NATIONAL QUALITY FORUM

+ + + + +

PEDIATRIC MEASURES STEERING COMMITTEE

+ + + + +

WEDNESDAY
DECEMBER 2, 2015

+ + + + +

The Steering Committee met at the National Quality Forum, 9th Floor Conference Room, 1030 15th Street, N.W., Washington, D.C., at 7:30 a.m., John Brookey and Jeffrey Susman, Co-Chairs, presiding.

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

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

NQF STAFF:

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

ALSO PRESENT:

MARY BARTON, MD, MPP, NCQA
SEPHEEN BYRON, MHS, NCQA
LARRY KLEINMAN, MD, MPH, CAPQuaM
MOLLY FINNERTY, MD,
KARISHMA KUMAR, MPH, University of California San Francisco
CASEY LION, MD, MPH, Seattle Children's Research Institute
MICHELLE MACY, MD, University of Michigan
RITA MANGIONE-SMITH MD, MPH, Seattle Children's Research Institute
SARAH REEVES, PhD, Q-METRIC - University of Michigan
EYAL SHEMESH, MD, CAPQuaM
SARAH HUDSON SCHOLLE, MPH, DRPH, NCQA
REBECCA SMITH-BINDMAN, MD, University of California San Francisco

*** present by teleconference**

CONTENTS

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

(7:31 a.m.)

CO-CHAIR BROOKEY: Good morning, everybody. We have a quorum, so we're going to go ahead and get started. Do we want our developers to come up?

DR. NISHIMI: No, we first have to go to Marcia. We have two new committee members.

CO-CHAIR BROOKEY: We have a few things to do first. First of all, we have a couple people here that weren't here yesterday, and we have somebody here today that was on the phone. Marcia, I'm going to turn it over to you for disclosures.

DR. WILSON: Thank you so much, John. We just have two committee members who have joined us today, and we need to do our disclosures of interest. Jim is here in the room, and I believe on the phone, do we also have our other committee person? Keith, are you on the phone yet?

MEMBER WHITE: I am. Can you hear me?

DR. WILSON: Yes, I can. Thank you very much. We're just going to do a brief oral disclosure of interest. This is any work, paid or unpaid, that is relevant to the measures coming before the committee. It doesn't need to be a complete summary of your experience, but any relevant activities that you have to disclose, and we do this in the spirit of transparency. Jim, if I could go to you first, if you would introduce yourself, tell us where you're from, and if you have anything to disclose.

MEMBER DUNCAN: Sure. I'm Jim Duncan. I'm a radiologist from Washington University in St. Louis. I have a conflict of interest with the measure submitted by UCSF. I believe it's 2820, CT Radiation Dose.

DR. WILSON: Thank you, Jim. Keith, if you could introduce yourself and disclose if you have anything.

MEMBER WHITE: My name is Keith White. I'm a radiologist, medical imaging director for Intermountain Healthcare, and I have no

disclosures.

DR. WILSON: Thank you very much, Keith.

DR. NISHIMI: Then we have one new member who was on the phone yesterday, who joined us. Marlene, did you just want to -- you don't have anything to disclose, but if you wanted to just --

MEMBER MILLER: No, but I am the voice on the phone.

CO-CHAIR BROOKEY: All right. We also got out of here so quickly last night we failed to allow for public input, so that's our bad, but we will make up. What I would recommend doing, because we're starting early, is to get through -- to finish up the FECC measures, and then we will have public input. We'll do that, Marcia, if you can help us with that, to remind us. I just want to say that I have been part of many groups/committees over the years.

I should say this at the end of the day, and not at the beginning, because I don't

know how you're going to be by 3:00, but I think that this is a very well-functioning group. I really appreciate working with all of you. The comments were really very valuable. I think everyone was respectful of each other, didn't talk over each other. I think the group dynamics have been very good, so I just wanted to say it's been a real privilege working with you. We do have a long day ahead of us, so we're going to try to finish by 3:00. We had this existential discussion yesterday about whether self-reported outcome can be a process measure, vice versa. I think we all struggled with this yesterday. Quite honestly, I think that from what I understand over the years, the committees have approached this differently and not consistently.

We clearly had an inconsistency with our approach yesterday, by no one's fault, but it just turned out that I think people were thinking one thing would flow down one pathway, and something else would flow in a different pathway. I don't think there's a right or wrong answer

here, quite honestly. I think we can argue that patient-reported outcomes can be -- if you're reporting a process -- I think that John said it well. We do this in Kaiser.

If you say you've had a mammogram, we will count that as a process measure for that woman. I think it's a very good example of that. But we're not going to debate that any further because I think we do need to move on. The good news is that I think that we could probably be consistent with our approach and not completely rehash every conversation from yesterday. What the staff have proposed, and what I want to see if you would agree to, is to take through the FECC measures that did not completely pass -- those are the gray zone or the fails -- and just revisit those and vote them up or down, going along the top pathway. That would be FECC Measures 7, 9, 14, and 17.

I will tell you that one that we had a pretty clear pass on -- pretty close to pass, 58 percent, so it was very close to passing. The

other three were much less favorable. Many of them were sort of felt to be insufficient. What I would recommend that we do is to go through those four. We'll go through how we voted on them. We'll ask for any further comment, and we will vote them up or down based on this being a survey of patient-reported outcome. Any thoughts about that process to get us moving forward? Marlene?

MEMBER MILLER: I'm really kind of confused why we would redo this, and then would we judge all the rest of the FECC measures? It'd be inconsistent to do half of them one way, half another. I'm really confused why we did this.

CO-CHAIR BROOKEY: I think the rationale is that if we pass them to begin with, the likelihood is -- if they passed on the vertical pathway, the chances are they would pass on the horizontal pathway. That's just our thinking about efficiency here. Any other comments?

CO-CHAIR SUSMAN: I think the bottom

line is it seems more stringent when we've done it down that vertical pathway and that the simple yes/no, up/down vote was much less stringent and, therefore, if things passed already, indeed they should easily pass again. The only ones that we'd reconsider are those that were on the borderline.

CO-CHAIR BROOKEY: Jon?

MEMBER FINKELSTEIN: I think as a way to move forward, that's fine with me, but I think somewhere in the minutes, I think there's a difference between outcome measures and process measures. We can't call everything that's patient reported a patient-reported outcome measure. My example is if the question is did you get a mammogram in the last 12 months, and you ask the patient, that doesn't make it an outcome measure. I don't disagree, but I think we need to keep that straight in our thinking at some point. Maybe that's feedback to NQF.

CO-CHAIR BROOKEY: I don't think we're going to resolve this variation that we've had in

approaching these measures. We can't really solve that today. I think the NQF staff need to kind of go back and noodle on that a bit. But I do want us to move forward so we can get to the other measures this morning. David?

MEMBER KELLER: I don't want to hold up that process, but just as an observation, what I think the difference is is the more complex pathway that involves reviewing evidence seems fundamentally different from a pathway that, in essence, is just asking us to propose a logic model, that we could draw a Driver diagram.

That's a different standard of evidence. We need to think through whether there's a way of making that more -- I think there's a way of making that more rigorous that involves evidence to support the Driver diagram.

CO-CHAIR BROOKEY: I think that's why we struggled so much with the adolescent measures yesterday is we felt like we had a different standard for that than we did for the FECC measures. Amy?

MEMBER HOUTROW: I guess I'd just like to echo that concern. It's not rigorous to say is there a rationale? It's not also how they were presented to us as measures. They were not presented to us as patient-reported outcomes, so I have a lot of discomfort with taking -- I understand why we needed to do it with ADAPT because that's how it was presented to us.

But taking out the ability to review the evidence for why we would do something by taking a pathway that wasn't the intended pathway by the developers makes me very concerned that we're doing something that undermines the basic process of what we were asked to do, which was to review the evidence for these things.

This up and down, is there a rationale for the relationship by the patient saying I had a mammogram in the process of them having a mammogram, it's just circular. You can't skate away from answering yes to that second question. If we're going to say that a patient-reported outcome is do you have a case manager, and they

say yes, that is circularly related to whether or not they have a case manager.

The rationale is a one to one, as Virginia was saying yesterday. So now we have to ignore all of the evidence, and I don't think that is appropriate for what we were tasked to do. I was uncomfortable yesterday with that process, and I'm even more uncomfortable with it now.

CO-CHAIR BROOKEY: Marlene?

MEMBER MILLER: I was just going to say the same thing. The developers didn't submit them as outcomes measures, and I feel really uncomfortable changing the rules midstream. I think ADAPT should not have been reviewed as outcomes because I think we have to rely on evidence for all of these measures, but to rejudge and change the way they were developed just doesn't seem right to me.

CO-CHAIR BROOKEY: Go ahead, Karen.

MEMBER DORSEY: Just to reflect on Helen's comments yesterday, I think that there's

a purposeful difference in the standard for an outcome measure versus a process measure. It's even more of a wrinkle to think about this as a patient experience measure, which I think is another bar.

I think when we talk about outcome measures, just to use that as an example, we do talk about the rationale or logic model. That's the bar for outcome measures. But we also talk about the importance, like why are we measuring this outcome? I think that these have to be held to that standard, even with respect to patient experience.

Is there value in measuring this aspect of patient experience? I felt yesterday that the adolescent measure met that bar because this is a structural fact of the healthcare system. Everybody's going to go through this. It's valuable to understand patients' experience of this, and we think that the healthcare processes can affect their experience.

All of those things, to me, made

sense, in terms of importance, but I think we would have to apply that same standard to the FECC group if we were going to say this is an outcome measure. It's not just the logic model, but it also is this valuable outcome or experience for us to be measuring and reporting on.

CO-CHAIR BROOKEY: Jeff.

CO-CHAIR SUSMAN: Just briefly, I think we're doing more than just that one to one. I think we're trying to assess whether the patient-reported outcome is something that is valued by the person who we're looking at and is linked to -- hold on just a second -- and is linked to a health outcome which is amenable to -- as they say, influenced by structure, process, blah, blah, blah.

Really, there is a tie between one, this idea that it's of interest. There's going to be some patient-reported outcomes that are of no interest, that really they say, "I don't care." There can also be patient-reported

outcomes that aren't really linked to any health consequence, so there is that tie to evidence.

It has to be amenable to change through any of those pathways. I think when you start to deconstruct it there is a tie back to the evidence. It's not as clear. It's not as black and white. I understand all of our confusing with this. I know Helen, when I talked with her afterwards, understands this is a murky area which needs cleaned up. I think we either need to get on with it or decide we're not going to rerate and move forward, just given the pressure of time here today.

CO-CHAIR BROOKEY: I think that the committee can decide what they want to do. I don't think, at this point, it will make a tremendous amount of difference if we leave it the way we left it yesterday.

I think the outcome may not be very different if we revote. I am comfortable with the committee's decision. Jeff is comfortable with the committee's decision with either

revoting or just letting it rest where we left it.

Because we still have to go through all the rest of the issues, which are gap and reliability and so forth. Can we just sort of get maybe a straw vote? How many people would be comfortable leaving it where we left it yesterday and moving forward with the discussion about gap and liability and everything else?

So leaving it alone, how many would be comfortable?

(No audible response.)

CO-CHAIR BROOKEY: Does anybody want to revote?

(No audible response.)

CO-CHAIR BROOKEY: I think that we should just move forward and let it rest. We're not going to go back and revisit adolescents, even though I think some of us struggled with the decision to take it through the horizontal pathway. What I'd recommend we do is move forward with gap. I believe that for the gap,

and correct me if I'm wrong, I believe we can just vote on these in block.

DR. NISHIMI: You can decide to vote on these in block.

CO-CHAIR BROOKEY: If anybody wants to pull one out in particular, we can do it that way, but those who have looked at this closely -- I'm going to ask those who were the experts on this measure whether they would agree with that. Which of you are the ones that were reviewing it?

You agree to take them in blocks? Any disagreement to reviewing these in blocks? Who would like to talk about the performance gaps here? I can get my notes out. Who's on point for this measure?

MEMBER MILLER: I'll start.

CO-CHAIR BROOKEY: Marlene, go ahead.

MEMBER MILLER: I struggled with this one, in that there is a clear sense that there is a gap here, but there was very little data -- none, really, that provided a lot of -- because we don't have a lot of information on this, so

understanding the amount of the gap is really unknown. I just didn't see any substantial data attesting that there is a gap in this for kids with special healthcare needs or however you want to define medical complexity versus not. I think we think it's there, but there's no proof.

MEMBER FEI: Overall, this is -- probably the reason we don't have data, it's a low N overall. But while it's low occurrence, it's high impact, high resolution, high risk population. For me, that brings to the light that it's something we should be.

CO-CHAIR BROOKEY: Any other comments? Oh, sorry, Karen.

MEMBER HOUTROW: I just had a quick question for the reviewer. There's a comment that children with medical complexity comprise 13 percent of the pediatric population. I was noting that number's markedly higher than previous estimates of complexity and how that's being used. It's on Page 9, I think.

DR. LION: The reference there was Jay

Berry's JAMA pediatrics paper in 2013 and Tamara Simon's paper in Pediatrics.

MEMBER SLAVIN: I just wanted to correct Nathan on his point, that the lack of measurement in an area that we all come to consensus on is important to measure, I think, is enough of an assessment of gap, without breaking out exactly the gap for the individual component parts. I'm comfortable with the way the developers rated it here.

CO-CHAIR BROOKEY: Jon.

MEMBER FINKELSTEIN: I think it's reasonable to use the field test results as evidence of gap. We have field test results on all of those things, and some are higher and some are lower, but they all, in my view, show some gap.

CO-CHAIR SUSMAN: It seems to me this is an area where there's sufficient evidence, given where the scientific evidence has gone. We have a field test. There may be some systematic biases in that. I suspect not.

CO-CHAIR BROOKEY: There is some data to support our intuition, I think, not a lot. Do we have enough information to vote on gap? Hearing no objections, we'll vote.

MS. CHAVEZ: We have 26, 24 in the room, two over the phone.

MS. ALLEN: Keith and Lauren, if you could please submit your vote via chat. Severa will be reading the instructions.

MS. CHAVEZ: We are now voting on gap for Measure 2770, family experiences with coordination of care measure set. The answer choices are 1 high, 2 moderate, 3 low, 4 insufficient. Voting's now open.

CO-CHAIR BROOKEY: I hope somebody is videotaping us voting because it's really pretty funny. From up here, it's hysterical. You wonder if you squeezed out that last vote.

DR. NISHIMI: We need to do a show of hands on this one. It's not just a plug it into the wall issue, apparently, so that didn't work. 1 high, moderate, 25 moderate, low, and

insufficient. We need one more vote, and that came in. Oh, Jeff is conflicted.

PARTICIPANT: He's always conflicted.

CO-CHAIR BROOKEY: We're going to move on to reliability. Robyn, I think you are recommending that we take these measure by measure, or are there any of them that can be lumped together, asking the reviewers if any can be combined, or should we just take them one at a time? Any recommendations?

DR. NISHIMI: During the workgroup's review, you ended up wanting to break out a little over half of them. That's why we went to, at that point, measure by measure, but it really would be up to the workgroup whether you want to consolidate or not -- not the workgroup, the committee; I'm sorry.

DR. BURSTIN: Robyn, in the past, we've allowed committees to just agree to review them all and then pull out by exception the ones you want to discuss.

CO-CHAIR BROOKEY: Right, as we did

last time, we can do them en bloc. Remember, this is a must pass. If we want to pull out any of them for individual consideration, we can. I'm going to recommend we do them en bloc. I'm going to ask if anybody wants to pull any of them out, if anybody has a preference to pull it out, we'll approach them individually.

CO-CHAIR SUSMAN: I have sort of a opposite thought here. Because I think the first one around has care coordinator is most broad and really the most evidence around it, probably the most reliability, though I haven't read extensively the detailed notes. I'm wondering if that one should be voted first, and maybe the rest of them en bloc, which are much more particular.

CO-CHAIR BROOKEY: I'm okay with that.

(Simultaneous speaking.)

CO-CHAIR BROOKEY: Jim.

CO-CHAIR SUSMAN: I'm not wed to that.

MEMBER BOST: FECC-3 and 15 have no reliability information presented, but as we

worked yesterday, if they provide valid data at the element level, which they do provide, we could use that as a substitute. If I was going to block, I would use those two because we have to evaluate them a little bit differently than the rest, where they do have either Spearman-Brown or internal consistency reliability estimates.

CO-CHAIR BROOKEY: Okay, any other requests? We'll just honor the individual committee member requests to pull them out individually. Let's go ahead and vote on one, and then we'll vote on the others en bloc, except for 3 and 15, is that correct? 1 is Marlene and Jim. Any comments about reliability for No. 1?

MEMBER MILLER: I have a question for the developers. When I look at this data, it looks like they had about 1,200 surveys, and then 900 or so were picked for reliability. That seems odd because if there's a giant sample, I can respect taking 20 or 30 percent sample, but if it takes 70 percent and not do all, but then

the back tables talk about the reliability, on Pages 84 and 85. Then I see Ns of 30, 50 and 100. I'm very confused. Overall, it struck me as a very small sample size, and why wasn't everything included in reliability testing?

