it is different.

DR. NISHIMI: Thinks it is different. Thank you.

MEMBER HOUTROW: In the last 12 months.

CO-CHAIR BROOKEY: David.

MEMBER KELLER: So again, the issue, if we are going to move this to a process measure, then you are right and anything I am going to say makes no difference.

If the issue, though, is whether there is evidence that this measure, as currently constituted, is changeable by an individual provider, I would argue that there is and it was presented to us. It makes -- even if somebody answered this question no: In the last year have any of your providers not attended to your needs? And I take special efforts in my practice to make

sure that I ask about the needs at every visit. Over time, that answer is going to change. We will have changed that outcome, even though I can't change what happened with another provider.

If it is an issue that they are going to see multiple providers within my practice, I can undertake a standardization effort to standardize that question so that it is always asked every time they walk into our office. And that will move that metric. It won't move it to 100 percent but it will move the metric.

So I mean, again, I am sort of with John on this. I think this is different than the questions that are specific to the 15-month-old visit that I am not going to be able change those because by the time they come in for the next one, they are going to be asked a different set of questions. And whether I can change the 15-month one or not is a different.

CO-CHAIR BROOKEY: Rajiv.

MEMBER MODAK: So I just was looking at the questions also. And I want to just point

out that it also says in the last 12 months did your child's doctor, or other health provider, including a nurse, ask if you have any concerns about your child's learning or development. So that is different also. That is no longer provider. You could -- okay.

MEMBER DORSEY: I think they phrase it that way because we fracture the way that we address these issues in clinical practice. Some of us have child life specialists who help the parents fill out development surveys, right? I think that is meant to bow to the structure of the way primary care is delivered differently in different settings. I don't think it is about -- they are not really asking that an emergency doctor to do that.

I think we might feel better about this if we saw some data element validity. That came up in the earlier discussion of the previous measure that if we could see this checked against what actually happened in the practice so that we were confident parents weren't going back to a

visit they had, a WIC visit four months ago and that was what was populating the answer. I think we would feel better about the way it was phrased but I don't think we have all that information from the developer about all the testing that was done during the development of this survey.

So I think some of that is you know just a lack -- we just don't have that in front of us.

CO-CHAIR BROOKEY: Jill.

MEMBER MORROW-GORTON: But I do think that the wording that says or other healthcare providers, such as a nurse, confounds the this may not just be one practice. And it broadens that question to it could be the WIC nurse. It could be the school nurse. It could be -- you know it then takes it out of are we just measuring what is happening within a practice.

DR. NISHIMI: Excuse me. Can I ask what question you are seeing that on the thing about a nurse? Because I don't see it on the version of the survey that I have. So could you

direct me to where you are seeing that, please?

I see doctors or other healthcare providers but I don't see including a nurse.

MEMBER MODAK: I'm looking at the NCF Quality Measure ask about and address parental concerns Data Dictionary.

MEMBER FINKELSTEIN: So the two things are he is looking at the Data Dictionary --

DR. NISHIMI: Oh.

MEMBER FINKELSTEIN: And I am looking at the screen shots that are on page 32 of the attachment.

DR. NISHIMI: Okay, thank you. Yes, our tool doesn't have nurses in there.

CO-CHAIR BROOKEY: Oh, ok. Amy.

MEMBER HOUTROW: So I mean I just want to bring us back. This is not being changed to a process measure. It is staying as an outcome measure, for which the criteria to judge evidence is not a lot of criteria. We just have to say yes or no, do we think that an answer to this survey has something that is actionable at any

sort of level in the healthcare system. It doesn't even need to be the personal provider, really and certainly it is.

If no one is doing anticipatory guidance on a kid and they are in your practice, well then start doing some anticipatory guidance. I mean it is actionable.

And I feel like if we don't change -- this is not moving to a process measure. I think we have to respect the fact that we have to vote on it as an outcome measure.

CO-CHAIR BROOKEY: So I don't know if there is anything new that can be said about this. I think we all understand exactly what we are talking about but this is an individual provider report card. So I just want to keep on saying that. This is going back as an individual --- they are going to open it up and look at this and think it reflects their care.

So unless there is something new, I would suggest that we vote on evidence, unless anybody disagrees with that.

Okay, for number 2 -- 3220.

MS. JUNG: This is for Measure 3220, so that would be the second question in the set of five or second measure.

Okay, so voting for Measure 3220 for evidence is now open. Option 1, pass; option 2, not pass.

Okay, voting is now closed. The results are 45 percent for pass with 10 votes; 55 percent not pass with 12 votes. And with that, consensus has not been reached.

CO-CHAIR BROOKEY: So we move forward?

MS. JUNG: We do.