DR. LION: Great question. Because of some limitations related to the IRB stipulations with Washington State Medicaid, specifically -- we had two states, Washington and Minnesota. With our Washington State data, we actually didn't have practice-level information for some of the participants.

While state is the level of intended aggregation, for the purpose of reliability testing, we looked at practice grouping, in order to have -- because our field test only had two states in it. Unfortunately, some of the sample from Washington State was not actually identifiable at the practice grouping level.

We did compare the characteristics of the people who were included in the interclass correlation coefficient calculations to the

overall participant -- the overall respondent group in Table T-1, which I don't think my page numbers match up with anyone else's, and overall, the practice groupings look -- the demographic characteristics of the groups looked very similar.

In addition, we compared -- we statistically compared the scores on the overall respondent sample to the respondent sample that we were able to use for the reliability testing. There was actually only one score that different significantly, and that was not actually one of the measures that we submitted for endorsement here today.

So we did -- although it is clearly a limitation, we felt comfortable moving forward with the analysis using the group for which we have practice information. Regarding your second question related to the sample sizes that we list in the interclass correlation coefficient table, those are actually the practice group numbers, and then the number of patients.

So the first column, with the 92, 48, 59, is the number of practice groups, and then the patients included in those. There were a couple of practices for which we only had less than five patients, so those, we weren't able to use in the interclass correlation coefficients either. Does that explain adequately for you the difference in the sample sizes?

MEMBER MILLER: Sort of, although now I'm more confused why, in Table T-3, the number of patients varies dramatically for reliability testing between each and every single question.

MEMBER BOST: I think I can answer that. The Spearman-Brown formula basically takes the interclass correlation coefficient and asks you if this was your sample size, what would your alpha be? The N equals 30, 50, 100 and 300 were Ns that they chose to allow you to see what the Spearman-Brown formula would then look like.

MEMBER MILLER: No, I'm thinking the N is in two columns or so prior to that.

MEMBER BOST: The 626 patients?

MEMBER MILLER: Yes, but when you go down -- this is, again, a global question against all of them. You're right. For FECC-1, it's 626. Why for all the rest of them were such variations in sample sizes of the patients when it's one survey?

DR. LION: The eligibility for individual items varied to some degree based on the responses to previous questions. For FECC-3, 5, 7 and 8, for instance, you had to endorse that you had a care coordinator in order to be eligible for those questions regarding specific functions of what your care coordinator did for you. In addition, because we chose not to include people for whom we had incomplete information -- and we address this later in the missingness section -- because we didn't feel comfortable imputing responses, we only scored measures for which we had complete information.

So for some of the measures, especially ones that actually had multi-item components, we have lower sample sizes because if

somebody didn't answer any of those, we didn't want to unfairly -- for the measures for which we rolled up the scores, we didn't want, in some cases, for each item to have a weight of 33 percent, but then if someone only answered two of them, for those to be weighted higher for those people. So we only used people for whom we had complete data for each individual item.

CO-CHAIR BROOKEY: Any other comments about reliability for Measure 1?

MEMBER BOST: Basically, the values are -- do demonstrate good reliability as presented for FECC-1.

(Simultaneous speaking.)

CO-CHAIR BROOKEY: All right, thank you very much. Are we ready to vote, just for FECC-1?

MS. CHAVEZ: Okay, we're now getting ready to vote on reliability for FECC-1. We have 27 votes, but one recusal, so we're expecting 26 votes, two over the phone. The options are 1 high, 2 moderate, 3 low, 4 insufficient.

Voting's open, 8, 24, two more, 24.

PARTICIPANT: Everybody vote.

MS. CHAVEZ: 25.

MEMBER WHITE: This is Keith White. I'm going to defer the vote on this set because I haven't been involved in the conversation. I'm not familiar with the measures.

MS. CHAVEZ: Thank you, Dr. White. That would give us 25 votes, so 5 voted high, 20 moderate, 0 for low, 0 insufficient. FECC-1 passes reliability.

CO-CHAIR BROOKEY: Okay, so the next set of measures will be the rest, except for 3 and 15. Marlene, do you want to start off with any comments about those?

DR. NISHIMI: I'm sorry. I just want to clarify for the committee because 14 and 17 have already failed, they are not included in your vote at all.

CO-CHAIR BROOKEY: Thank you, Robyn. Would you please out the ones we're voting on, then?

DR. NISHIMI: We're not voting on 3, 5, 14, and 17. You are voting on 7, which is --

PARTICIPANT: Care coordinator assisted with specialist service referrals.

DR. NISHIMI: Right, 8, which is care coordinator was knowledgeable, supportive, and advocated for the child's needs, 9, which is appropriate written visit summary content, 15, which is related to the medical interpreter, and 16, which is shared care plan. Those are the ones.

(Simultaneous speaking.)

MEMBER BOST: Did you yes, we're doing 5, or no? I couldn't --

DR. NISHIMI: No, we're not doing 3 and 5 because -- I'm sorry, 3 and 15. So you're not voting yet on access to medical interpreter.

CO-CHAIR BROOKEY: So we're voting on 7, 8, 9, 15, and 16.

MEMBER BOST: And 5.

DR. NISHIMI: And 5, not 15.

CO-CHAIR BROOKEY: Not 15? We're

voting on 5, 7, 8, 9 and 16? Okay, 5, 7, 8, 9 and 16.

CO-CHAIR BROOKEY: Any comments about these measures, in terms of reliability, from those who reviewed them closely? Jim?

MEMBER BOST: So the reliability estimates provided for 5, 8, and 9, because they are multiple items, they did the within item set alphas, which were actually very high, ranging between .73 and .86. The rest of them, again, were the same Spearman-Brown formula associated with the interclass correlation coefficient. Each one of those were fairly good.

Again, potentially the least good, if you will, was FECC-9, but was still at .7, when you had an N of 100. It was a little confusing why some of these items did not have the ICC, like Item 5, which seemed to have enough of a sample size, but they did provide the within item alpha, which was good. All of the data provided on reliability, where it was provided for these items, was good.

CO-CHAIR BROOKEY: Any other comments about reliability of these measures? Okay, should we vote for 5, 7, 8, 9, and 16?

MS. CHAVEZ: Okay, reliability for FECC-5, 7, 8, 9, 16, 1 high, 2 moderate, 3 low, 4 insufficient. Voting's open -- 23, we're expecting 25 votes, 24, 25 -- 2 voted high, 23 voted moderate, 0 for low, 0 insufficient. So FECC-5, 7, 8, 9p, 16 pass reliability.

CO-CHAIR BROOKEY: Thank you. Moving on to validity, I'll ask the same question. Jim, Marlene, Kerri, others who reviewed it, are there those that we would vote en bloc, or are those that we should call out individually? Any objection to -- go ahead, Jim.

MEMBER BOST: I think all of them except 15 showed good validity, so I would, again, probably consider 15 separate.

CO-CHAIR BROOKEY: Any disagreement with that? So we'll be voting for all except for 15, but we also will not be voting on 14 and 17. Let me just see if I can get this straight,

voting on --

MEMBER BOST: And we have to do the validity or reliability.

CO-CHAIR BROOKEY: -- 1, 3, 5, 7, 8, 9 and 16.

DR. NISHIMI: We need to pull 3 out separately, too, because you would want -- if you're going to, you would want to use your validity vote for the reliability vote.

CO-CHAIR BROOKEY: Okay, so we'll just do it the way we did last time, then. We're voting on -- but we're including 1 in the vote. We're including 1, 5, 7, 8, 9 and 16, is that correct? Jim is supporting the validity of these measures. Any other conversation about validity? Okay, we'll vote.

MS. CHAVEZ: Okay, 1 high, 2 moderate, 3 low, 4 insufficient. Voting's now open, 23.

PARTICIPANT: One more.

MS. CHAVEZ: 24, 25.

CO-CHAIR BROOKEY: It sounds like we need to vote on 3 and 15 separately. Let's vote

on 3, which Jim had recommended be part of the other block, but we need to pull it out because we pulled it out for reliability. So we're voting only on 3. Any other comments about 3?

DR. NISHIMI: Before we start, Jim -- I'm sorry to cut you off -- Severa needs to read the results out for the transcript.

CO-CHAIR BROOKEY: I'm sorry, go ahead.

MS. CHAVEZ: Yes. For the record, validity on FECC-1, 5, 7, 8, 9, 16, 2 voted high, 21 voted moderate, 1 low, 1 insufficient.

CO-CHAIR BROOKEY: Thank you. Any other comments about validity for Measure 3?

MEMBER BOST: So the item level or element level assessment that we're being asked to review for 3 was done on the denominator, where they looked at 700 children using the algorithm associated with the measure and compared whether the denominators were the same using clinical chart review as the gold standard. The results were very good, in terms of

sensitivity, specificity, at both Seattle Children's and Washington Medicaid.

CO-CHAIR BROOKEY: Any other comments about validity for Measure 3? Should we vote?

MEMBER HOUTROW: I have a question.

CO-CHAIR BROOKEY: Amy, go ahead.

MEMBER HOUTROW: For 3, it says it did not achieve good reliability at the performance score level, is that right, Jim?

MEMBER BOST: I'm sorry, I can't hear the question.

MEMBER HOUTROW: On Page 14, Item 3 did not achieve good reliability at the performance level, which the developer attributes to small sample size.

MEMBER BOST: Right, that was the fact that they had no reliability data and so were using this element level one instead.

CO-CHAIR BROOKEY: Any objections to voting? We're now voting on FECC-3 for validity.

MS. CHAVEZ: Yes, voting on FECC-3 for validity, 1 high, 2 moderate, 3 low, 4

insufficient. Voting's open, 15, 24, 25. Okay, 1 voted high, 22 voted moderate, 2 voted low, 0 insufficient. FECC-3 passes validity.

CO-CHAIR BROOKEY: So moving forward -- actually no, we need to vote on one more. We need to vote on 15 for validity. Comments about 15?

MEMBER AGORATUS: This is Lauren. There were some comments on validity, first in specifications, the not accepting the sample size as a reason for lack of reliability, not reporting results, and then under validity testing, again, small sample size. It should be looked at separately.

Another comment, validity assessment was good, unfair to deny high score due to lack of convergent validity, difficulty with validity testing due to small sample size, although there was face validity, and finally, test sample was inadequate. The result did not show validity. That's it.

CO-CHAIR BROOKEY: Okay, Jim go ahead.

MEMBER BOST: That's exactly right. The validity associated -- the convergent validity was to look at an adult CAHPS measure comparison, and for all the rest of the items, the validity showed acceptable levels of validity, except for FECC-15, which only showed acceptable validity, I believe, in one of the assessments. The thing to think about with FECC-15 is that this is the did you have an interpreter question.

I'm not surprised it had lower validity because as we saw in earlier evidence assessments, interpreters are used for everything -- or every patient in need besides those with complex kids. I think you have to decide whether we should have expected them to do some other kind of validity assessment for this question or give it a pass because it wasn't as appropriate for the items they chose for validity.

CO-CHAIR BROOKEY: But there was high face validity.

MEMBER BOST: I'm sorry?

CO-CHAIR BROOKEY: There was high face validity.

MEMBER BOST: Yes, their face validity was good, the convergent validity not so good.

CO-CHAIR BROOKEY: So we could still pass it based on face validity, but we can't give it higher than a moderate.

PARTICIPANT: A moderate.

CO-CHAIR BROOKEY: Any other comments before we vote? Okay, we're voting on FECC-15 for validity.

MS. CHAVEZ: 1 high, 2 moderate, 3 low, 4 insufficient. Voting's open, 20, 24, 25, 0 voted high, 18 voted moderate, 5 voted low, 2 voted insufficient. FECC-15 passes validity.

CO-CHAIR BROOKEY: Thank you. We're moving on to feasibility. I suppose unless we're sending out ten different surveys, which I hope we don't do, that it should be the same for all measures. I would recommend that we vote en bloc. Any opposition to that? Any comments about feasibility? It's a survey. It's probably

feasible. This is at a plan level or practice level? I forget now. Plan level, it's at a plan level survey, so not too challenging. Any comments before we vote? Hearing none, we'll vote for --

MS. CHAVEZ: Okay, voting on feasibility for the whole FECC measure set, except for FECC-14, correct?

CO-CHAIR BROOKEY: Actually --

PARTICIPANT: And 17.

MS. CHAVEZ: And 17.

CO-CHAIR BROOKEY: Correct, 14 and 17 will not move forward.

MS. CHAVEZ: Okay, voting on feasibility, 1 high, 2 moderate, 3 low, 4 insufficient. Voting's open, 24, 25, 1 voted high, 21 voted moderate, 3 voted low, 0 for insufficient, so this passes feasibility.

CO-CHAIR BROOKEY: Thank you. We move on to usability and use. Again, I would propose that we vote en bloc, unless there's any recommendations to vote for a measure

individually. Jon?

MEMBER FINKELSTEIN: I just have one comment on one of the measures that's giving me pause. If I'm the -- it's discharge instructions visit by visit. What's giving me pause is not so much the usability of the measure, but we're in this meaningful use period, where the requirements for discharge summaries have entrained a whole bunch of non-productive gaming and EHR things, so people get out the right discharge instruction with the right elements. I'm just a little concerned about doubling down on that when I think we don't really know what we're doing, as a field. I don't know if this is exactly the right place to say that, but that one is sticking out for me.

CO-CHAIR BROOKEY: Do you want to provide that as input, or do you want to pull that out for a separate vote?

MEMBER HOUTROW: Aren't we separate voting all of them at the end?

CO-CHAIR BROOKEY: We can. You want

to do it that way?

MEMBER FINKELSTEIN: That might make sense because -- yes.

CO-CHAIR BROOKEY: I'm okay with that. Any objections to that? We will vote on usability en bloc, but we'll have the opportunity to do overall by measure for the two that didn't move forward. Any other discussion? Otherwise, we'll move to vote.

MS. CHAVEZ: Okay, now voting on FECC-1, 3, 5, 7, 8, 9, 15, 16, usability and use. 1 high, 2 moderate, 3 low, 4 insufficient. Voting's open.

MEMBER MILLER: Could I ask a question, though, thinking about use and usability? I was wondering if the developers -- because this is meant to be how would purchasers and developers can comment to me -- what I see on Page 102 is only three measures at a large state level showed any differences, but I know there was subsets and plans. Did any of these measures show any differences? I think that affects

usability, that they can identify gaps. I'm just trying to understand that only three of them showed any difference.

DR. LION: We were limited by just having a two state sample at this point. In our field testing, we did four of the -- although, I guess one of them was FECC-17, which is now out -- so three of the remaining measures did show differences at the state level. We also saw differences by English proficiency and race/ethnicity at -- in one, two, three, four, five, six of the measures, which were not necessarily the same ones that we saw differences by state in, and we expect that with more states and more points of comparison, we would likely see more differences.

MEMBER MILLER: I'm just curious, did you try looking at the health plan level, because then you had hundreds of health plans, or you didn't do that analysis?

DR. LION: We had Washington State Medicaid, and we had Minnesota Medicaid, and that

was it.

CO-CHAIR BROOKEY: Any other comments or questions? We'll go ahead and continue the vote.

MS. CHAVEZ: Okay, voting on usability and use for FECC-1, 3, 5, 7, 8, 9, 15, 16, 1 high, 2 moderate, 3 low, 4 insufficient. Voting's open, 24, 25, 2 voted high, 18 voted moderate, 5 voted low, 0 insufficient. This group passes usability and use.

CO-CHAIR BROOKEY: Okay, so as recommended, we're going to vote for overall, for all of the measures except 14 and 17. Remember this is a must pass, so this is your opportunity, if you have reservations about one particular measure, that you can vote yes or no. I'm going to just go measure by measure, ask for comments, and then we'll vote. FECC-1, has care coordinator, any comments before voting? Marlene?

MEMBER MILLER: Just a curiosity for the developers I want to think about. It seems

to me if you grouped 1 and 8, where 8 says, "Was the care coordinator knowledgeable and supportive," you get at what you really want than just the person being assigned to you.

CO-CHAIR BROOKEY: That came up yesterday. We talked about not just having a care coordinator, but having the care coordinator actually do something for you, so comments for the developer.

DR. MANGIONE-SMITH: Right. Performance in FECC-1 was only 44 percent in the two states among eligibles, so we felt pulling it out as its own measure was important. We didn't only want -- if you combine it with 8, then you can only ask that measure of people who have a care coordinator. That's why they're separate. I know you want to double whammy them. You mentioned that yesterday.

MEMBER FINKELSTEIN: It's just waiting to come up -- is to use, in a sense, the denominator for 1 with the numerator for 8. That's just input. I understand that's not the

measure.

CO-CHAIR BROOKEY: Right, we talked about that yesterday. Any other discussion about overall for 1? Okay, we'll vote for 1 overall.

MS. CHAVEZ: Now voting on FECC-1, overall suitability for endorsement, 1 for yes, 2 for no. Voting's open, 24, 25, 23 voted yes, 2 voted no, so FECC-1 has been recommended for endorsement.

CO-CHAIR BROOKEY: Thank you. We're going to FECC-3, care coordinator helped to obtain community services overall. Any comments? Hearing none, we'll go to a vote.