CO-CHAIR BROOKEY: Okay.

So now let's talk about gap. And I have lost track of who are the discussants. Is it Lauren or it is Deborah?

MEMBER AGORATUS: It's me. It's Lauren.

CO-CHAIR BROOKEY: Okay.

MEMBER AGORATUS: So under the gap, it is noted that, again the top five providers had

the range of 64.9 to 92.3 percent. Then they looked at the Kaiser study and it said approximately half, 53.3 percent of parents reported that they were asked whether or not there were any concerns. Disparities were race and ethnicity. And the level for risk for at-risk for developmental, behavioral, social delays and another study showed age and birth order as having an effect.

Let me just look here. I thought it was a little vague and I noticed some of the other folks also commented on this that you need to clarify this not just concerns but specific to learning development and behavior and that there also needs to be a response. So I think asking is one thing. What is the response or what resources are the parents being given after they are being asked this.

And that's it. That's all I have.

CO-CHAIR BROOKEY: Debbie.

MEMBER FATTORI: I think my only question goes back to the language issue. So

there was a disparity noted between Hispanics but I am curious to know how we can validate that or verify that if the survey isn't going out in Spanish.

CO-CHAIR BROOKEY: So is that -- that is sort of a question mark for a gap.

MEMBER FATTORI: Yes.

CO-CHAIR BROOKEY: That is also a question mark for validity, right?

MEMBER FATTORI: Right.

CO-CHAIR BROOKEY: Ok.

MEMBER FATTORI: Well, it could be both but for this segment of the conversation, the developer has listed this as a gap. However, I am not quite sure how you can tell.

CO-CHAIR BROOKEY: I'm going to go ahead and let the floor make comments and then I will go around.

So, John, I think you are next.

MEMBER FINKELSTEIN: I don't think I have anything to add. I think they demonstrated some gap.

CO-CHAIR BROOKEY: Okay. And then Jeff?

CO-CHAIR SUSMAN: Gap.

CO-CHAIR BROOKEY: Okay. David?

MEMBER KELLER: I was just going to say that it is likely that a large number of Latino or Hispanic folks speak English. So you can, at least in our surveys, we often are able to identify a Hispanic population that is English speaking. I agree that it would be better to have also data on the Spanish-speaking Latino population but it is not essential.

MEMBER FATTORI: Agreed. We can't make that determination. That should be -- the survey they get should be based on their primary language, not necessarily race.

CO-CHAIR BROOKEY: Their preferred language, yes.

Any other questions about gap or should we vote on gap? Jill, I'm sorry.

MEMBER MORROW-GORTON: I just wanted to clarify that I read this right that this was

done, the gap was done only on five providers. Just the top five. Is that correct?

MEMBER FATTORI: It was the Kaiser Permanente Northwest study as well.

MEMBER MORROW-GORTON: Yes, well the Kaiser Permanente study just has one number.

MEMBER FATTORI: Right in that study about half didn't get it.

MEMBER MORROW-GORTON: Right.

CO-CHAIR BROOKEY: Okay, vote on gap.

MS. JUNG: Okay, the voting for Measure 3220 for performance gap is now open. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

Okay, voting is now closed. The results are 5 percent for high with one vote; 82 percent for moderate with 18 votes; 14 percent for low with 3 votes; and zero percent for insufficient with zero votes.

And with that 82 percent for moderate, the performance gap passes.

CO-CHAIR BROOKEY: Okay, so let's go

to reliability. Is that Lauren or Debbie?

MEMBER AGORATUS: Yes, it's Lauren and I have under the specifications just a little reminder that the developer stated the missing data about the asked parental concerns questions were excluded from analysis but that NQF did not consider this as an exclusion.

Also under the reliability testing, the developer mentioned there were three different patient-centered strategies used. They did not report on the data element, the item level reliability. They did assess the difference between measured objects and inter-unit reliability. The ICC was 0.78, where 0.74 is considered excellent.

And I think -- well, am I supposed to stop there or are we supposed to do all of validity? I'm sorry, I'm losing track of where we are.

CO-CHAIR BROOKEY: Reliability.

MEMBER AGORATUS: I'm sorry. Done. Okay.

CO-CHAIR BROOKEY: Okay, Debbie.

MEMBER FATTORI: No, my comments were the same as Lauren's.

CO-CHAIR BROOKEY: Okay.

MEMBER FATTORI: I had nothing new to add.

CO-CHAIR BROOKEY: John.

MEMBER FINKELSTEIN: Just quickly, I think this goes under reliability. I understand they provided psychometric data but I think test/retest reliability in this case means it gets back to the concern raised earlier about the recall period. So if you send this out two days after the visit, after the index visit, I believe you will get different results than if you send it out three months after the visit. And since there is no limit on the time after the visit, you can send it out. I think that is a significant problem in reliability.

CO-CHAIR BROOKEY: So just to clarify, developer, what time period do we have? Is it unlimited or is there a three month, six month

window?

DR. SOLLOWAY: Well, we don't have control over when the providers send it out. It is sent out at their convenience. They decide when they are going to do that. That said, the studies that we did do were sent closely after the visit. So, it wasn't like -- I don't think it was a three-month time period. It was probably sent within a week or two. I don't know the exact time frame.

Naraa, do you know what the exact time frame might have been on that?

CO-CHAIR BROOKEY: But it sounds like there is no control over it.

DR. SOLLOWAY: Right.

CO-CHAIR BROOKEY: Probably most people will send it out soon.

MEMBER FINKELSTEIN: Right but the point is if this is for accountability and there is no control over when it gets out, I think there is a reliability problem.

CO-CHAIR BROOKEY: Okay, Jeff.

CO-CHAIR SUSMAN: Nothing to add.

CO-CHAIR BROOKEY: Open it up for committee.

So we are voting on reliability.

MS. JUNG: The voting for Measure 3220 for reliability is now open. Option 1, high; option 2, moderate; option 3, low; and option 4, insufficient.

Voting is now closed. The results are zero percent for high with zero votes; 32 percent for moderate with 7 votes; 59 percent for low with 13 votes; and 9 percent for insufficient with 2 votes.

And with that, consensus has not been reached for reliability.

CO-CHAIR BROOKEY: So we do not move --

MS. JUNG: Oh, apologies. Yes, correct, it fails.

CO-CHAIR BROOKEY: We do not move forward.

Okay, 3221, Family Centered Care. And

I will just briefly ask the developer to mention if there is, in terms of the instrument as well as the evidence, are there differences with this measure compared to the previous two?

DR. SOLLOWAY: No, this asks in the last 12 months did your child's doctors or other healthcare providers and then there are seven items: take time to understand your needs, listen carefully, show respect, et cetera. So it is still asking about you and other providers.

CO-CHAIR BROOKEY: Where is Kraig? Oh, Kraig isn't here, is he? Oh, there's Kraig. Kraig, you are the primary discussant, right?

PARTICIPANT: No, I think it is Carol. Carol Stanley.

CO-CHAIR BROOKEY: I'm sorry. I'm looking on the wrong page.

MEMBER STANLEY: Yes, it's me.

CO-CHAIR BROOKEY: Carol, it is you.

MEMBER STANLEY: Yes, it's me.

CO-CHAIR BROOKEY: I believe that I am looking at the wrong, literally looking at the

wrong page. Okay, Carol.

MEMBER STANLEY: Yes, the comments that came up were identical to the issues we have raised regarding attribution to a single provider. And you know the same limitations are there.

CO-CHAIR BROOKEY: Okay, there was some distinction between the first measure and the second measure, some minor distinction.

MEMBER STANLEY: Right.

CO-CHAIR BROOKEY: What is your assessment of this third measure?

MEMBER STANLEY: I mean it has a 12-month look back period when it goes to the parent and it seems very similar to the first question that we addressed.

This talks about whether the healthcare provider understands specific needs of the child and concerns of the parent, if the provider has helped build confidence in the parent, things are explained in a way that the parent can understand and if the provider shows

respect for a family's values, customs, and how they prefer to raise their child.

CO-CHAIR BROOKEY: Okay. All right, I'm sorry, is it Tarra or Tara?

MEMBER BRISTOL-ROUSE: It's Tara.

CO-CHAIR BROOKEY: Tara, okay.

MEMBER BRISTOL-ROUSE: But I will answer to either, especially at this point in the day.

You know I echo what Carol said. I think an additional concern for me is calling this kind of family centered care because family centered care is an approach to care. So, it includes a lot more than is even represented in these five or seven items.

And additionally, I think this maybe even more so than the other ones that we had issues with getting this and not knowing if it was specific to you as a provider will be very difficult to address for improvement.

CO-CHAIR BROOKEY: Okay, good comments.

Let's see, we also had Jill on the group. I'm sorry.

MEMBER STANLEY: And this is another, you have to answer at least four of the questions out of the seven, the seven areas it addresses.

MEMBER MORROW-GORTON: Yes, and I was going to say that having seven areas and having them sort of they are not -- I mean they are sort of all in the same ballpark. I think it would be -- it might be difficult for people to answer them and there is seven. And then what do you do with -- you know do you get them as each individual one? Is that helpful? And how do you actually use that information.

CO-CHAIR BROOKEY: Right but you may not get the item level result.