MS. CHAVEZ: Okay, voting on FECC-3, overall suitability for endorsement, 1 yes, 2 no. Voting's open, 25, 22 voted yes, 3 voted no. FECC-3 has been recommended for endorsement.

CO-CHAIR BROOKEY: Thank you. We're going to FECC-5, care coordinator asked about concerns in health changes overall. Any comments? Hearing none, we'll go to vote.

MS. CHAVEZ: Voting on FECC-5, overall

suitability for endorsement, 1 yes, 2 no. Voting's open, 23, 24, 25, 21 voted yes, 4 voted no. FECC-5 has been recommended for endorsement.

CO-CHAIR BROOKEY: Okay, moving on to FECC-7, care coordinator assisted with special in service referrals. Any comments about overall? Okay, going to vote.

MS. CHAVEZ: Voting on FECC-7, 1 yes, 2 no. Voting is open, 24, 25, 19 voted yes, 6 voted no, FECC-7 has been recommended for endorsement.

CO-CHAIR BROOKEY: Somebody yesterday said that FECC-8 is what we really want, which is a care coordinator was knowledgeable, supportive, and advocated for child's needs, so any comments about this measure? Okay, we'll go to vote.

MS. CHAVEZ: Okay, voting on FECC-8, overall suitability for endorsement, 1 yes, 2 no. Voting's open, 24, 25, 24 voted yes, 1 voted no. FECC-8 has been recommended for endorsement.

CO-CHAIR BROOKEY: Thank you. FECC-9, appropriate written visit summary content. Amy,

do you want to make a comment about this?

MEMBER HOUTROW: I have similar concerns that John had about the expectation of this measure, and also how it doesn't relate so much to care coordination, but what an actual physician does in the moment of providing care, so conceptually, it seems a little bit different than the rest of the ones that we just voted on. In particular, upping the expectation past meaningful use, I think, is going to be hard on people, so I do have a couple of concerns about this particular one.

CO-CHAIR BROOKEY: Rita.

DR. MANGIONE-SMITH: The only thing I'd like to point out, in our validation analyses, where we looked at the association between increases in scores on that measure and increases on four different CAHPS measures, that one had the strongest validation results.

Just to give you a sense, there was one validation metric that we looked at from health plan CAHPS that asked about did you get

needed help coordinating your child's care, or did you get all the help you felt you needed. That's on a 0 to 100 scale, as is FECC-9. If you were go from failing that measure being a 0 to passing it being 100, you would move 36 points up on the 0 to 100 scale on the CAHPS measure. That was literally one of our strongest results in validation. So although I hear what you're saying about the meaningful use thing, it does seem to matter, in terms of people feeling like they got needed care coordination. That may have to do with the fact that we pushed the envelope a little bit on the NCQA criteria.

We were pushed by families, actually, in our center to add the criteria that you would put information about who to contact if problems came up related to the visit. That extra content you guys noted yesterday was actually very driven by the families, so I would just ask you to keep that in mind as you think about this.

CO-CHAIR BROOKEY: Any other comments? Okay, we're going to vote on FECC-9.

MS. CHAVEZ: Now voting on FECC-9, overall suitability for endorsement, 1 yes, 2 no. Voting's open, 21, 24, 25, 18 voted yes, 7 voted no. FECC-9 has been recommended for endorsement.

CO-CHAIR BROOKEY: Thank you. We're going to go to FECC-15, caregiver has access to medical interpreter when needed. Any comments? Okay, let's go to the vote.

MS. CHAVEZ: Voting on FECC-15, overall suitability for endorsement, 1 yes, 2 no. Voting's open, 23, 25, 22 voted yes, 3 voted no, FECC-15 has been recommended for endorsement.

CO-CHAIR BROOKEY: We're on the last measure. It's FECC-16, child has shared care plan. Any comments? Okay, move to vote.

MS. CHAVEZ: Okay, voting on FECC-16, overall suitability for endorsement, 1 yes, 2 no. Voting's open, 22, 25, 22 voted yes, 3 voted no. FECC-16 has been recommended for endorsement.

CO-CHAIR BROOKEY: Great. We can all take a sigh of relief now. We've gotten through the FECC measures. Thank you to our developers.

This has been a great discussion. I know that you'll take all these comments to heart, very good discussion about the measurement set. Jim, you want to --

MEMBER BOST: I just wanted to make one closing comment, and that's about the fact that we're doing these separately, as opposed to in a group, for the developers to think about how they would approach benchmarking for these items if a health plan gets to choose which ones. Perhaps that's fine for internal benchmarking, but if they ever want to develop more national benchmarking, you might want to consider that you need to do them all in order to get your unbiased benchmarks.

CO-CHAIR BROOKEY: Great feedback. Kerri.

MEMBER FEI: One other closing comment, as well, along those lines, that regardless of what the measures are that come out of the survey, keeping the survey together, as a whole, and not just picking out the ones that

have measures associated with them.

CO-CHAIR BROOKEY: Agreed.

MEMBER FEI: I think that's important. Otherwise, you're going to skew how people respond.

CO-CHAIR BROOKEY: Agreed, thank you. All right, thank you very much. Oh, one more comment from Jeff.

MEMBER SCHIFF: I just want to comment, now that voting's done, just to say that -- just a couple things. One is I wanted to compliment Rita and her staff for really keeping this together in a good process. Some of the interesting comments about Minnesota and Washington are a lot alike, which is why there's not a lot of variability, but I think that's a challenge. The other thing I just really wanted to point is from my point of view, in Medicaid -- I'm sure a lot of you who work with medical homes and care coordination -- the whole movement to care coordination happening at the practice level in a measurable, identifiable way, this body of

measures moves that.

So as we think about this in a health plan level, it helps the health plans be accountable for something that's much more meaningful going forward. I've just kind of been sitting here all day. I just wanted to throw that in.

CO-CHAIR BROOKEY: We want to make sure we capture your comments, since you didn't vote on anything, so thank you very much. Thank you so much. We are going to now go to public comment. You want to take --

DR. NISHIMI: Operator, can you announce that it's public comment and see if there's anyone who wishes to do so?

OPERATOR: At this time, if you would like to make a public comment, please press star, then the No. 1 on your telephone keypad. We have no public comments at this time.

DR. NISHIMI: Is there anyone in the audience who wishes to make public comment? Okay.

CO-CHAIR BROOKEY: Thank you so much. I'm going to pass the baton to Jeff now to pick up on 2820, so take it away.

CO-CHAIR SUSMAN: All right, John. Go get yourself a cup of coffee, a slug of whiskey. You're off the hook for a bit. We're going to turn to 2820, the pediatric CT radiation dose measure, No. 2820. I believe we have developer representatives dialing in. Are they there?

MS. KUMAR: Yes, Karishma Kumar from UCSF is calling in.

DR. SMITH-BINDMAN: Hi, and this is Rebecca Smith-Bindman from UCSF.

CO-CHAIR SUSMAN: Okay, and we have one recusal in James Duncan. Any others who recuse themselves? Okay, so for our measure developer, I wonder if you could give us three to five minutes, just brief overview of the measure and the evidence and rationale gap.

DR. SMITH-BINDMAN: Yes, of course, thank you. Good morning. The purpose of this measure is to encourage facilities to assess the

radiation doses they use for CT, computed tomography, in children, and to standardize and lower these doses if they learn their doses are higher than benchmarks.

Just a little background, most medical imaging tests use ionizing radiation, and the doses that are used for CT, which has become the workhorse in radiology, are far higher than conventional X-rays, many orders of magnitude higher.

The doses that are typically used for CT are in the range that they will cause cancer in a small, but significant number of children. These risks are particularly elevated in children compared with adults. Thus, it's important to keep these doses as low as possible. Unfortunately, the doses that we typically use for CT in process are far higher than needed for diagnosis and are highly variable across institutions. As an example, it's well known that you should use a low dose -- 2, 3, 4 millisievert -- CT to diagnose kidney stones.

But we just completed a 15-center randomized controlled trial comparing ultrasound versus CT for kidney stones that we published in the New England Journal last year, and we found that fewer than 7 percent of patients actually received low dose examinations.

The average dose was three times higher, and the doses ranged up to 75 millisieverts for a single study, even though the patient should have gotten a dose of 2. So the variability is profound, and we've published extensively on this. The work has shown, both ours and others, that children have doses that are even more highly variable. While higher doses can lead to more detailed images, there's no evidence at all that these lead to more accurate diagnosis.

One of the primary reasons for the variability is that institutions tend not to know to dose that they routinely use. There's no one who tends to look at, assess, or track their dosage. This has changed dramatically in

California over the last two years, where state law now requires reporting of dose in the medical record, and this has really increased this awareness across the state. This is also going to change across the country in hospitals that are covered by the Joint Commission because they're beginning to require institutions to begin assessing their typical doses.

However, this won't cover outpatient facilities where most children are imaged. What this measure does is provide a very simple framework to allow facilities to assemble the doses they use for CT and compare these doses to typical benchmarks that are published. It calls for facilities to look at their average doses, as well as their high doses, and both of those will contribute to the total dose that their patients receive.

It asks them to assemble their doses, and this can be done in various ways, both automatic and using manual techniques, and we have published comparing effort required using

automatic versus manual techniques, and it's relatively modest, no matter what the approach. Lastly, there have been several papers and abstracts, including observational studies, as well as a randomized control trial, that have shown that assessment of doses, as described in this measure, results in meaningful lowering of doses, both average doses and typical doses. The average institutional doses can be reduced 25 percent or more, and the outlier high doses above the benchmark, in a paper we recently completed, went down between 25 percent and 80 percent at each individual institution after reviewing their doses, as described in this measure.

I appreciate you considering this measure. This measure was previously endorsed by the NQF. It lost its endorsement because there was concern that simply assessment of doses is not enough to change the doses. I've resubmitted the measure because now we have several publications, and several more in press, that show, in fact, looking at the doses using this

very simple framework does result in measurable reduction. Thank you very much for considering it.

CO-CHAIR SUSMAN: Thank you very much for that succinct summary. Let me ask for clarification up front about your intent. There are essentially two parts to the measure, or perhaps two measures, depending on how one is thinking about this. The first is the mean dose, and the second is the proportion of CT exams with doses greater than the 75th percentile. Do you mean this to be a composite, two separate measures, or reporting just the latter or the former? What is the intent?

DR. SMITH-BINDMAN: The intent was for a facility to comply with this measure, they have to collect data on their CT scans. They collect data on 100 CT scans and they describe the distribution. The benchmarks tend to be at two points in that distribution.

The benchmarks tend to be published at the 50th percentile, the average, and the 75th

percentile. So it's really just one measure. Facilities can improve by comparing either or both. They're not separate measures. They're different points on a single distribution. So if you know the 50th percentile, you also know every other percentile, including the 75th.

It's just two points of comparison, and those points are assembled because that's where benchmarks are. As another example, you could look at the 95th percentile or the 99th percentile. There just tend not to be as many benchmarks at those locations that you can compare with.

CO-CHAIR SUSMAN: Maybe I'm not clear, but let me try it again. In your reporting -- let's say I'm out there at a health plan or integrated delivery system. What would I be reporting, the percentage that are above the 75th percentile, or both those measures, or something else?

DR. SMITH-BINDMAN: You could report both of those measures or either of those

measures.

CO-CHAIR SUSMAN: Okay. Joe.

MEMBER BOST: When you're talking about 75th percentile of a benchmark, are you just using the data collected at that time period for your benchmark, or are you using all the historical data from multiple time points, or are you using national benchmarks from everybody that does this measure?

DR. SMITH-BINDMAN: The benchmarks are national benchmarks. The national benchmarks, for example, these are within age strata -- we'll say that the 50th percentile of a child's abdominal CT is 5 millisieverts, and you'll compare your 50th percentile to that benchmark, or you'll look at the 75th percentile benchmark created nationally, and say the 75th percentile is -- by definition, you should have approximately 25 percent of exams above the 75th percentile benchmark if you're doing an average.

If the 75th percentile for children nationally is 8 millisieverts, and in your data,

the 75th percentile you have 40 percent above that benchmark, you're doing less good than average. If you have 10 percent above that benchmark, you have been doing better than average.

MEMBER BOST: Do these benchmarks already exist because you've been collecting this historically, or will they be developed going forward, based on who participates?

DR. SMITH-BINDMAN: There are several existing publications on benchmarks in children, including our own and internationally, but I anticipate that as our group is publishing more benchmarks, other groups will also publish benchmarks. These benchmarks will, therefore, be updated on a regular basis. For example, in 2015, hospitals might compare their data to benchmarks published in 2014, and in 2016, you'd compare your benchmarks to those of the year prior.

CO-CHAIR SUSMAN: Okay, I'm going to get Marlene, and then we'll turn to the

subcommittee that looked at this.

MEMBER MILLER: I'm wondering can you just clarify for me again the denominator? What I see is that it just says all these consecutive CT scans, but can you clarify for me again three specific populations of whether they're included or not, the inpatient setting, the emergency department setting, and then all other ambulatory? I'm just confused from the document of what is exactly in the denominator pool.

DR. SMITH-BINDMAN: The denominator is every CT scan done within the organization that chooses to participate in the measure. So if the organization is an outpatient facility, it would be every CT scan done at that facility. I highlight the executive exams because the measure doesn't call for looking at doses in typical patients or subsets of patients. It calls for looking at doses in all patients. An outpatient facility would look at doses in every head CT, every chest CT, every abdominal/pelvic CT. If the institution participating is a health plan

that has both inpatient and outpatient and hospital settings, it would assemble all of those doses together and do the same with the consecutive doses in every head, every chest, every abdomen and pelvic CT. We have assessed doses in a range of institutional size from very tiny to very large, integrated healthcare systems, with 3 million plus members, and have shown --

CO-CHAIR SUSMAN: I'm sorry to break in, but because we're really constrained on time, I'm going to want to move us forward. I think you nicely answered the question. I'd like to turn to Keith and the members of the committee who reviewed this, and then I will get your comment. Yes? Good, okay, so please.

MEMBER FATTORI: I think the first question that our workgroup wanted to clarify was whether or not this should be evaluated as a process measure or an outcome measure, so maybe that's something that we can discuss at the beginning, before we launch into the rest of the

conversation.

CO-CHAIR SUSMAN: Sounds good. I invite the developer to discuss -- I think this was presented as an intermediate clinical outcome, and whether we determine or think of it as such is going to influence which path -- the infamous vertical or horizontal path we use thereafter.

DR. SMITH-BINDMAN: I think we -- I believe it's an intermediate outcome measure. In order to assess the outcome of cancer, we'd have to wait a very long time in these cohorts -- 5, 10, 20 years. There's a very strong relationship between the radiation dose and cancer. Patients and their parents care a great deal about the radiation by itself as an outcome. So I believe it's an intermediate outcome measure.

CO-CHAIR SUSMAN: Keith and others on the subcommittee like to weigh in on that particular issue?

MEMBER WHITE: I would agree with that assessment. I think that if you believe the

science that links the administration of radiation to the belt and cancer, even though, for an individual patient, you can't actually show a direct outcome or a direct impact on cancer, on the population basis, you can infer a direct outcome from amount of radiation given, so I would agree that this would be an intermediate outcome measure.

CO-CHAIR SUSMAN: If I'm following correctly, if this is accepted as an intermediate outcome measure, it would take us more on that horizontal path of pass/no pass. Virginia?

MEMBER MOYER: As they were beginning to talk about this, to me this is almost exactly the same thing as immunization rates, which are an intermediate clinical outcome.

CO-CHAIR SUSMAN: Okay, Keith or other of the subgroup members want to comment particularly, now, with regard to the specifications and evidence and whether this is an intermediate clinical outcome, as proposed?

MEMBER WHITE: Just on the

specifications, I don't know if this is the right place to talk about this, but I'd like to maybe ask a couple of questions about using the site specific measure. The specifications were that there would be measurements made of CT and BLT and also the size specific. The size specific, as was stated, I think, in the workup, it's going to require either a lot more manual intervention, or it's going to require systems that want to employ this measure to have computer programs to automatically do that modeling for them.

Could the developers comment on how strongly they feel that the size-specific dosing should be included, and if so, their belief as to how that could best be operationalized?

DR. SMITH-BINDMAN: I think there's a lot of attractiveness to the size-specific dose estimate because it allows numbers to take into account how large the patient is, a tiny infant or a very large teenager. Once you've established the age category, however, it becomes far less important.

Quite honestly, in all of our work, we have found no difference, whatsoever, in identifying outliers, whether you use size-specific dose estimates, or whether you use the other metrics that it outlines. You get exactly the same result. So there's a theoretical attraction --

CO-CHAIR SUSMAN: Hello?

MEMBER WHITE: Hello?

CO-CHAIR SUSMAN: Our link to the sky.

MEMBER AGORATUS: Did Rebecca cut out?

CO-CHAIR SUSMAN: Yes.

DR. SMITH-BINDMAN: Can you hear me?

CO-CHAIR SUSMAN: Yes, we can hear you. We can hear Keith.

(Simultaneous speaking.)