MEMBER STANLEY: And the developer specifically says in here that there was no improvement among providers who participated in the study with these questions over time.

CO-CHAIR BROOKEY: I don't have any additional comments. Amy, are you in this group,

too?

MEMBER HOUTROW: No, I was just going to make the point that in fact the family center care measures are conceptually very similar to the ask about parental concerns, which is a part of family centered care. So it is the same discussion that we would have for this set would be appropriate for the discussion that we just had on the previous one.

CO-CHAIR BROOKEY: Any other comments?

MEMBER AGORATUS: This is Lauren. I just have one clarifying question. On all of these, and I know we didn't get this far, but on all of these they discussed there was no automated reporting system and there was supposed to be a website launch in February 2017. Do we know if that happened?

DR. SOLLOWAY: It did not. We got a little bit delayed but it will happen this month.

MEMBER AGORATUS: Okay.

CO-CHAIR BROOKEY: All right.

DR. SOLLOWAY: The way -- okay.

CO-CHAIR BROOKEY: Any other questions or comments from the group?

Okay, so we will vote on the evidence as an outcome measure for 3221.

MS. JUNG: Okay, voting for Measure 3221 for evidence is now open. Option 1, pass; option 2, not pass.

We're looking for one more vote. Did anyone step out? Oh, there we go.

Voting is closed. We have the results are 27 percent for pass with 6 votes and 73 percent for not pass with 16 votes. And with that, the measure does not pass for evidence.

CO-CHAIR BROOKEY: So, I am going to preface this by saying that I don't want to rush through this. If we can't finish on time, we will finish on the phone call because I want to be sure that we get everyone's input and not just kind of try to get this done by 5:00. And if people need to leave and we don't have quorum, we will conclude early.

So having said that, this is a similar

measure. And Kraig, I think you are primary discussant on this one.

MEMBER KNUDSEN: All right. So this is Measure 3222: Assessment of Family Alcohol Use, Substance Abuse, and Safety.

In terms of evidence, again, this is the same issues that we have experienced with the other ones. This measure will evaluate the proportion of kids whose parents report being assessed for three items, alcohol use, substance abuse, and fire arms in the home. It can be used by providers to determine the level at which they discuss these issues with the parents.

The developer provided a logic model with all of the measures, the same thing, which connects the outcomes to provider behavior in terms of quality improvement.

In terms of this one, the developer indicated that the American Academy of Pediatrics and Child Health Bureau Bright Futures Guidelines include assessment of alcohol and drug use, the presence of guns, family violence, and other

safety issues in the family.

So, that's what they said.

CO-CHAIR BROOKEY: And your assessment, based on previous conversations, does this relate to this measure as well?

MEMBER KNUDSEN: Yes, it does.

CO-CHAIR BROOKEY: Okay.

MEMBER KNUDSEN: I believe it does.

CO-CHAIR BROOKEY: Okay. Let's see, Karen, I think you are the --

MEMBER DORSEY: I agree, it is the same.

CO-CHAIR BROOKEY: Okay, very good. And I agree. No more comments from me.

Any other comments from the group or on the phone? David.

MEMBER KELLER: One question for the provider or for the developers. I was wondering why you split this from the next measure that we are about to talk about, which is also about family well-being. It would seem to me that we actually ask these questions routinely as part of

a screener that we have but we have it all lumped together as a single thing because we are asking about home safety and basically how the parents are doing.

CO-CHAIR BROOKEY: Yes, developer, any rationale why these were split, 3222 and 3223?

DR. SOLLOWAY: I don't know what that is. I know there was one but I don't know what that is. I can't really speak to it.

Naraa, do you know if there is any rationale?

DR. GOMBOJAV: No.

DR. SOLLOWAY: It may have just been specific content. You know one is really about psychosocial screening and emotional well-being versus something outside of that.

CO-CHAIR BROOKEY: Okay, that's fair enough.

Okay, any other comments about the evidence on this particular measure? Any objection to moving to vote?

Okay.

MS. JUNG: Voting for Measure 3222 is now open for evidence. Option 1, pass; option 2, not pass.

Again, looking for one more vote. Do we have everyone in the room? There we go.

Voting is now closed. The results are 32 percent for pass with 7 votes; 68 percent for not pass with 15 votes. And for that, it does not pass on evidence.

CO-CHAIR BROOKEY: Okay, thank you.

Why don't we go ahead and start 3223? If we can't finish, we will do it on the phone call but I suggest we go ahead.

I believe, David, you are -- are you able to stay for a few minutes?

MEMBER KELLER: Yes, no, I'm here. I'm tracking. I have got Google Maps set and I'm tracking how long it will take me to get to Dulles for my 7:00 plane and I can still make it.