DR. SMITH-BINDMAN: I thought I cut out. I got back. Okay, so if you use these data automatically -- if you collect them automatically, it's very easy to do SSDEs. If you have to do them manually, it's an enormous amount of work and not practical at all. It

requires calculations. That's just not feasible for most people to do. It's not impossible.

So the bottom line is I don't think it adds that much, but I think it will enhance radiologists and medical physicists' adoption of this measure because they like it very much. I think it is equivalent. It takes extra work if you have to do it manually, but if you do it automatically, it's the same. In terms of how well the measure will work, though, it doesn't change whether you're using the SSDE or the other metrics that I outlined.

MS. KUMAR: Don't most modern CT machines calculate it automatically?

DR. SMITH-BINDMAN: That's a great question. It calculates all the other measures that I specified, and that's why the other ones are easier to get. You get direct output of CTDIvol or DLP from the machine SSDE is not calculated automatically. It's basically the CTDIvol divided by the abdominal circumference, but you have to measure the abdominal

circumference. If you're doing it automatically, it's a piece of cake. If you're doing it by hand, it will take time.

CO-CHAIR SUSMAN: David, do you have a question?

MEMBER EINZIG: It's just a general question. Is there any risk for other clinical times where a higher dose of radiation is warranted, or is there a risk that something will be missed? Are there times where something could be missed because of using a lower dose?

DR. SMITH-BINDMAN: This is looking at institutional averages. If there are some patients who get higher doses and some patients who get lower doses, that will average at the institutional level. This measure does not require every patient to be below the benchmark. It requires the averages.

There's a lot of variability in dose depending on clinical indications, but that variability is dwarfed compared to the variability based on institutional preference. A

concrete example is that some facilities use single-phase settings going for an area once for a clinical question, others use multiple phase.

If you use multiple phase, your dose is about twice as high. So in general, if an institution uses multiple phase on all their patients, their doses will be twice as high. It's not to say that if you usually use single phase, you can't use double phase in some patients for some clinical institutions.

CO-CHAIR SUSMAN: Okay, going up the line, Martha.

MEMBER BERGREN: I didn't understand that last explanation, so I'm going to ask the question that I need to understand. Would the requirement or the motivation to the lower the doses lead to an unintended consequence of having to repeat scans?

DR. SMITH-BINDMAN: If an institution has doses that are, in general, very, very high, and to lower those doses they make a concerted effort to bring the doses down in all patients,

there could be examples where for an individual patient, those images are no longer diagnostic because the dose is too low, and they'll have to repeat an individual case.

I don't think that is an ongoing problem because if they realize they have to start repeating cases, they will raise the doses up. The strategy that most institutions use for getting the right dose is they lower them as low as they can until they start seeing that they can't make good diagnoses, and then the radiologist will you you've got to bring the dose back up because we're starting to get non-diagnostic scans.

MEMBER BERGREN: Thank you.

DR. SMITH-BINDMAN: That's the typical approach that facilities use to get the right dosage.

MEMBER WHITE: This is Keith. I would concur. That's a very common approach. That's basically the hard reality of dose reduction is that at least at one level, there's a subjective

acceptability of the image noise that radiologists have.

What is clearly true, based on published evidence and also by practical experience of institutions that are trying to manage actively their radiation dose, is that there's a tension in the dynamic between the dose that's administered and the acceptability of the images to the radiologists.

It's clear that by measuring -- for an institution measuring their dose and contrasting it to national benchmarks and coming back and saying we're using high dose, that will force that organization to address that tension, and experience shows that they will decrease radiation exposure to children.

The other thing I would point out is that the data clearly shows, as well, that on the average, the doses that are being administered across the board in the country for children are higher than are really, truthfully needed to make diagnoses. I would echo that there's potential

for unintended consequences from doing this, but I agree that it would be more on an individual level, and that the overall benefit to the population, as a whole, would far exceed any compromise that's being extended for the care of an individual.

MEMBER BERGREN: Thank you.

CO-CHAIR SUSMAN: Okay, there are a number of questions over here. I'm going to start with Jeff, who is one of our subcommittee members.

MEMBER SCHIFF: I just wanted to ask the developer to talk a little bit about -- this states modern CTs. I think the number we got from your information was that's 90 percent. I'm curious if you have any data in the amount of CT scanning that's done by the other 10 percent? I guess what I'm really asking is are we looking for the keys under the streetlight here, where the other 10 percent may be the ones we should be most concerned about with regard to this, and if you have any data on those?

DR. SMITH-BINDMAN: Are you asking how many facilities have CT scanners that don't report these measures?

MEMBER SCHIFF: Or don't have the ability to report them, and what the pediatric population in those, if you have any information on the pediatric population in those sites?

DR. SMITH-BINDMAN: The only number that's available for how many facilities don't have these measures automatically recorded was a survey done in 2005, so ten years ago, by the FDA. That number, at the time, was 95 percent. I believe that number is closer to 99 percent. Really, these numbers -- any new scanner, and any scanner within the last ten years, reports these data.

The number of facilities that don't have the capacity to do this -- there are no data, but it should be in the ballpark of 1 percent, so very, very few facilities. There's no way to really get a handle on it. Truth be told, those facilities could still do this. It

just would require using a free computer software program, where they put in --

(Simultaneous speaking.)

CO-CHAIR SUSMAN: This sounds more like a feasibility issue, and I think you've well described it. Thank you very much. Jon.

MEMBER FINKELSTEIN: Mine is actually just a very quick follow on to that. I just want to make sure, for the record, that places don't need a particular brand of CT, that machines from all the different companies do this the same way, that we won't be pushing people toward something proprietary that there's a conflict here. I doubt it, but I think we should ask that question.

DR. SMITH-BINDMAN: Absolutely not. You can get these data from every vendor's CT scanner, and there are a half dozen dose software products out there. You can get these data from any of those products.

CO-CHAIR SUSMAN: Ricardo.

MEMBER QUINONEZ: My question relates

to access of care if this were to pass. Contrary to popular belief, most children are not taken care of in children's hospitals or children's ERs. Most are seen in community settings and community ERs, where there's no pediatric radiologist. I would imagine that lowering the dose to its most minimal need would make the need for pediatric radiologists reading studies a little bit higher, since they would be more comfortable with the lower-dose results. Did you consider whether doing that would limit access to care for children, since most of them are actually seen in outlying ERs?

DR. SMITH-BINDMAN: I think your question is an excellent one. I think most of the reason the small, non-academic, non-pediatric hospitals use higher doses is they don't tailor the doses they're using for children. This measure would encourage them to tailor their doses to children, but at some point, they may not have the expertise to read those studies.

I would say the way the access issue

can be addressed is these studies can easily and electronically be sent to other facilities, larger academic facilities or pediatric facilities. But it's certainly possible, at a facility that has no trained pediatric radiologist or no radiologist who's experienced with using these doses that they could have difficulty using them.

CO-CHAIR SUSMAN: Okay, thank you.

MEMBER WHITE: Can I make one other comment there on that question?

CO-CHAIR SUSMAN: Yes.

MEMBER WHITE: That is that if you really think about this, if you have a child in an adult facility and that child has a scan done at an adult protocol, so the adult amount of radiation administered, those images on that child are going to be extremely sharp and really very, very high detailed, more high detail than what would be a comparable scan done with that same protocol in an adult.

Those radiologists would have been

used to looking at those more noisy images on the adult, so if they were to tailor and to decrease the radiation exposure on the child, such that the quality of the images was the same as what they were generating on the adult, it would reduce the dose for the child, and they would still be looking at studies that are basically of the same quality as what they're reading daily for the adults.

CO-CHAIR SUSMAN: Marlene.

MEMBER MILLER: I just wanted to make three points. In the State of Maryland, our state considers an amoxicillin radiation dose error just like a medication error. I often train our residents that you have to think of X-ray dose or radiation dose just like amoxicillin dose. If you put that lens on and think about it, we would never want a child going to a community ED and be given an adult amoxicillin. This is one of those rare opportunities we have to move that field to the right thing for kids.

You've heard it from the developers. The scans can do this. We know the right doses. It won't impede the reading, and we would not want children exposed to these unnecessary doses, just like a medication wouldn't be acceptable at a community hospital. Maybe they shouldn't be seeing the children if they can only do adult amoxicillin.

CO-CHAIR SUSMAN: Okay, one final question, at least, that I have is you say that the actual measurement provokes changes in behavior, but from the studies that I've seen that you provided, it looked like it was always associated with some educational or quality improvement process. In much of quality improvement work, just measuring something doesn't always evoke behavioral change. I just wonder if you could comment briefly on that as the developer?

DR. SMITH-BINDMAN: I think facilities are starting, largely, with no knowledge of how they're doing. So giving them a framework, which

is what this measure does, for looking at their dose and comparing it, I think, is a strong motivator.

The intensity of our educational intervention has been very modest, to say the least. It's been basically encouraging people to speak on the phone and share best practices, so we have not made extensive materials available to optimize dose. That being said, as part of two large federal NIH grants that I have, we'll be making a lot of materials available online, for example, optimized pediatric protocols.

Those don't currently exist very often. There are very few protocols out there, basically instructions, so we will provide those in the next few months. They'll be available freely for anyone to download, which I think should help. But our several projects that lower doses, there was no gauge it was basically encouraging calls to share best practices, and that led to lowering doses dramatically in some settings.

CO-CHAIR SUSMAN: Okay, thank you. As we come to a vote, really, there's two or three paths we can take. One is to consider this as a single measure, if you would, and vote on both components, mean dose and proportionate CTs greater than 75th percentile of benchmark together, or we could separate those out and vote for each. Obviously, you need to have a mean dose before you can benchmark it. Just like the committee's input about how we consider that. Yes, Dave.

MEMBER KELLER: I'd suggest that we consider these two together. As I hear the two measures, what it reminds me the most of is how you look at a distribution and you're really doing is identifying what would be the standard deviation. While one of them is identifying sort of your general level, and the other is identifying the degree of variation within your practice, I think they're really all of a piece when you're trying to look at how this material is used. I'd suggest we look at them together.

CO-CHAIR SUSMAN: Ricardo.

MEMBER QUINONEZ: The only caveat to this is that I'm not convinced, from hearing the developers are looking at this measure, that there are general accepted national benchmarks. It seems like there are a few. There's some national ones. There's some being developed. There's some that might come out of this measure, but having it together when there are no nationally accepted benchmarks that everybody agrees on would be -- how do you implement this if there are none?

CO-CHAIR SUSMAN: Keith, maybe you could comment from a radiologist's perspective on are there nationally accepted benchmarks? How would radiologists who are actually doing this work look at this?

MEMBER WHITE: I'm going to go back to the statement I made earlier that what they say what is acceptable imaging quality is in truth, in all honesty, is a subjective assessment. There is really very little objective data

evidence that directly correlates, for example, radiation dose for a specific indication what the diagnostic accuracy of that study. So to a certain degree, these are surveys of educated and dedicated teams that are working in this area. But, for example, the work that Dr. Goske did has been referred to. I know that particular study, where they were establishing some national benchmarks for CT dose, involved a number of different children's facilities, all of them coming together.

It included not just an assessment of the doses that were used, but an assessment of some observer studies looking at the sharpness and detail of certain anatomic structures to try to put some objectivity into whether or not the images were of adequate diagnostic accuracy. I don't believe these benchmark studies are devoid of any objective assessment of image sharpness or image quality.

The truth is that they're the best we've got and I'd like to improve and mature over

time. But I think it would be a shame to not have the measure pass simply because we don't have really rigorous measures of image quality, which I think is extremely difficult to do and will take a lot of effort to put together.

CO-CHAIR SUSMAN: Okay, thank you very much, Keith. We have two more comments, and then we'll see if we can vote here. Virginia.

MEMBER AGORATUS: This is Lauren. I have a question.

CO-CHAIR SUSMAN: Sure, please.

MEMBER AGORATUS: I'm just wondering if -- first of all, if there's any difference between CT scans with and without contrast, and also if anyone is looking at -- and this may be beyond the scope of this measure -- in terms of reducing dose length and frequency of scans medically complex children who have multi-system involvement and may be receiving multiple scans within days, or even the same day?

CO-CHAIR SUSMAN: Keith, maybe you could comment on the first part.

MEMBER WHITE: Because she asked two questions, and due to time, I may have to answer first question only.

CO-CHAIR SUSMAN: You can answer both if you like.

MEMBER WHITE: Restate the first question.

MEMBER AGORATUS: If there's any difference in outcomes between CT scans that are used with and without contrast?

MEMBER WHITE: In terms of radiation dose, typically, hospitals do not have distinct radiation profiles for with or without contrast. They don't differentiate at that level. Generally, there's not going to be a difference in the radiation administered, whether contrast was given or not. I'm sure there are exceptions to that, but as a general rule, that's not done. Second question, repeat that one.

CO-CHAIR SUSMAN: It's about medically complex children who might be getting multiple scans, even on the same day.

MEMBER WHITE: The question is risk in those patients? The right way to think of -- the whole theory behind radiation dose here is -- I think the best analogy, it's like flying in an airplane. Every time you get on an airplane, there is a risk that something bad's going to happen.

The fact that you have flown 1,000 times doesn't change the risk that on the next flight, you could have an event. Patients that had repeated examinations cumulatively over time have an increase of developing cancer, but the increase in the risk of cancer that occurs from any single CT event is the equivalent for them as it would be for any patient.

CO-CHAIR SUSMAN: I appreciate your insight, since we're all flying home. Virginia.

MEMBER MOYER: I had has to do with the benchmarking issue. As I understand this measure, the measure is are you measuring? It isn't are you meeting a particular benchmark because the benchmarks are currently not really

out there. I think that -- and the documentation provided, I thought, by the developers that simply measuring is associated with decrease in dose, and at this point in history, that probably is where we are with this. Maybe ten years from now we will be there, our benchmarks --

(Simultaneous speaking.)

CO-CHAIR SUSMAN: As I understand this, there are really two components. One is indeed the mean dose, and the other is the proportion of CT scans with doses greater than 75th percentile.

MEMBER MOYER: Right. The question still is are you measuring? It's two different measures. It's essentially the mean and standard deviation. The question is not how many patients are and we're going to judge you on that; the question is did you measure? Because just measuring alone seems to result in improved in dosage.

CO-CHAIR SUSMAN: Again, I'm fine, and if the will of the group -- let's hear Jim -- is

to just consider this as a single measure that's closely tied, obviously as more and more data are accumulated, that'll allow us to establish more rigorous, nationally accepted benchmarks. Jim?

DR. SMITH-BINDMAN: Can I say just one thing about the last comment? It's exactly right that just measuring is useful, and even in the absence of any external benchmark, institutions can compare their doses on a monthly or quarterly basis to their own doses from the same prior period, and then encourage doses to go down. That's what we happened to do at our institution. Our doses are quite lower than national averages, and yet we keep trying to achieve lower doses for our patients by having our doses be compared monthly, over time.

CO-CHAIR SUSMAN: Okay, I'm going to get Jim and John, and then see if there are any final questions.

MEMBER BOST: For the second measure, I'm just not sure how you can collect this measure unless the developer provides or endorses

a value that is considered the 75th percentile, so that everybody who does this measure can report it consistently.

CO-CHAIR SUSMAN: Thank you. John.

CO-CHAIR BROOKEY: Yes, I just had a question. Just measuring this and reporting it out publicly, it could be reported out comparatively with percentiles and have really the same effect as comparing it to benchmarks which have not really yet be established.

So to me, the first measure can really effectively fulfill the intent of both, depending on the way it's reported and how it's displayed, so people could compare themselves to other institutions. I'm not sure if that was the intent or not, but I'm just looking at it and thinking that would sort of solve the issue. That's a question for the developer.

DR. SMITH-BINDMAN: I'm sorry, can you repeat the question part of it? Sorry.

CO-CHAIR BROOKEY: If the first measure is to measure, and then to report, it

could be reported comparatively, institution to institution, and even reported by percentile, irrespective of whether there's a benchmark or not. It can just be comparative data. I don't know if that was your intention or not.

DR. SMITH-BINDMAN: No, that's absolutely true. This provides data that currently don't exist.

CO-CHAIR SUSMAN: Okay, I would suggest we move to a vote on evidence, considering the whole discussion that we've had. Is there any comment? Robyn, you --

DR. NISHIMI: I just need to clarify for folks the submission set outcome, the developer said today intermediate outcome, and what I just heard now was the measure is did you measure, which is a process. So just putting it out there that we need a little bit of clarity here.

CO-CHAIR SUSMAN: I heard earlier in the discussion that this was an intermediate clinical outcome of interest from many of you,

whether I personally agree with that or not -- if that's the committee's wisdom, it's going to send us down one path. If it's a process, it's going to send us down something else and so forth, so we need to be really clear.

MEMBER FINKELSTEIN: I'm very comfortable with this as an intermediate outcome, and I think the benchmark -- I think it's did you measure, but I think the benchmarking is part of it. It's implicit in it, and I would go across the top pathway on this. That's my two cents.

DR. NISHIMI: Actually, an intermediate outcome sends you down anyway.

CO-CHAIR SUSMAN: Yes, it sends you down.