CO-CHAIR BROOKEY: I should have put you out of order but go ahead.

MEMBER KELLER: I enjoy living close

to the edge. What can I say? To my wife's eternal regret, but still.

And actually this, given the last vote, I think this will take very little time because really this is the same. It is, as I said in my question, it is a parsing of similar questions that were aimed at family -- at parental well-being. We think they are important questions for assessing the environment in which a child is being raised. And they are part of Bright Futures. They are endorsed by the Academy of Pediatrics, as well as HERSA in its Bright Futures Manual.

And basically, but in the same way they ask over the last 12 months has your doctor or any other provider addressed these issues.

And so the evidence presented is really from the same surveys that we have already talked about in detail, I think. And I don't think I have anything else to add -- this really falls into that same bucket of similar surveys, the last four that we voted on.

CO-CHAIR BROOKEY: Maureen, do you have comments?

MEMBER EDIGER: In the interest of time, no. I will keep them to myself.

MEMBER KELLER: Well, we talked about this at the break a little bit, too.

MEMBER EDIGER: Yes.

MEMBER KELLER: Because we are just starting actually some -- you all know Maureen she is at our hospital. And we are starting now, in one of the clinics she goes to, to ask these questions routinely of every family that comes in.

She hasn't come in since we started doing that.

MEMBER EDIGER: But boy, will I!

MEMBER KELLER: I suspect she will have some feedback for us when she does.

MEMBER EDIGER: Yes, I think this is just -- I will do a plug for parents being involved in the process because I think the more that families are involved in the process,

families are going to raise their hand and go if you are doing something like this, what is in it for me as far as why would I spend 20 minutes on this if you are giving me a tool that is going to possibly bring up a lot of emotions.

Just having a child with a mental health diagnosis, if I sit down and I have to start kind of like regurgitating and telling somebody about how hard things are at home sometimes and then I hit send, I am going to be sitting there pretty emotionally raw. And it is like and then what.

So, that's all.

MEMBER KELLER: For the record, we did have parents involved. We just didn't have you involved.

MEMBER EDIGER: Right.

CO-CHAIR BROOKEY: Thanks, Maureen. Other David.

MEMBER EINZIG: I don't have anything.

CO-CHAIR BROOKEY: You have nothing to add? Okay, Karen.

MEMBER DORSEY: I have nothing to say.

CO-CHAIR BROOKEY: Okay. Any other comments from the group? You all look a little worn down.

PARTICIPANT: Bring us home, John.

CO-CHAIR BROOKEY: Yes, I feel kind of depressed after this whole string of measures we are not endorsing. But I hope that we field the process before we vote on the last measure. I guess after we vote on the last measure we can take one minute to get feedback about process.

So, I think we are ready for a vote, unless I am hearing any objections.

So this is the final measure, 3223.

MS. JUNG: Yes, the voting for Measure 3223 for evidence is now open. Option 1, pass; option 2, not pass.

CO-CHAIR BROOKEY: Do we have member comment at the end?

DR. NISHIMI: Yes.

CO-CHAIR BROOKEY: Okay.

MS. JUNG: Voting is now closed. The results are 36 percent for pass with 8 votes; 64

percent not pass with 14 votes. And with that, it does not pass on evidence.

CO-CHAIR BROOKEY: Okay, if you need to go to the airport, we are, at this point, going to open it up for member comment and then I would like it if people could stick around for a couple minutes just to talk about feedback.

So, you are excused.

MEMBER KELLER: Well may I make one comment --

CO-CHAIR BROOKEY: Absolutely.

MEMBER KELLER: -- as I pack up to head off to Dulles.

Just I wanted to express my appreciation for the thoughtfulness of this committee and I wanted to thank CMMI for reconvening us and I hope that we continue these discussions because we really do need measures of these things and we need to be able to force them going forward.

So, thanks to everybody for today. I appreciate your work.

PARTICIPANT: Sing it, sister!

CO-CHAIR BROOKEY: Thank you, David.

MEMBER FINKELSTEIN: John, do you want to excuse the developer?

CO-CHAIR BROOKEY: I'm sorry.

MEMBER FINKELSTEIN: Is the developer still on the line?

CO-CHAIR BROOKEY: I'm sorry, the developer, we are concluding these measurement discussions. So you can sign off, if you haven't already.

DR. SOLLOWAY: Okay, thank you so much. I appreciate your feedback and we will definitely take it and see what we can do. Thank you.

CO-CHAIR BROOKEY: Thank you.

Do we have member comments?

PARTICIPANT: Public also.

CO-CHAIR BROOKEY: Public comments?

OPERATOR: If you would like to ask any public comments, please press *1. Press *1 to ask a public comment.

And there are no public comments.