DR. NISHIMI: But we like to characterize our measures to our members correctly. That's why I'm pressing it.

CO-CHAIR SUSMAN: Virginia, or did you have another comment, Jim? Okay, anybody else -- yes, please.

MEMBER MORROW-GORTON: I just want to

sort of echo what you said. I think this is not did you measure, it's what was the measurement, which is an outcome, not a process.

CO-CHAIR SUSMAN: Let's consider this, then, an intermediate outcome measure, which is going to send us down the blue boxes in our typical fashion here for evidence. That will allow us to be clear for anyone using this measure, as well as for all of you in voting on it. We have evidence and high, moderate, low, or insufficient. We're going to vote for this en bloc, the two parts of this, if you will, measurement.

MS. CHAVEZ: We're now voting for Measure 2820, pediatric CT radiation dose, 1 high, 2 moderate, 3 low, 4 insufficient. Voting is open, 20, we are expecting 26 votes, 23, 26, 7 voted high, 16 voted moderate, 1 voted low, 2 voted insufficient. This measure passes evidence.

CO-CHAIR SUSMAN: With regard to gap, then, opportunity for improvement. Deb, you're

one of our subcommittee members.

MEMBER FATTORI: Our discussion at the subcommittee really was around that the gap was illustrated by the variability in radiation doses. We also discussed -- the developer provided some of their data through their work -- that there was -- in their work, they found a bigger gap, if you will, higher radiation in county hospitals as compared to more academic hospitals. We felt that was significant and further illustrated the gap.

CO-CHAIR SUSMAN: So we've had one member who looked at this closely concur that there is a gap here. Other comments before we vote? Okay, let's move to vote on the performance gap, 1 is high, 2 moderate, 3 low, and 4 insufficient.

MS. CHAVEZ: Voting on gap 2820. Voting is open, 24, 26, 11 voted high, 14 voted moderate, 0 voted low, 1 voted insufficient. This measure passes gap.

CO-CHAIR SUSMAN: Let's move on to

reliability. Keith, you want to talk to reliability of this measure, perhaps?

MEMBER WHITE: I thought there were questions -- I think they've already been discussed -- about the clarity on the data elements that needed to be defined. I think that I would agree with what was said earlier, that it would really be helpful if, as a part of the measure, there was at least some attempt to put a benchmark standard that could be used in the measure. I think that I agree with what was stated earlier that if there's not an accepted benchmark that putting in the measure the requirement to compare with a benchmark is not very useful.

CO-CHAIR SUSMAN: And the Kappa scores on a limited number of locations, I think, were certainly adequate. Any other comments on reliability? Let's go ahead, then, and vote, 1 high, 2 moderate, 3 low, 4 insufficient.

MS. CHAVEZ: Now voting on reliability. Voting's open, 21, 24, 26, 5 voted

high, 17 voted moderate, 1 voted low, 3 insufficient. This measure passes reliability.

CO-CHAIR SUSMAN: And that brings us to validity. Keith, any comments on validity?

MEMBER WHITE: Nothing that hasn't been discussed already.

CO-CHAIR SUSMAN: Deb, any further comments on validity?

MEMBER FATTORI: The only question that came up was in regards to exclusions. There was one area in the measure summary that said there were no exclusions, and then there was another area that listed the exclusions out, so if the developer would just clarify that.

DR. SMITH-BINDMAN: I think that was my error in how I described it. The measure looks at consecutive exams in the specified anatomic areas, head, chest, abdomen, pelvis. It does not look at exams in other anatomic areas. It does not look at radiation oncology-guided CT scans, and it does not look at procedure-related CT scans. Those are not really exclusions, as

much as areas that are not covered by this measure.

CO-CHAIR SUSMAN: Thank you. Kevin.

MEMBER SLAVIN: Just a question about the indications for scanning and if there's a particular facility or area that has a specialty where there's a high volume of patients with a particular indication, how would that affect where they score on a benchmarking compared to other facilities, and would that then sort of be problematic for the validity of it as a quality measure?

DR. SMITH-BINDMAN: I think that is a theoretical concern, but not a practical concern because there are no such indications that would be expected to have such a profound effect on the average doses. If a facility did, however, have, let's say, only obese children -- if it was a facility dedicated to studying obese children, then their doses would be higher than the averages, and they probably should not compare to those benchmarks.

But other than that example, I can't come up with any other indication that should lead to such a profound shift in typical doses. We have a large paper coming out in radiology that looks at 800,000 exams. In fact, the indication for imaging is much less important predictor of dose than just institutional preferences and variations.

CO-CHAIR SUSMAN: Thank you. Any other questions on validity? Hearing none, seeing none, let's go ahead and vote on validity, 1 high, 2 moderate, 3 low, 4 insufficient.

MS. CHAVEZ: Voting is now open, 5, 21, 24, 25, 26, thank you.

CO-CHAIR SUSMAN: Robyn has a comment.

DR. NISHIMI: I can do it at the end.

CO-CHAIR SUSMAN: Okay, at the end.

MS. CHAVEZ: Okay, 6 voted high, 16 voted moderate, 0 for low, 4 insufficient. This measure passes validity.

CO-CHAIR SUSMAN: Okay, I think we've talked about the feasibility of this in some

detail. Kerri, you have a question?

MEMBER FEI: I do have a question. I was trying to think of the right time to bring this one up. I noticed that in the documentation it says the measure could be used at the health plan level. It has not been -- it doesn't appear that it's been tested at the health plan level.

The other problem is most health plans right now are not going to have access to any of this data. They're going to have to go to providers to get it, either through contracting with a service or getting direct access to EMR or having providers report data to them and the labor intensive and burdensome at a health plan level right now, in our current state, to be able to even get this measure.

CO-CHAIR SUSMAN: So you'll be happy to know that Robyn was right on this issue, and that was the concern she had, so the NQF is working for you.

MEMBER FEI: Yes, I think in the long term, this is something health plans would like

to be able to have access to, but it's just not -- we're not there yet.

DR. SMITH-BINDMAN: Can I make a comment about that?

CO-CHAIR SUSMAN: Sure.

DR. SMITH-BINDMAN: We actually have assessed these doses at many health plan levels. If an HMO is considered a health plan, we've assessed these doses across six large integrated health plans. We've included those data in two JAMA papers. I currently have a project that's collecting similar data across 100 hospitals across the country, including several very large health plans. So I agree these data take some work to assemble. It's relatively modest, and we have assembled them across many small and large health plans.

CO-CHAIR SUSMAN: Thank you. Kerri?

MEMBER FEI: There's a difference between a large integrated health system that has a health plan associated with it, so your Kaisers or Intermountains or Geisingers that yes, they

can do this. Your run-of-the-mill private health insurer right now cannot.

CO-CHAIR SUSMAN: I think the point is well made. Thank you. Other questions about feasibility?

DR. NISHIMI: I just want to make the point, though, that under the NQF rubric, we will be asking the developer to remove that checkbox.

CO-CHAIR SUSMAN: Certainly, that's work I think we'd all like to be able to see in the future, but let's go ahead, then, and vote on feasibility.

MS. CHAVEZ: Voting on feasibility, 1 high, 2 moderate, 3 low, 4 insufficient. Voting is open, 9, 23, 25, 25, 26, thank you, 9 voted high, 12 voted moderate, 3 voted low, 2 insufficient. This measure passes feasibility.

CO-CHAIR SUSMAN: Okay, now let's go to usability. I think again, we've discussed this broadly. Are there any other new comments? Seeing and hearing none, let's move to vote on usability.

MS. CHAVEZ: Voting on usability, 1 high, 2 moderate, 3 low, 4 insufficient. Voting's open, 18, 24, 25, 26, 10 voted high, 14 voted moderate, 1 voted low, 1 voted insufficient. This measure passes usability and use.

CO-CHAIR SUSMAN: Finally, the overall voting on this measure, overall suitability for endorsement, 1 yes and 2 no. Are there any final comments? Seeing none, let's vote.

MS. CHAVEZ: We're now voting on measure --

CO-CHAIR SUSMAN: Wait a minute, excuse me. Maureen, you did have a comment.

MEMBER EDIGER: Sorry. I have a kid who this totally applies to. He has Spina Bifida, so gets CT and MRI scans very frequently. You had asked about changes in behavior. Just the more transparent this information is, it absolutely will change behavior. It's been great being part of this process.

CO-CHAIR SUSMAN: Thank you very much

for that comment. Any other final comments? Okay, let's vote, 1 yes, 2 no on overall suitability for endorsement.

MS. CHAVEZ: Thank you.

CO-CHAIR SUSMAN: Thank you.

MS. CHAVEZ: Voting on Measure 2820, overall suitability for endorsement. Voting is open, 24, 25, 26, 24 voted yes, 2 voted no. This measure, 2820, pediatric CT radiation dose, has been recommended for endorsement.

CO-CHAIR SUSMAN: Okay, thank you very much. That was good work, smoothly done. We're going to plow ahead, trying to keep us on schedule here.

PARTICIPANT: Thank you.

CO-CHAIR SUSMAN: Thank you so much, appreciate all the effort on this.

DR. SMITH-BINDMAN: Thank you very much.

CO-CHAIR SUSMAN: So we're going to move to the overuse element. If you need to take a break for coffee, the restroom, please feel

free, but in an effort to keep us on schedule, I'm going to plow further. This is overuse of imaging for the evaluation of children with post-traumatic headache, a Q-METRIC proposal from the University of Michigan. We should have, on the line, either Gary Freed or Julie McCormick or both. Are you there?

DR. MACY: Hi, this is Michelle Macy. I'm a pediatric emergency physician at the University of Michigan and the Q-METRIC topic lead for this measure. We appreciate the opportunity to provide opening remarks on behalf of the entire Q-METRIC team, who's here with me.

CO-CHAIR SUSMAN: Okay, any recusals on behalf of the committee? Seeing none, we'll go ahead and hear from our measure developer for a brief description.

DR. MACY: The measurement of imaging overuse for headache and seizures was determined by CMS processes to be of high national priority, especially for Medicaid patients. This topic was assigned by CMS to Q-METRIC for measure

development and testing. Reduction in overuse can benefit children by lowering immediate risks associated with sedation and anesthesia and lifetime risks associated with radiation exposure, which include leukemia, brain tumors, and other malignancies. Reduction in cost is an additional benefit.

Our proposed measure is crucial, in light of the strong evidence for the existence of imaging overuse in the evaluation of children with head injury, as demonstrated by the Pediatric Emergency Care Applied Research Network or PECARN studies that led to the development of two age-stratified clinical decision rules for the identification of children at very low risk for clinically important traumatic brain injury. As this measure is discussed, it's essential to keep in mind a few important points. First, our assignment from CMS was specified to children with headache. Children with post-traumatic headache are a subset of a broader group of children who have experienced a head injury.

Children 2 years and older were selected as the initial population of focus for this measure because we were not assigned the task of -- because we were assigned the tasks of measuring imaging overuse for headache. Younger, pre-verbal children cannot reliably report this symptom. Future measure enhancements can broaden the eligible population. Second, the sole focus of this measure is overuse. Overuse is distinct from use, under use, and inappropriate use.

Overuse is defined as any patient who undergoes a test or procedure without an appropriate indication. Overuse calculations are therefore derived from the population of patients who have undergone the test of interest. Specifically for this measure, the denominator consists of children who underwent neuro-imaging for the evaluation of post-traumatic headache in the emergency department within 24 hours of injury. The numerator includes a subset of children from the denominator who received the neuro-imaging test without an appropriate

indication. Our measure inclusion criteria are aligned with the PECARN study.

We also relied on the PECARN study to determine indications for neuro-imaging that exclude cases from the numerator. Children with an indication for neuro-imaging documented in the medical record were excluded from the numerator, leaving behind children without an indication for imaging. Through testing, we found that administrative claims can be used reliably and feasibly to narrow the population eligible for chart review.

At the present time, chart review is necessary for the calculation of this measure because there are no ICD-9-CM or ICD-10-CM codes for key measure inclusion criteria, specifically neuro-imaging obtained within 24 hours of injury among children with normal mental status. More than half of the 200 charts sampled for testing were excluded due to lack of documentation of injury occurring within 24 hours prior to the ED visit in which imaging was obtained. We did

include ICD-9-CM codes for concussion without loss of consciousness in the population eligible for the chart-review sampling. Based on comments from the workgroup call, we have recalculated overuse in our test sample, including children with concussion and no documentation of headache in chart review.

The inclusion of these children resulted in higher overuse at 18 percent, and just for reference, overuse in the PECARN study was around 20 percent. In closing, we believe that this evidence-based measure is highly important and has potential to make substantial positive impact on the quality of care for children.

Our measure focuses on the overuse of imaging, which is a national concern and priority for CMS. Our testing found this measure to be valid, reliable, and feasible through chart review. Efforts to enhance this measure and expand on its utility can be pursued in the future. We look forward to addressing any

questions or concerns you may have, and thank you for your time and consideration.

CO-CHAIR SUSMAN: Thank you for that very succinct and thorough presentation. One of the things to orient us here, this is an overuse measure, and it looks a little different because the denominator is those who have undergone neuro-imaging, and the numerators are those with no appropriate indication, as opposed to -- I think it's the PECARN study was laid out kids come in, they have a variety of concussion, blah, blah, blah, and measuring how many have an appropriate indication.

It's a different denominator. It gets a little bit fuzzy when you start thinking about that, so just keep that in mind as we discuss this measure. We're looking at overuse, and the denominator is those who have undergone neuro-imaging. Did I get that right, developer?

DR. MACY: That's correct.

CO-CHAIR SUSMAN: We have a lot of questions here. Let's have these four questions,

and then probably it would be useful to have those who reviewed this most thoroughly -- maybe I'll start with Jeff, since you were on the subcommittee that did that.

MEMBER SCHIFF: I was on the subcommittee. I think the developer already talked about one of the concerns we had about the limitation around headache and addressed that, so I think that's been addressed. I think that the -- I have one major comment and one smaller one.

The major one I'm concerned about is the numerator exclusion for absence of documented neurologic examination because I just feel that would mean that -- and I need the developer to help with that because unless I'm missing a double negative, that seems illogical to me. Then the other one I just want to talk about is I'd like to hear about why that is not a hospital-level measure.

CO-CHAIR SUSMAN: Why don't we get responses from the developer on those issues while they're fresh at hand.

DR. MACY: With regard to a chart where there is no documented neurologic exam, we didn't feel that was a chart that was appropriate for inclusion in the measure. I think that ends up being a rare occurrence for kids who are undergoing neuro-imaging events. It might be a point that we could look at specifically in our chart review elements and get back to you as to the number of cases that we had that occur. Then the second question?

MEMBER SCHIFF: Was about health plan level.

DR. MACY: And why we went with a health plan level? We went through the health plan level because that was the Medicaid focus for the CMS measures, and our approach for testing was designed to be in line with that. We think that future enhancements of the measure could be looked at at the hospital level, but we do not have testing results at this time for that.

MEMBER SCHIFF: Can I just ask a

follow-up question?

CO-CHAIR SUSMAN: Please.

MEMBER SCHIFF: Do you have numbers on the -- you mentioned that it was a rare occurrence to have an undocumented neurologic examination. My concern was that obviously would mean that there was absolutely no indication documented. What was the rate on your review without a documented neurologic examination, do you know, or did I miss it?

DR. MACY: About 7 of the 200 cases that we did chart review on had no documented neurologic exam.

CO-CHAIR SUSMAN: Thank you. Let me turn to Jim, who is our expert in this area. We want to make sure he has a chance to say wise words.

MEMBER DUNCAN: I don't know if I'm an expert. I would actually think Jeff, in the emergency room, probably sees these children on a regular basis. I have personal experience with my son going through this process. I guess I'm

guilty of carrying a slightly different nomenclature for overuse in my head. It sort of goes to antibiotics. Antibiotics for viral symptoms are often cited as an example of overuse of antibiotics.

So, therefore, my mental model was looking for a measure that was looking for head CTs, much like PECARN, in children presenting with head trauma, and not with the exclusion of headaches, which I also think begins to sort of pull apart the strong linkages back to a considerable amount of data, which I think no one will argue with. That was part of what we discussed as a subgroup before is part of my concerns.

CO-CHAIR SUSMAN: I think, again, getting our heads around, so to speak, that we're starting out with neuro-imaging in the denominator is a different mental setup than, say, how PECARN study was set up, so an important point. I'm going to go back to Carol, and then we'll get the rest of the comments in order.

MEMBER STANLEY: I'm just curious why children who are suspected of neglect are going to be excluded? With the Medicaid population, I would think that the exclusion, if you tested it with any Medicaid health plan specifically, if you would have had some different results?

DR. MACY: In our committee meeting within the Q-METRIC group to design this with the national panel, it was strongly expressed among the emergency providers that CT imaging of the head is an important step in the evaluation of children with suspected child abuse and neglect and, therefore, we felt that being in a situation where providers could be doing that appropriate evaluation of children with that clinical concern should be excluded from this measure.

CO-CHAIR SUSMAN: Okay, Deb?

MEMBER FATTORI: I'm a little reluctant to bring this up, but I'm just a little confused as to why this is categorized as a process measure, where our last imaging measure was an intermediate outcome measure, so I'm

trying to get my head around that and appreciate any input.

CO-CHAIR SUSMAN: I think the process is obtaining the image, and as everybody sitting on this panel knows, the classification of these measures remains something that troubles us. I'm not sure there's much more to say about that, unless the staff wanted to weigh in on it. As Helen said, it's up to us, guys and gals.

MEMBER FEI: I could offer one thing. It's a process of whether or not something was done. It's not the outcome or the report or the result. It's yes or no. It's a yes or no answer.

CO-CHAIR SUSMAN: This, to me, is a pretty classic process measure, but I understand the discomfort. Did you have anything else, Kerri?

MEMBER FEI: I did, actually, I just --

(Simultaneous speaking.)

CO-CHAIR SUSMAN: Go ahead.

MEMBER FEI: -- just for the fun of it. Understanding that this is a health plan measure and it was tested Medicaid data and had chart abstraction that's part of it, again, generalizing to health plans across the board can be done, but would incur expense and, in the present time, is difficult.

CO-CHAIR SUSMAN: Thank you. Ricardo.

MEMBER QUINONEZ: I was part of the subcommittee that looked at this measure. My only -- for sure, there's no question that the evidence strongly supports this, except that as has been discussed before, but I'd like a little bit more clarification from the developers, by including imaging in both the numerator and the denominator, you're excluding most of the kids that would benefit from this measure.

Most kids who show up to the ER after a minor head injury don't have a post-traumatic headache. I'm just a little bit confused and concerned that if we pass this measure, this will be seen as the measure of overuse for minor head

injury when the numerator could be all children presenting to the ER with minor head injuries, which would actually capture the spirit of what overuse is. This is looking at a subset of a subset, and it would capture a minority of the children who have had imaging done improperly.

CO-CHAIR SUSMAN: I think that's a tremendously important point, but probably goes in our usability and, ultimately, to our vote on this measure. I'm going to suggest we sort of stay on the point of evidence. Recognizing that this may not be the best measure we could think of, we still have to vote on the measure that's before us. Thank you for those comments. Yes?

MEMBER MORROW-GORTON: I just want to go back to the question about children who've been abused. I think they have to be excluded. One, you can't rely on the history from the family or whoever brings them in. Two, you don't know whether they had a head injury or not.

There are a lot of legal things you have to do. I just think they're a completely

different -- it's a completely different work up, a completely different way to think about it. You're not thinking overuse. In that population, you're really not thinking overuse.

CO-CHAIR SUSMAN: So essentially, the pretest probability of a positive test is going to be much higher. There's a much higher prevalence of bad things. Go ahead, John.

MEMBER MORROW-GORTON: I absolutely --

(Simultaneous speaking.)

MEMBER STANLEY: I was specifically thinking about neglected children, not necessarily abuse.

MEMBER THACKERAY: That's what I was going to say. I absolutely agree with Jill that if there's a consideration of physical abuse, none of this should apply. It doesn't apply for PECARN, and I'm glad to see it doesn't apply here. But I do think saying child abuse or neglect is different, there's very few indications to be doing head CTs when you have a concern for neglect. So it might be just more

clear to specify suspected child physical abuse and not say abuse or neglect.

CO-CHAIR SUSMAN: It's a good comment, and maybe, if you will, sloppiness in our language or documentation that drives some of that, but we'll certainly make sure that's taken into consideration. Jim.

MEMBER AGORATUS: This is Lauren. I have a question.

CO-CHAIR SUSMAN: Go ahead.

MEMBER AGORATUS: I'm a little concerned because it says in the absence of neurological symptoms, yet prolonged headache is listed as a neurological symptom, so I'm not sure, again, if we're going to be not testing people that should be tested.

CO-CHAIR SUSMAN: Does the developer want to respond?

DR. MACY: It's the absence of neurologic symptoms that would be indicative of clinically important traumatic brain injury, as outlined by the PECARN studies, so not acting

like themselves, persistent vomiting, those sets of symptoms that were identified in the PECARN study were the ones that we were including in that statement.

CO-CHAIR SUSMAN: So that did not include prolonged headache?

DR. MACY: I think headache was one of our sets of symptoms, and because we're looking at within 24 hours of injury, I don't think you would cross that threshold of prolonged, necessarily.

CO-CHAIR SUSMAN: Fair enough, thank you. Jim.

MEMBER DUNCAN: It might be best left to the validity discussion, but I think this idea of when the denominator gets small, rather than all the children with head trauma, you start introducing that the abuse case can sort of tweak the numbers, and that's the advantage of having a large denominator is it's going to improve the validity. I think the developer found no cases that needed the child abuse exclusion in their

test set. So I would think that, again, if you expand the denominator, again, it's going to have a very small impact.

I also think -- and it's part of the usability discussion -- the more exclusions -- trying to calculate these within our department, it makes it really tough because you have to get the data, and then you have to run the algorithm. The simpler the measure, the more it fits -- I think we're trying to impact the decision making in the ER, by the physician, about whether this child needs a head CT. You want to be getting credit for making a good decision which is, again, outside this measure.

CO-CHAIR SUSMAN: So I think again, good points, probably more apropos to usability. Any discussions further on evidence here? Seeing none, let's go ahead and vote on evidence, which is 1 high, 2 moderate, 3 low, and 4 insufficient.

MS. CHAVEZ: Now voting on evidence for Measure 2802. Voting's open, we are expecting 27 votes, 19, 23, 25, 26, okay, 26

votes, 8 voted high, 15 voted moderate, 2 voted low, 1 voted insufficient. This measure passes evidence.

CO-CHAIR SUSMAN: So opportunity for improvement or gap? Is there a gap here? Keith, any members of the subcommittee? Jim?

MEMBER DUNCAN: Again, it goes back to how they calculated their measure, the other examples that I've seen calculated using different -- clearly shows overuse of head CTs, but it's, again, hard to say that this measure would parallel what's been seen by other calculations.

MEMBER QUINONEZ: I agree. I don't think there's a gap that has been shown by the developers because of the inclusion of the denominator of only kids with headache. So the gap is in all children who present to the ER with minor head injury. I don't think they have proven that for their specific measure there's a gap.

CO-CHAIR SUSMAN: Does the developer

want to respond or have any further input? Then I'll get Jeff.

DR. MACY: It has been shown that CT utilization ranges widely across emergency departments and free-standing children's hospitals from about 20 to 60 or 70 percent of kids are imaged, and we have no reason to believe that imaging inappropriate or overuse measurement would not also have a similar spread.

CO-CHAIR SUSMAN: Jeff, and then John.

MEMBER SCHIFF: In Minnesota Medicaid, we looked at CT scans by Level 1 and 2 trauma centers and found a range from 27 percent to 80 percent for head injury. This is probably seven or eight years ago. So there was a big variation, but I think the concern that Ricardo brought up is really the relevant one, is are we going to identify a significant gap in a subset of kids just with headache? I guess what I'd like to know from the developer is is there any -- can you tell us any larger numerator and how many kids, of all the kids that were looked at,

had -- what percent of kids had headache? Because we, in my ER, almost never would code headache.

CO-CHAIR SUSMAN: Developer?

DR. MACY: One moment please.

CO-CHAIR SUSMAN: Okay, maybe in the meantime, Jon you want to comment?

MEMBER FINKELSTEIN: Yes, as I'm thinking of this, there are two separate issues here. One is is the patient population that the measure addresses too narrow, and is it either unimportant -- that's not the evidence part -- or there isn't enough evidence in this subgroup of a subgroup, if you will.

The second is the way the measure's specified, which I think is a separate issue, it is -- and the developer's laid this out pretty well. The way it's specified is as a fraction of scans, inappropriate scans as a fraction of scans. The other way you can think of it is as a fraction of people. I agree that it's more natural to think of it as a fraction of people.

It's what they call an inappropriate use measure. They're calling this an overuse measure. In both cases, the target should be zero, but that's a separate issue. That's how they've specified it. I don't think that's the evidence or gap.

CO-CHAIR SUSMAN: I think it boils down to we know there's a lot of overuse of scans, but we're looking at this as an overuse measurement and is there a gap there? Questions of whether that exists. Developer, do you have an answer for Jeff?

DR. MACY: Yes. First I'd like to note that this was a determined priority of CMS for the nation. We have a denominator of children with post-traumatic headache of 5,912, based on ICD-9 codes that include the specific code for post-traumatic headache, as well as concussion without loss of consciousness.

CO-CHAIR SUSMAN: Okay, Jim?

MEMBER DUNCAN: I don't think we've yet seen a performance gap at the health plan level. Most of us will quote data showing

hospital-to-hospital variation or physician-to-physician ordering variation. I don't know if I've yet seen evidence that there's performance gaps at the health plan level, nor did, I think, in the submitted materials did we see that.

CO-CHAIR SUSMAN: Yes, so there's, I think, some important issues that have been raised. Is there any other further comment? If not, I'd suggest we vote on gap.

MS. CHAVEZ: Now voting on gap for Measure 2802, 1 high, 2 moderate, 3 low, 4 insufficient. Voting's open, 25, 27, 2 voted high, 9 voted moderate, 11 voted low, 5 voted insufficient.

CO-CHAIR SUSMAN: My reckoning of this is it does not pass on this, is that correct?

MS. CHAVEZ: This puts us in the gray zone.

DR. NISHIMI: No, we have 27 votes now, though, so it's --

CO-CHAIR SUSMAN: Please.

DR. NISHIMI: You need 17 to pass.

(Simultaneous speaking.)

CO-CHAIR SUSMAN: Do you want us to continue, I guess, is the operative question?

It's right on the border. It's gray zone. We will continue and let NQF staff sort out how they treat this. Moving right along, we've considered gaps, and now we get to reliability. Any comments on scientific reliability? Jim?

MEMBER WHITE: This is Keith. One of the concerns I have, because this relies on chart review, and because -- and maybe the developers could clarify a little bit more, when it talks about identifying the neurologic findings that would warrant the testing, there's, I think, a great deal of subjectivity there. I think there's a question about how reliably the data could be collected with chart reviews, especially where the measure is somewhat vague on what constitutes the specific neurologic indications for concern.

CO-CHAIR SUSMAN: Does the developer want to comment?

DR. MACY: Yes, we had a high degree of reliability between the chart review elements across our data abstractors. It was 99 percent, with a Kappa of .98.

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

MEMBER DUNCAN: I always worry about very, very high values in agreement on something that I think is going to be very subjective in reviewing charts. Also, especially when you blow this up to try and review charts across the country, and with a pretty low number of charts being reviewed at any particular site or with any particular health plan, you're probably going to get different reviewers and abstractors. I don't know if that reliability would hold.

CO-CHAIR SUSMAN: Okay, Jeff.

MEMBER SCHIFF: The numerator exclusions are on the top of Page 28, and they are abnormal neurologic signs and symptoms of

intracranial pressure increased, intracranial pressure, e.g. decreased alertness, altered mental status, GCS less than 14, gait disturbance.

I have to tell you that my concern is that if this becomes a health plan level measure by which ER docs are somehow held accountable to health plans, we already document these things when I want to do a CAT scan, and if I know I have to meet the criteria, they will always meet these. So it may or may not be that the Kappa will be variable depending on how well documented it is. My concern is that we're going to create a market or a process by which people will document well, rather than get to an exclusion, if that makes sense.

CO-CHAIR SUSMAN: Probably use your microphone if you're going to make comments. I'm sorry.

MEMBER SCHIFF: We never imagine otitis in the emergency room, Dr. Moyer.

MEMBER MOYER: Got it.

CO-CHAIR SUSMAN: Okay, that's all off the record, thank you. Any other comments about the reliability? I think if we go -- yes, Kevin?

MEMBER SLAVIN: Maybe I missed this in what Jeff had just said, but is there a list or a tool to identify the history of the other medical conditions that would warrant neuro-imaging? I'm specifically thinking of things like genetic collagen disorders that prone vessels to rupture, or other types of medical conditions where neuro-imaging might be indicated, even in the absence of specific findings.

DR. MACY: We do have an extensive list based on ICD-9-CM codes that were translated over to ICD-10-CM codes that are included in our documentation.

CO-CHAIR SUSMAN: And the testing here was also at the element level for validity, which I think will reflect back to the reliability testing that we've discussed in prior measures here, and not just the inter-rater reliability. David, did you have a comment?

MEMBER KELLER: Yes, just a question, based on what you just said. So the exclusions are applied at the coding level, not by the chart abstractor? I would have thought you'd also want to have a list of diagnoses or trigger words that you'd be looking for in chart abstraction that were not code based, or at least most of us don't write codes in our notes.

DR. MACY: We used the ICD-9-CM codes to screen the administrative data for the presence of those conditions, and then with the chart review, those different kind of categories, like sickle cell disease and blood dyscrasias were included. Brain tumors were included for things that people were looking for in the chart.

CO-CHAIR SUSMAN: Thank you. Any other discussion of reliability? Okay, let's move to vote, 1 high, 2 moderate, 3 low, 4 insufficient on reliability. I think Robyn has a comment.

DR. NISHIMI: Yes, I just wanted -- a reminder that this is at the performance element

level, so the highest eligible score is moderate.

CO-CHAIR SUSMAN: Is moderate, yes. Thank you.

MS. CHAVEZ: Now voting on reliability. Voting is open, 20, 26, 27, 0 voted high, 12 voted moderate, 13 voted low, 2 voted insufficient, and this takes us in the gray zone.

CO-CHAIR SUSMAN: Another measure in the gray zone. Let us press on to validity. Comments on validity, Jim or Keith, others? There was empirical validity testing done at the element level, for what that is worth.

MEMBER DUNCAN: The validity testing, as I read through the measure, was not on the data level. It's really the expert panel face validity, is that correct?

CO-CHAIR SUSMAN: Measure developer, please clarify. I thought there was element validity.

DR. MACY: We do have element validity.

CO-CHAIR SUSMAN: Thanks. Ricardo.

MEMBER QUINONEZ: My biggest concern with validity here is that it -- to me, again, it just fails to pass that this measure would adequately measure quality of care provided, again, because of the numerator, the way it's written. That, to me, would make it not valid for the spirit and the intent of the measure.

CO-CHAIR SUSMAN: It's going to be difficult, I think, for some of us to separate out how we view this and where we vote our conscience on this, so thank you. Any other comments about validity? If not, I'd suggest we go ahead and vote, 1 high, 2 moderate, 3 low, 4 insufficient.

MS. CHAVEZ: Now voting on validity. Voting is open, 18, 25, 27, 1 voted high, 8 voted moderate, 17 voted low, 1 voted insufficient. This measure does not pass validity.

CO-CHAIR SUSMAN: I think we actually have gone beyond the gray zone to the death zone. This measure didn't make it. I think everybody agrees that overuse of imaging is an important

concern. What I've heard is that by limiting it to those with headache who have neuro-imaging, the definition, whether it be by CMS, at least those of us in the field feel like we have rarefied the sample so much that it will have much less impact, will be much more prone to error, and have some problems.

While we recognize when you're given a task order from CMS you have to follow their instructions, there might be some feedback from us that would be useful. Let me go ahead and get further comment first from Jim, then Jeff.

MEMBER DUNCAN: As I said before, personal experience, when I was a radiology resident, my 2-year-old son fell, hit the floor, loud thunk in the other room, screaming. Scoop him up, drive --

Yes, ludicrous levels to the emergency room. Luckily I went to a children's hospital, where they looked at them, calmed me down, recognized that he was calming down, and gave me the list of instructions and sent me home. My

son is now fine. I guess the thought was that I want to thank the people that made a good decision at that point and recognize that they just wouldn't have gotten credit for that good decision in this algorithm. Again, if we're trying to improve best practice, that's, I guess, the tweak that I was thinking.

CO-CHAIR SUSMAN: Thank you. Jeff.

MEMBER SCHIFF: This is both to the committee and to the developer. I think that there's a little bit of a message that I think we need to have the ability to say to CMS, which is along those lines. One is that we need to look at a broader message, as Jim said, that doesn't cut that denominator so small that it becomes not relevant. That's not the developer's fault. It's the task order they were given.

Then the second thing is that measuring this at the health plan, in many states where the health plans overlap, especially of children's institutions, is gigantic seems to be an inappropriate level of accountability. I

think those are two messages that I just wanted -- more for the fact that I think the developer tried hard to get a peg to fit into a certain sized hole.

CO-CHAIR SUSMAN: Yes, I think to make it clear to the developer, the data, the information, the forms were very well filled out, clear. You were given a thankless task, if you will. I'm going to go around the room, get the final comments. John.

CO-CHAIR BROOKEY: Yes, I just want to kind of restate the obvious, but I deal with peer review across all disciplines where I work. I continually challenge our ED physicians and our family practice doctors to provide the same level of care that a pediatrician would provide. Pediatricians do exactly what you just described every single day.

Every single day they're in clinic they have a kid that bumps their head. But if the kid shows up in a pedes clinic, they're going to get a different recommendation, often, than if

they show up in urgent care or family practice or the ED. So the one aspiration that I have is to try to have the same care for the children wherever they go. So for developers, I hope they keep that in mind because that really would be my goal.