CO-CHAIR BROOKEY: There are no comments? Okay.

I don't know if it was because I got up at 3:00 California time or it is just this meeting but I am kind of spent.

So I would like to first just ask for any feedback about the fairness of the process today, if there is any concerns about anything. These meetings are always very interesting. I just appreciate the group process because if we didn't have individuals on this group, we wouldn't have known as much as we know about these measures now. I don't think individually any of us are able to figure this all out. It took the group for us to figure this out. And so I appreciate everyone's contribution.

But let me just throw it back at you. Thoughts, comments, just for a couple of minutes, if we could.

Oh, John has his --

MEMBER FINKELSTEIN: So I would just

echo that I thought it was really a great day and people being thoughtful all around the table.

For this last set of measures, which was kind of an unusual circumstance, I think it is only fair to feedback to the developers that if they want to come back as a process measure with the caveats that -- bottom line is I think they might get through on evidence as a process measure but they have real validity problems as an accountability measure. So I just think there should be a discussion and counseling with them and they should be able to come back.

CO-CHAIR BROOKEY: So do you think that if they changed it, perhaps to a process and maybe changed the level of reporting maybe it might fly?

MEMBER FINKELSTEIN: Right and they would have to provide all the evidence that is behind the Bright Futures ---

CO-CHAIR BROOKEY: Right.

MEMBER FINKELSTEIN: --- elements in order for us to look at them. Anyway, I just

think we should give them a chance.

CO-CHAIR BROOKEY: I think they deserve our feedback and I agree with you. And I hope that we captured that.

Okay. So, Rajiv.

MEMBER MODAK: So I just wanted to say this is my first meeting and I really found it interesting and valuable and the discussions were right on.

And what I just want to express is that just even the discussions just gave me a lot of ideas to take back to my practice and to implement in my practice. So, I just want to give you that feedback that you will help my quality improvement.

CO-CHAIR BROOKEY: Alright, Jim.

MEMBER BOST: Yes, I think that for the last set of measures that if a rate is low, there is something you could do to improve. But I don't think these rates, with the questions the way they are written could ever be used for something like Pay-for-Performance or

accountability.

CO-CHAIR BROOKEY: Yes. Amy.

MEMBER HOUTROW: My comment is just in general about this issue about the level of evidence that you need for an outcome measure versus a process measure, which is an NQF issue.

But choosing a patient-reported outcome as a measure of a process of care is kind of a backdoor way to not to provide the rigorous amount of evidence that is required for a process measure and it makes me just a little bit uncomfortable.

And then it also makes these discussions in the room really challenging because we are cognitively thinking about it in a way that is different than they are potentially presented to us.

CO-CHAIR BROOKEY: Yes, thank you.

Let's go this way. Debbie.

MEMBER FATTORI: So I am often accused of falling on the sword and I am going to do it now too, again, because I feel like I have some

personal accountability in this. Because I think that these were some of the questions in my head as I was reviewing this and I am just wondering if there is something that we can do as individuals or a team that we are kind of stopping the line, if you will, before we end up at this point in time.

I don't know if that makes any sense but I went through the process and some of those questions are nagging. Maybe I am not experienced enough in this work but I am just wondering if there is something that we can do before it gets to this point because we had this issue I think with a measure or two last time around.

CO-CHAIR BROOKEY: Yes.

MEMBER FATTORI: So, that is my only comment.

CO-CHAIR BROOKEY: I think all of us had the same thought, what could we have done to have prevented this. And again, I think that this was a situation that none of us really saw the huge big picture of what was going on here.

But I think it took the group process to figure that out. Unfortunately it was for five measures. So that is the unfortunate part.

Kerri.

MEMBER FEI: Yes, so back to those five measures, when it comes to putting on my health plan hat, I suppose, and if you are talking about use and accountability, what we also need to see is that I can know for a fact that an individual provider or a group is statistically worse, better, or the same as some type of benchmark. Right?

When we use them at the Association, that is how we look at accountability. The way it is now, we couldn't -- we would not be able to do that.

So, with always having this need of the NQF measures should have --- be appropriate for accountability, the way it is now, they will need to address that if they want them used for accountability as well.

So you will need to be able to see

that statistical significance that you can discern that, I think, between whether it is provider or group, or however they decide to do it, you need to see that back to the insured coverage measure as well.

I feel bad I think he was a little bit upset with me.

PARTICIPANT: I think he was upset with everybody.

MEMBER FEI: And I feel terrible because that was not my intent. But from our perspective, I don't know with that chunk that this state is better --- statistically worse, no different, or better than the ones next to it. And that is what I was trying to get at.

CO-CHAIR BROOKEY: From my seat --

MEMBER FEI: I feel bad that he was mad. That was not my intent.