CO-CHAIR SUSMAN: A laudable one. Finally?

MEMBER MORROW-GORTON: I wanted to echo what Jeff said about doing this at a health plan level, as opposed to a hospital or provider level. But the other thing I wanted to say was that I think in -- while your numbers ended up being very small, having a concentration on a single reason actually makes it easier to remediate -- and sorry to use CMS language -- to improve, in the end, because you have a single reason -- a single way to address it, rather than having multiple causes all lumped together. So for whatever that's worth, I think there was some value in thinking about it from that vantage point.

CO-CHAIR SUSMAN: Thank you very much. I think these were very useful comments and appreciate the committee's hard work. We're making the executive decision here to take a break. I can see people squirming in their seats. Let's try to be back by 25 after, so that we can continue to get through this large dose of measures. Thank you.

(Whereupon, the above-entitled meeting went off the record at 10:11 a.m. and resumed at 10:22 a.m.)

CO-CHAIR SUSMAN: Okay, we're going to go ahead and get started. The next measure we're going to be considering is transcranial Doppler ultrasonography screening among children with sickle cell anemia. This is another Q-METRIC measure from the University of Michigan, No. 2797. Gary Freed, Julie McCormick, or whoever, are you on the phone?

DR. REEVES: Hi, this is Sarah Reeves. I'm an epidemiologist on the Q-METRIC team. I'll be the developer for this measure.

CO-CHAIR SUSMAN: Thank you, Sarah. Are there any people who asked to recuse themselves? I saw Jonathan Finkelstein.

No, forget that. It's wrong here. Developer, would you please give us an overview of the measure and the information behind it? Thank you.

DR. REEVES: Absolutely, thank you. This measure assesses the percentage of children ages 2 through 15 that have sickle cell anemia who received at least one transcranial Doppler screening within a year. Currently, there are no NQF-endorsed quality measures for children with sickle cell disease, and development of quality measures for sickle cell was determined by CMS, through a public process, to be a priority for the nation.

This measure specifically is supported by strong evidence because without intervention, 11 percent of children with sickle cell anemia will have a stroke by the age of 18. Importantly, these strokes can largely be

prevented. Transcranial Doppler or TCD screening is a non-invasive ultrasound method to identify children that are at high risk of stroke.

TCD measures the blood velocities in the brain and among those with highest risk, receipt of chronic blood transfusion can dramatically reduce the risk of stroke. In a randomized control trial, a 92 percent reduction of risk was observed. Given the fact that TCD is the only method in which to identify which children with sickle cell anemia are at the highest risk of stroke, an expert panel at the National Heart, Lung and Blood Institute, or NHLBI, strongly recommends that all children with sickle cell anemia should receive one TCD screening per year from ages 2 up until 16. Therefore, in concordance with these NHLBI guidelines, our measure uses administrative claims to assess the proportion of children ages 2 through 15 with sickle cell anemia that receive a TCD screen within a year.

As you know, there are many hemoglobin

variations, and our measure focuses specifically on HBFS or sickle cell anemia cases, and this is consistent with the NHLBI recommendations. Our denominator is the number of children with sickle cell anemia identified through the presence of at least three sickle cell anemia claims within a year.

Our denominator definition has both high sensitivity and specificity to identify these children, as compared to the gold standard of newborn screening records. Our numerator is the number of children with sickle cell anemia that receive a TCD screen, and this is identified using administrative claims. Again, we found that this method was valid as it was highly correlated with the TCD screens found within the medical record. In addition to the evidence that our definitions of numerators and denominators were valid, we performed a signal-to-noise analysis using Medicaid data across six different states. We found that the reliability of this measure was extremely high using this method, as

all reliability coefficients were over .96 on a scale of 0 to 1.

We then used the same Medicaid data from the six states to assess the TCD screening rates from 2005 to 2010. Across the period, TCD screening rates ranged from 7 percent to 51 percent, depending on the state, and they did increase over time. However, the average screening rate in 2010 was still only around 38 percent, indicating that there is substantial opportunities for improvement.

In closing, we do feel that this measure is highly important, and it has the potential to make a positive impact on the quality of children's lives. Our measure focuses on TCD screening, which is currently the only method in which to identify children with sickle cell anemia that need to initiate stroke prevention efforts. We found that this measure is highly reliable and valid, and that the data elements to calculate this measure are readily available and administratively straightforward.

Finally, we found that there is an important performance gap, and we strongly believe that endorsement of this measure can have a positive impact on the health of these high-risk children. We very much look forward to addressing any questions or concerns that you may have. Thank you for your time.

CO-CHAIR SUSMAN: Thank you so much. We should record that as an example of a great succinct and complete presentation. With that, let me turn it over to Jim, first, to see if he has some comments or questions, and then maybe some discussion around evidence.

MEMBER DUNCAN: I think there's a very strong level of evidence, given the NHLBI's recommendation and systematic review. There's also, I think, quite feasible, as their data extraction showed the only concern, and they really did the testing to assess the difficulty of identifying children with sickle cell anemia and requiring three separate iterations of the appropriate code in the medical record. I think

the subgroup, as we talked about it, said just like their summary was so succinct and straight on that the measure seems that way.

CO-CHAIR SUSMAN: Very good. Keith, any further comments? Hearing none from Keith, John, do you want to add anything?

CO-CHAIR BROOKEY: No, I think this is one of the measures that seemed to have the strongest evidence and reliability and validity. The only thing that we did talk about -- it's not going to be addressed in this measure -- is that there may be variation from center to center, in terms of the quality of the studies, but that doesn't preclude having this measure go forward because of the importance of the measure and, really, the ease of measuring it and the importance of identifying these children early.

CO-CHAIR SUSMAN: I invite questions from the group. I see Ricardo.

MEMBER WHITE: I'm sorry to interrupt. This is Keith. I was talking before on mute. I apologize.

(Simultaneous speaking.)

CO-CHAIR SUSMAN: It was fabulous comments. We loved your insights.

MEMBER WHITE: They were brief, weren't they? I just was interested again in the methodology used for the transcranial Doppler and the availability of that across the country. Are there any issues or questions with access of skillset and expertise in performing this exam geographically?

CO-CHAIR SUSMAN: Any comments from the panel? John, you were mentioning --

CO-CHAIR BROOKEY: -- whispering over here. Our sickle cell experts, even within Los Angeles, feel there are definitely some centers that they rely on more than others, so I think there is intercenter issues, in terms of -- I think there's some variability that, again, this measure is probably not going to address. But again, as I mentioned before, I don't think it precludes moving forward.

It's just I think that is an issue

that will have to be addressed separately. I was thinking that it's probably going to hopefully coordinate the care of children with the severe sickle cell because the imaging is only part of a therapy. Trying to monitor and treat these children on an island would be extraordinarily difficult.

CO-CHAIR SUSMAN: Okay, Ricardo.

MEMBER QUINONEZ: Just a quick clarification of whether we can vote all four, since I didn't see any mention of grading of the evidence or performance of a comprehensive review, other than the NHLBI guideline.

CO-CHAIR SUSMAN: I think, actually, there was based on a systematic review of the evidence, and then was ranked a strong recommendation with moderate quality evidence. When looked at the guts of that previously, I think it's actually there pretty explicitly. I don't know if our developer wants to comment on that.

DR. REEVES: Absolutely. The NHLBI

guidelines do exist, but also, there was a systematic review of about at least 50 observational studies in two randomized control trials, and the evidence from all of that collectively was rated as moderate quality evidence with a strong recommendation.

CO-CHAIR SUSMAN: Thank you. Is that okay, Ricardo? Okay, Kerri.

MEMBER FEI: I have no doubt that this measure is very appropriate for the Medicaid population. I'd like to see testing in the commercial, to make sure that it would function the same in the commercial health plan population.

CO-CHAIR SUSMAN: Okay, any other questions about evidence? Debbie.

MEMBER FATTORI: Just to add to Kerri's comment, I believe after a relatively short period of time, most patients with sickle cell disease do qualify for Medicaid, so I think the vast majority of those patients do get shifted to that insurance carrier pretty quickly.

DR. REEVES: That's actually true. We found at least 70 percent of children with sickle cell anemia are enrolled in Medicaid.

MEMBER FATTORI: Right, and that's -- if it's going to be listed in the NQF site as a health plan measure if it does get endorsed, it might not be all health plans that are able to implement, use, and get reliable results.

DR. BURSTIN: It's a really interesting point, Kerri. We don't typically make that distinction. I think the key thing would be people would need to consider sample size and appropriateness. Not every measure should be used for everything, but we would say it's appropriate for the health plan, and obviously you'd need to consider whether it's logical. As those measures get selected for use, there'd be a lot of discussion of exactly the point you raised, but not so much on the endorsement side.

CO-CHAIR SUSMAN: Other questions, particularly now, about the evidence? Kevin?

MEMBER SLAVIN: Just in terms of the evidence and the timing of the transcranial Doppler studies and the recommendation, is it once per calendar year, once per year depending on when the last study was done, just some stuff about that, in terms of what the recommendations actually specify?

CO-CHAIR SUSMAN: I'll let the developer respond.

DR. REEVES: The recommendations are specifically for one TCD annually, from ages 2 to 16 years old.

CO-CHAIR BROOKEY: Meaning a rolling 12 months, right?

DR. REEVES: Correct.

CO-CHAIR SUSMAN: Any other clarifications, questions, comments? Okay, let's go on then to vote on evidence. We have 1 for high, 2 for moderate -- oh, I'm taking your job.

MS. CHAVEZ: Thank you. Now voting on Measure 2797 on evidence, 1 high, 2 moderate, 3 low, 4 insufficient, and we are expecting 27

votes. Voting's open, 23, 27.

CO-CHAIR SUSMAN: I'm happy to trade jobs with you for a while.

MS. CHAVEZ: Okay, 18 voted high, 8 voted moderate, 1 voted low, 0 voted insufficient. This measure passes evidence.

CO-CHAIR SUSMAN: As John was saying, it's probably the strongest one we've had thus far. Okay, so let's move on to the gap, opportunity for improvement. I think there's fairly clear documentation. Jim.

MEMBER DUNCAN: Right. I thought the Medicaid data that they submitted, while it isn't current, was showing that there were gaps from state to state, and also a pattern towards improvement.

CO-CHAIR SUSMAN: Okay. Amy, did you --

MEMBER HOUTROW: I just wanted to say it was pretty obvious, given the range of reported percentage of children that received them, that there's a large practice area of

variation that is evidence of the gap. We're not going to find ethnic disparities, and we're not going to find likely socioeconomic disparities, given that a vast majority of these children will be in a Medicaid population.

CO-CHAIR SUSMAN: Okay. Other questions about gap? If not, let's go ahead, 1 high, 2 moderate, 3 low, 4 insufficient.

MS. CHAVEZ: Now voting on gap. Voting is open, 22, 25, 26.

CO-CHAIR SUSMAN: Lauren, we're waiting on your vote, please, if you're there.

MEMBER AGORATUS: Sorry, I did it, but I'll put it through again.

DR. NISHIMI: Okay, we got it.

CO-CHAIR SUSMAN: Okay, so maybe just add the one vote to the total count.

MS. CHAVEZ: So it would be 23 high, 4 moderate, 0 low, 0 insufficient.

CO-CHAIR SUSMAN: Okay, thank you, so passes on gap. Let's move forward to reliability. Any comments on the reliability? I

think there was signal-to-noise testing. There was actual testing at the performance level.

MEMBER DUNCAN: What I saw was that the exam is coded, very reliable. The only concern was identifying the children with sickle cell, and again, they worked through a process to show their reliability.

CO-CHAIR SUSMAN: Any questions? Ricardo, do you have a question, or are you just -- yes, okay, no problem. Then let's go ahead and vote on -- oh, excuse me. I'm sorry.

MEMBER MORROW-GORTON: I just have a procedural question.

CO-CHAIR SUSMAN: Please.

MEMBER MORROW-GORTON: The notes say that the rating may be high or moderate, and those are the only choices, correct?

DR. NISHIMI: I'm sorry. It should be all four. It's eligible for the highest, though, because it's at the performance core.

CO-CHAIR SUSMAN: Vote your conscience here on reliability, and it can be any of the

four ratings.

MS. CHAVEZ: 1 high, 2 moderate, 3 low, 4 insufficient. Voting's open.

CO-CHAIR SUSMAN: Thank you, Vanna.

MS. CHAVEZ: Twenty-four.

CO-CHAIR SUSMAN: We're having the sidebar. Is Vanna still out there? Does anybody know?

MS. CHAVEZ: Twenty-six, 26. Okay, so we are expecting 26 votes. Seventeen voted high, nine voted moderate, zero for low, zero insufficient. This measure passes reliability.

CO-CHAIR SUSMAN: Let's move on to validity. Comments, Keith, Jim, others who took a look at this?

MEMBER WHITE: From the articles cited and the research that's done, it's clearly indicative that there's a valid, indisputable link between the abnormal ultrasound findings and the incidence of stroke, so I'd say it's a very valid measure.

CO-CHAIR SUSMAN: And the validity

testing about case definitions seem to be done, so -- any questions? Okay, if not, let's vote on validity, 1 high, 2 moderate, 3 low, 4 insufficient.

MS. CHAVEZ: Voting is open, again expecting 26 votes, 24, 26. Twenty voted high, six voted moderate, zero low, zero insufficient. This measure passes validity.

CO-CHAIR SUSMAN: Feasibility. Comments on feasibility? This seems like --

MEMBER WHITE: I think with the administrative claims data, it's going to be almost plug and play. It would be very easy to collect the data.

MEMBER DUNCAN: I think their Medicaid data that they collected clearly shows that.

CO-CHAIR SUSMAN: This is a health plan measure, administrative data, documented. Let's go ahead and vote on feasibility, 1 high, 2 moderate, 3 low, 4 insufficient. You almost beat me to it.

MS. CHAVEZ: Voting's open.

CO-CHAIR SUSMAN: This is going to be like -- who's Vanna's partner on the show, since I don't -- it'd be like Pat Sajak spinning the wheel or -- gosh. Yeah, turning the letters over.

MS. CHAVEZ: Okay, we did get our 26 votes. Twenty-four voted high, two voted moderate, zero for low, zero insufficient. This measure passes feasibility.

CO-CHAIR SUSMAN: Usability now. Usability.

MEMBER DUNCAN: The only comment I saw was that it would be -- if the New York State data is available, and also sort of seeing it go forward. I think that's a minor -- it doesn't impact usability.

CO-CHAIR SUSMAN: Any questions on the usability of this measure? Your turn.

MS. CHAVEZ: Okay, voting on usability, 1 high, 2 moderate, 3 low, 4 insufficient. Voting's open, we're looking for 26 votes, 22, 26. Twenty-two voted high, four voted moderate, zero for low, zero insufficient.

This measure passes usability and use.

CO-CHAIR SUSMAN: So we can vote on the overall measure. I think there's been pretty consistent feeling about this one. Is there any final comment? If not --

MS. CHAVEZ: Now voting on Measure 2797 for its overall suitability for endorsement, 1 yes, 2 no. Voting's open.

CO-CHAIR SUSMAN: Can I buy a vowel?

MS. CHAVEZ: Twenty-three, 26. Twenty-six voted yes, 0 no. This measure, 2797 --

CO-CHAIR SUSMAN: Congratulations.

MS. CHAVEZ: -- has been recommended for endorsement.

CO-CHAIR SUSMAN: Thank you to our measure developer.

DR. REEVES: Thank you guys very much.

CO-CHAIR SUSMAN: That's probably the first person who leaves real happy.

(Laughter.)

CO-CHAIR SUSMAN: Including -- yes, John and I are ecstatic over here. Am I doing

this one, or are you doing this one?

(Pause.)

CO-CHAIR SUSMAN: So we're going to move on and do the mental health follow-up measure, Timeliness 1, delayed coordination of care following mental health discharge. This is 2815. It's collaboration for pediatric quality measures, CAPQuaM.

MEMBER WHITE: This is Keith. I'm going to sign off the meeting now. I don't think I'm going to be contributing much to the remainder.

CO-CHAIR SUSMAN: Thank you very much, Keith. We really appreciate your help and expertise.

MEMBER WHITE: Thank you.

CO-CHAIR SUSMAN: Do we have the developers either here in person -- yes, here they are, materializing from the ether.

So, introduce yourselves, and then provide us a brief overview. I like that contraption you have for your glasses. That's

interesting. Good design.

DR. KLEINMAN: It's a wonderful adjunct to aging.

CO-CHAIR SUSMAN: There you go.

(Laughter.)

DR. KLEINMAN: I'm Larry Kleinman. I'm the PI of the Collaboration for Advancing Pediatric Quality Measures, one of the seven CHIPRA centers. I'm joined by Dr. Eyal Shemesh, who was the co-lead with me in developing this measure.

Dr. Shemesh is a pediatrician, a psychiatrist, and a child psychiatrist, so clearly overqualified for most tasks, but ideally qualified for this one. I'm very pleased to be here. First, I actually, having been here for much of yesterday and some of this morning, I want to express my appreciation for the work all of you are doing and acknowledge the challenge that you're facing.