CO-CHAIR BROOKEY: From my seat, that was a very unexpected but rich discussion and I think that he listened to the feedback. So I think --

MEMBER FEI: Well and again, I think it is one of those where maybe that measure isn't right for that type of accountability but it could be so fruitful and useful from other perspectives.

CO-CHAIR BROOKEY: Thank you.

Jill.

MEMBER MORROW-GORTON: And I actually wanted us to think about accountability from a broader perspective. Because I think we had a number, it wasn't just kind of the last set or the last set plus the state measure but we had this conversation even this morning in some of the measures that were a little more straightforward.

Should we be adding accountability, or a question about accountability, or does this need an accountability criteria or something like that as we are thinking about these? They are likely to be used for accountability.

CO-CHAIR BROOKEY: Yes.

MEMBER MORROW-GORTON: And if they are

not designed for accountability and they can't distinguish, the people that are sometimes choosing them don't know that and don't understand that. And that is a piece of information that I think would be really helpful.

CO-CHAIR BROOKEY: Yes.

MEMBER FEI: That's absolutely true because I am the one that gets tagged with that. I have folks say to me I want to use this measure. No, you can't, it is not right for that. But I want to. But no, you can't. It's not right. It's not fair. So, not that everything is fair but they have meaning.

MEMBER MORROW-GORTON: Well at least try to be fair.

MEMBER FEI: Well absolutely, especially if someone is going to tie your money to it.

CO-CHAIR BROOKEY: Karen.

MEMBER MORROW-GORTON: Right and that is the biggest piece of it is tying money to it.

CO-CHAIR BROOKEY: Believe me, I would

have 9,000 doctors chasing after me if some of these got out of Kaiser. So, yes.

Karen.

MEMBER DORSEY: So just I want to thank you both for your leadership. I thought this was really as smooth as it could have been today, given what has already been said about some of these issues just coming out when we start to talk in groups. I appreciate that. And as usual, it was a wonderful experience to convene with this committee.

I think two things jumped out at me today. One is that I was a little hazy on the guidance around the reliability -- the statistical approach to assessing measure reliability for a couple of these measures. It came up with more than one of the measures. And you know I think we have had this experience this time and the last time where we have a lot of folks who are coming new to the NQF process, to the measure endorsement process, they are newly into measure development. That is just where we

are in pediatric measures right now. And I think the extent that we and NQF staff can really give people guidance about what we mean by a precision assessment and how they should approach it, I think that will be a good service to people.

Because I think there were some good measures that came forward but there were some questions about the specifics of the testing. So I think that is one thing that stood out to me.

And the other is that there is still a lot of haze. And I think this speaks to the last two comments about use and usability issues around these pediatric measures. They are attributed at all different levels. We think about their use in different contexts. We think about maybe assessment of managed care organizations, for the enrollment, and we are thinking about providers for some of these last measures we discussed. And I think that that creates a lot of sometimes confusion and issues for us to work out when we are talking about the measures around whether we think they are

valuable and if they can be used in ways that are going to be productive.

And because it is so complex, I just sort of long for us to have some structure, some conceptual structure around it. And I think I am hearing that from you all, too, around issues of accountability but it is a very different thing to think about parent reported survey information at the provider level and a plan that is going to be held accountable for how well it keeps continuity and enrollment. You know they are just totally different universes. Right? And I think, hopefully, as a standing committee, we will start to develop some guidelines around how we think about it.

CO-CHAIR BROOKEY: So, Karen, to your first point, and I am just going to get shot by the NQF staff here, but if there was an opportunity for those who were interested to have a little tutorial like an hour or two WebEx on how to interpret some of this stuff, do you think some folks would be interested in that? I mean

just some feedback to the staff.

I think your point is really well -- I think the orientation was good but it probably is not deep enough for some of us. And so I think that might be helpful.

All right, did everyone have a chance?

Jeff, go ahead.

CO-CHAIR SUSMAN: Well first of all, let me thank all of you. You have just been splendid. As usual, I have learned a lot more from being here. I also want to thank the NQF staff who worked just incredibly to try to put these materials together.

In response to what I heard around the table, I do think that the NQF, as an organization, has to look at this issue of outcome versus process measures and really get it down better than we have now. It is difficult but we shouldn't be resolving these issues at the time of our meetings. It is not fair to the people who put a lot of energy and time into the development.

Secondly, I think it really would be very helpful to have some materials that really define, as you have suggested, Karen, like precision. Here are the three ways that we look at this issue. This is what it means and this is what we find acceptable. Or at least some guidance, recognizing that each panel might have certain nuances that need to be brought to bear.

And then finally, the issue which I talked to Shantanu is about the idea of accountability and having wonderful measures, which we all think would help improve health within our systems, or within our states, but are concerned about accountability. And having a more nuanced approach to that issue, so that as we vote, as we are trying to make some difficult decisions, it is not a black and white situation where we are either discarding measures which could really help improve individual performance or plan performance, what have you, or perhaps really should be leading the way and pushing people toward greater accountability.

Again, overall, thank you so much. It is really a pleasure to work with such smart and dedicated people.

CO-CHAIR BROOKEY: Robyn.

DR. NISHIMI: The last word? On behalf of NQF, I really want to thank you. Your discussions today were spot on. You struggled with the issues that I will tell you, having been the founding chief operating officer, every single committee struggles with these things because there are no bright lines or firm guidance.

The first project we ever did at NQF was the so-called Never Events List. Adverse events that should be publicly reported. We weren't even talking about payment then.

So, it is the same conversations and you shouldn't feel good or bad that you had to have them. It is just the way the field is and is evolving. So, I did want to mention that.

The same thing with the PRO versus PROP, patient reported outcome on a process. We

did recognize that this was the committee that raised this as an issue last year that there is something about the level of evidence that should be required for these that is different than a yes/no.

And so some internal processes have begun to discuss those. They just haven't been completed. We haven't landed on one place. And so to be consistent across projects, frankly, that is why you saw what you saw.

But I think we heard it again very loud and clear it is a struggle and the groups that are responsible internally for looking at our algorithm and our criteria and what we require, we are going to take it back and say it happened again and I had to sit there and listen to it again. So let's -- which is not to say that we shouldn't have or couldn't have moved faster but, again, it is the whole milieu we have felt the consistency. And so I did want to let you know that we had heard you the first time and there are people absolutely thinking about how

best to approach this.

And then the last thing I wanted to just stress because it peeked my interest was the use and usability. I congratulate you on the robustness of the discussions that you had because that is an often sort of set aside, kind of blow through the criterion discussion but you were very thoughtful there. And it is a struggle. And it is especially a struggle because all these measures, with the exception of the CAHMI and for them not for purposes of accountability, and at that level of analysis, but they haven't been in use. But you raised really good, thoughtful questions. I think they were good for the developer to hear and they were certainly good from the NQF staff point of view as we think about that whole criterion. So, that I was especially appreciative of and thank you for that.

That's it from me.

CO-CHAIR BROOKEY: So, I will just echo everything that Jeff said. You know this is

the second time that he and I have shared this meeting together and it is really a privilege to be able to do this and to interact with all of you. It is amazing to watch how things sort of go this way and that way and then back to some area that you didn't think it was going to go into. But at the end of the day, I think it is a good process and it has to be all of you together, including our parents in the back. They are sitting way in the back but next year you can be up here, okay? It is important to have you in the room and I appreciate that.

So, again, thanks, everyone and we are just going to go through a few points for coming up agenda items, our time line and then we will be dismissed.

So, go ahead.

MS. JUNG: I promise I will be brief.

Okay, so upcoming for the time line for our measures, we just have the comment period after this, which we will open to the public and to NQF members. Then we will have the NQF member

vote. And then after that, we will have the Consensus Standards Approval Committee or CSAC Committee vote for endorsement and that will be on June 27th. And after that, there will be a 30-day appeals process.

Special dates for you to note is there will be a post-meeting call Friday, March 10th and that will be from noon to 2:00 p.m. That should be on your calendars. If any of these events aren't on those calendars --

DR. NISHIMI: Madison, we don't need to have that one.

MS. JUNG: Oh?

DR. NISHIMI: We went through all the measures today.

(Off-microphone comment)

DR. NISHIMI: Right but you need to get your comments back in and then you will -- yes. So we don't expect them to bring it back in the next week.

(Off-microphone comment)

DR. NISHIMI: That is correct, Gary,

we were much harder on you.

CO-CHAIR BROOKEY: So, cancel next week's meeting.

MS. JUNG: Yes, okay. You will also get a calendar invitation for that. So, congrats.

After the comment period, we will have a post-comment call and that will be Wednesday, May 31st and that will be from 2:00 to 4:00 p.m.

And with that, I just want to say there is our contact information and the SharePoint link and the meeting materials will be posted on the SharePoint for you all to find.

And with that, I would just like to say a huge thank you from the NQF team for your high level of engagement today. I know it was a very long day and a bit tedious but we thank you very much for all your hard work.

Thank you.

CO-CHAIR BROOKEY: All right, safe travels.

(Whereupon, the above-entitled matter went off the record at 5:01 p.m.)

**M**

**U**

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Pediatrics Performance Measures Committee

Before: NQF

Date: 03-02-17

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