There are seven centers of excellence that emerged from a peer review competition among

many worthy applicants. Some of you may have been among them. We're all presenting, really, new science that is trying to move the field forward, and you're being asked to do an assessment of the standards of that, which I think is an incredibly hard task. I want to share with you the pediatric quality measures program, some of the tasks that we were asked to do, because I think this context will be helpful in understanding some of how we went about our measure development -- was to improve and strengthen children's healthcare quality measures, including expanding on existing pediatric quality measures and advancing their development, with the goal of increasing the portfolio of quality measures available to public and private purchasers of children's healthcare services, providers and consumers.

That's all language from the act that funded us. I want to start with -- our approach began with the IOM definition of quality, which is the degree to which health services for

individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge.

In this view, quality is a continuum, and not a dichotomy of good and bad, which is one of the things that also makes it more challenging in this endorsement context. It also acknowledges that uncertainty exists and we're making decisions in measures that reflect the best of what we know. I'll also note that the IOM created -- or, identified six key attributes or domains of quality, and timeliness is one of them.

The peer review process that we used -- and I want to spend a little time on this; I know we don't have very much time -- because on the call, it was clear that some of you didn't know why this didn't come out of thin air. Our process began with the principle that it's possible to develop excellent quality measures even when there's meaningful uncertainty, and that medicine is a clinical practice for which

the evidence is rarely dispositive for any specific patient.

One corollary to that is there may be more than one good measure. The challenge for you is to identify whether this is one of them. Our 360-degree process starts with an assignment from AHRQ and CMS, so these were topics that the feds thought were important. We did a scoping literature review, started with several thousand articles, leading to 653 articles being abstracted and reviewed and incorporated into our expert process. We had a parent focus group. We had discussions with clinical practitioners, and then we had a RAND-style expert panel that had pediatricians, a family doc, a discharge planner, hospitalist, psychiatrist, child psychiatrist, psychologist, and a patient navigator all on the committee.

We asked the committee to develop explicit criteria to guide the measure development, and from that, we developed something we called a boundary guideline. I'm

not going to go into it, but we had a formal and pre-specified process for doing this. Then our staff worked with our stakeholders and our steering committee to develop and refine the measures.

We had a broad array of stakeholders, and then we would consult the expert panel as necessary and appropriate along the lines of that path. Outside the process -- and this is not in anything you have, but in response to the committee and to NQF's feedback, we sent the measure to the expert panel, as it's specified, and a strong majority of them have said yes, that it is a good measure and that they support its use. We build off the NQF-endorsed HEDIS measure as good stewards of our federal funding. We did not reassess data at the element level, rather used administrative data as the established standard that had been demonstrated in the NCQA submission previously.

The construct validity, as provided by the expert panel process, who assessed the

capacity -- I'm sorry, and then we assessed the capacity of the measure to distinguish between different levels of care, which we established in an iterative set of testing with our partners in New York State Medicaid using a year of New York State Medicaid data. We did this in consultation with our steering committee.

We went through nine rounds, that I could identify, of assessment using a full year of New York State Medicaid data to end up where we are today. There was a question from the committee about why primary care physicians. We found in the New York State data that the absence of primary care physician visits and, independently, the absence of mental health visits, and jointly even more so, was associated with a higher re-admission rate. The smallest of the signal-to-noise ratios measured as a difference over co-efficient of -- I'm sorry, over standard error was four, of those three categories. We also analyzed national data and have shown high levels of comorbidity in

concomitant use of mental health and physical health drugs, also providing a clinical justification and rationale for why both are important.

This is a novel contribution to the measure, one of the specific enhancements, in addition to simply optimizing for a child age range. We sent to the committee, in our written response, the expert panel results regarding timing, use of ICD-9 and other things. I'm happy to get specific if you have questions and go over any of that.

But I want to acknowledge that it's not the only meaningful timing. We were given the opportunity to submit one measure, so we submitted this one. But there was a timeliness measure, there was an absence measure, and there was a delay measure. This is the delay measure, which we chose to submit after consulting with some clinical experts. There was a question specifically regarding follow up the same day. We addressed it in the response. But some of the

reasons that's problematic is it provides no insight as to the status of the patient as an outpatient, so it reflects the inpatient care.

It's a measure that can be easily gained, and in our data, we found that people who had visits on the day of discharge were very unlikely to be followed up with a second visit, even within another 30 days, so it doesn't seem to be establishing, actually, follow-up care. It's just a visit.

That wasn't the construct we were going for, and our panel has endorsed this. In conclusion, we set out to optimize the HEDIS measure for children. We're convinced we've done so in a rigorous, clinically meaningful and evidence-grounded fashion, and I look forward to being a resource to help you answer whatever questions you have. Thank you.

CO-CHAIR SUSMAN: Thank you very much in your responsiveness to our subcommittee's request for information. Let's go ahead, then, and ask our chief discussants here, Jim maybe

first, and then we'll go down the line here.

MEMBER DUNCAN: Going through the algorithm, I think one -- and on the workgroup call, one of the primary concerns is that is there strong evidence showing that this measure is related to an outcome, and a lot of concerns that 30 days -- clearly an arbitrary number. I know it's a little bit different, also, from your expert panel. Really, it's driven by largely the opinion of an expert panel, rather than testing data from the literature.

DR. KLEINMAN: The re-admissions data was actually using the 30-day measure, so that actually -- that's with a full, large state of Medicaid data, which is one of the obvious important consumers of this. There's good evidence that delayed follow up is associated with re-admissions.

We sent you one of the articles in the summary from their abstract about that, that absence of visits hasn't -- might be in there, plus the expert panel did endorse it. I want to

say that they actually did -- the expert panel said 30 days was okay, both a priority and in our assessment of the overall measure. They also would've been fine with 28 days, but in an effort to harmonize with existing data, we were not looking to gild the lily. We were looking to make a practical, useful, helpful measure.

CO-CHAIR SUSMAN: Deb.

MEMBER FATTORI: Just to speak first to the evidence, in the summary that we have that NQF put together for us and part of the discussion that we had in our subgroup meeting, there wasn't a systematic review done for evidence. Is that -- could you describe that a little bit?

DR. KLEINMAN: There was a systematic review. We didn't grade evidence. We felt, as do many, that the grading of evidence into levels of A, B, C, D or otherwise creates false distinctions. The question is what is the power of the information that's there? We let our expert panel -- because we gave them a 40-page

literature review.

It was systematically done, and we outlined some of that in the response. We actually read 653 articles in the literature, so we looked at things like does follow up make a difference, can interventions change follow up? Specifically, we asked is IT a requisite part of that, as part of our review. It was done very systematically, planned, with a research librarian from Columbia University guiding us, and ultimately doing the literature searches. It just -- the question is did we do a systematic review and grade the literature? The answer is we did a systematic review. It was a scoping review.

Unlike other, more typical development processes, we didn't start with a measure and set out to demonstrate that that measure was valid. We set out with a topic and looked to the panel to tell us what the measure ought to be. The literature review actually preceded the development of the measure, and the measure was

based and extracted from the expert interpretation of the literature review. That's why I said we don't quite fit within the lines. We colored outside the lines a little bit.

CO-CHAIR SUSMAN: I think that's a good description of maybe the challenge all of us will face in evaluating this particular measure. Ricardo, and then David.

MEMBER QUINONEZ: Since the 30-day follow up was not really informed by evidence, but mostly decided by an expert panel, what would you say to a mental health clinician who evaluates a patient in the inpatient setting and after their evaluation and treatment, decides that the correct follow up for this patient is six weeks, based on what they see with the patient? They would be penalized for this, and the evidence for the 30 day is not very strong, so do you --

DR. KLEINMAN: I'm going to let Eyal answer the clinical, and then I'll answer from a measurement perspective.

DR. SHEMESH: I think it's just a very, very unlikely scenario. From the clinical perspective, do we really want to capture all the very, very unlikely scenarios in a measure is a question that you should consider. But there is an answer in the data that I think that Larry will describe.

DR. KLEINMAN: I'm not sure I know which one that's going to be.

DR. SHEMESH: I will help you if you need.

DR. KLEINMAN: My answer is to my eye, there are no -- I shouldn't probably say no, but nearly no measures with clinical nuance -- with any level of clinical nuance or substantial levels of clinical nuance -- that are going to be right for every patient. We did not suggest that this measure ought to be used at the level of the individual practitioner, rather at hospital level or system levels because those are the ones who really control both the path and the availability, potentially, of physicians.

I would say it may be the case. In that case, this measure will get it wrong. Every measure gets some of them wrong. But at the population level, the levels of aggregation that we are suggesting this be used at, it's a very powerful measure with a clinically reasonable -- and with content validity established by a panel in a rigorous process. Every dichotomization is, at some level, arbitrary. It could have been 28 days. It could have been 31 days.

DR. SHEMESH: If I may?

DR. KLEINMAN: Please.

DR. SHEMESH: When I speak of data, there is two things that were in -- one of them is we look at the expert panel as data. We actually scored them and so on. In that sense, the expert panel did not at all think that there's anything that's more than a month, than 30 days, that's reasonable. It was to the contrary. From the perspective of where they put the range there, 30 was pretty up high there. Again, as a clinician, as somebody who's a

division chief, I think that the likelihood that 30 days is too little, that you need more than 30 days, is about the same as the likelihood that you don't need to sterilize a surgical instrument.

You may not need to, and you may not develop an infection. That's what our expert panel told us. But the other thing in the data was that this amount of time was correlated with less re-admissions, which you didn't mention. So there was some evidence that it makes sense to put it there, but I agree that you could have put it somewhere else, also.

CO-CHAIR SUSMAN: To be clear on the measure, it's primary care and psychiatric follow up?

DR. KLEINMAN: Yes, it's primary care and mental health, not necessarily psychiatric, with a list of clinicians. Those list of clinicians also came directly from the panel, both the primary care list and the other. I'll note that a behavioral and developmental

pediatrician actually satisfies both criteria, which is not something, personally, I would have chosen, but it's what the panel said, and we listened to the data.

CO-CHAIR SUSMAN: Okay. I'm going to down and around here, so starting with David.

MEMBER EINZIG: I hate to bring this up, but the vertical path or the horizontal path? This is listed as a process measure, but maybe we should be looking at it as an outcome measure that routine follow up does improve global outcome. Out of conscience, I just -- it's the right thing to do.

DR. KLEINMAN: Yeah, and I think that's a reasonable approach. I actually -- again, I'm uncomfortable with some of the categorizations we're asked to make because I think the world is different shades of gray. I think on the horizontal path, on the validity, I can see a very clear path going across, at least to moderate. I could understand the argument for not making it high, although I think it's

actually high. But I think if you think it's an outcome measure, then I'm happy to discuss it in those terms, if that would be better for the panel.

CO-CHAIR SUSMAN: I think it would be probably best if we consider this as a process measure. While, again, there can be some nuance here, for sake of expeditious review, unless people feel strongly otherwise. David two.

MEMBER KELLER: I'm David two, really?

(Laughter.)

CO-CHAIR SUSMAN: T-O-O.

(Laughter.)

MEMBER KELLER: I think not.

CO-CHAIR SUSMAN: Now I'm in trouble.

MEMBER KELLER: First off, I just wanted to echo and support the measure developers in setting the 30 day. It's hard for me to imagine a child who has been hospitalized for mental health reasons who does not require a follow up at least within 30 days.

My question was more about the follow

up with both mental health and primary care for a couple of reasons, one being that there's an increasing -- and also to challenge the comment that the hospitals control the access to the providers. Because that may be true in New York State, but in many places, the hospitals have absolutely no control over access to mental health providers and very little control over access -- a little more, but less control over to primary care providers than you might have thought. The two challenges I'd have to that, is one is that there's a number of systems now that are beginning to integrate behavioral health and primary care in different ways.

I'm referring specifically to the MCPAP program in Massachusetts, where a lot of follow up is being done by -- that's a program where child psychiatrists are available to consult by telephone with primary care providers. For many, given the shortage of mental health services, that's become a safety valve for a lot of those folks, where they can be seen by the

primary care provider, who will then talk to a child psychiatrist or to a mental health counselor while they're working on getting in.

I'm not sure that that is the -- I'm not sure that this measure would be able to capture some of the more innovative systems that are being developed to handle the mental health follow-up problem. I guess I'm wondering why it was an and, and not an or?

DR. KLEINMAN: Happy to share --

MEMBER KELLER: Maybe I misunderstood. I thought it was an and. Is it an or?

DR. KLEINMAN: No, it is an and, although we stratify, so that you can actually -- for purposes of improvement or understanding. Let me start --

CO-CHAIR SUSMAN: It's and that we're considering.

DR. KLEINMAN: It is and. Unequivocally it's and. So let me answer this. I think there are three questions, really. The first question is I said two things quickly that

clearly were able to be heard conflated as one.

Hospitals don't control the accessibility of the physicians, but they do control processes that have been found -- such as follow-up reminders, scheduling before discharge, things like that, which have been found in the literature clearly to be associated with follow-up rates. I'm sorry for not speaking more clearly. The systems obviously handle accessibility, staffing, etc. That's one thing. The second question related to why both? The simple and reductionistic answer is that our panel said it should be both, and our panel said that both should be in person. Telephone wasn't adequate. Now, I would argue that there are probably exceptions and innovations that would be quite good care.

I will also say that our other work, as we were doing this, was on medication reconciliation. Some of the evidence that drove our thinking about this is work I did in New York State -- I'm sorry, in Massachusetts Medicaid a

decade ago that related to failures of coordination and awareness between clinicians and psychiatrists.

I think it takes full attention to do the med rec piece, and that's likely to have some -- and that it's a part of what goes on at follow up, and a critical part of what goes on at follow up, since it's associated with ER visits and hospitalizations, the unsafe medication practices, which we also found are common and, in fact, have a new AHRQ grant to look at that. I think the feeling was face to face, in the current environment is the right way. Remember the IOM definition of based on current professional knowledge. The bleeding edge may have a better way, but it is really the way the world is today. This is important.

CO-CHAIR SUSMAN: Okay, thanks. There seems to be a discrepancy in the documentation in reviewing. One part it says and, one part it seems to say or, but just to be clear, your intent is and, both primary care and mental

health?

DR. KLEINMAN: Yes, I apologize. I will say we originally had it as distinct and combined measures, but in the need to put it as one submission, and in consultation with NQF staff and our clinical experts, this is how we did it. It's supposed to be and, and then broken down so that for purposes of improvement, or if some organizations only cared about one or the other, they could choose to focus where they wanted to.

CO-CHAIR SUSMAN: But the reported measure would be on the and?

DR. KLEINMAN: The top line measure is and. We hope that actually everything gets reported.

CO-CHAIR SUSMAN: Okay, thanks. I'm going down to Kerri, and we'll continue.

MEMBER FEI: Sure. Hopefully this is a quick question. In the denominator, the inclusion of the birth to five-year range, are there lots of newborns that get discharged from

the hospital for mental health issues?

DR. KLEINMAN: Yes.

MEMBER FEI: Okay.

DR. SHEMESH: We had that struggle, too.

MEMBER FEI: I'm asking from a place of probably ignorance, but --

DR. KLEINMAN: No, here's the issue. There were at least two issues. One is neonatal abstinence and the resulting problems actually trail through. They're relatively common in the epidemiological data. They were among the most common single reasons that children of zero to 21 are admitted with a mental health diagnosis, so that's there.

We also felt -- and we pushed our panel on this because we actually initially thought it would be two or six. But what our panel said to us was if a kid is given a mental health diagnosis as an inpatient, a clinician may view that as a license to use a potentially very dangerous and/or inappropriate mental health

drug. So if they carry the diagnosis, we expect the number to be low, but there it transcends from a quality to a safety issue, so that's why we've done it. But it's a great question, and we were skeptical, as we were going through the process ourselves, until we handed it on. Thank you.

CO-CHAIR SUSMAN: Let's move this along. Kevin.

MEMBER SLAVIN: Can you explain the logic or the reasoning behind a patient with multiple discharges, each discharge counting as a next 30 days, rather than there being a rolling 30 day from a subsequent discharge within that 30 days?

DR. KLEINMAN: Yes, practicality. The issue comes in are we going to -- what's the unit of analysis, is it the patient or the discharge? What we did is consistent with the unit of analysis being the discharge. It gets problematic when you're trying to figure out what counts, what doesn't. After talking with our

partners at New York State Medicaid, other places, in terms of the steering committee, this was the ultimate decision that we made. We looked at it a number of different ways. This was part of -- when I said we did those eight iterative or nine iterative rounds of testing -- it was nine ultimately -- this is part of what came out of that.

CO-CHAIR SUSMAN: Okay, thanks. I'm just going to try to move us along, so we can keep -- Carol?

MEMBER STANLEY: I have some questions about denominator statement, one being the 180 days following date of discharge. Because if this is going to be used at a health plan level, why wouldn't you be looking at continuous enrollment before the date of discharge?

Because if we're going to hold health plans accountable for this, we need to look at if they had any continuous enrollment before the admission and discharge. I'm not really understanding why it's truly 180 days following

date of discharge. Then another thing is you have an exclusion for children who are readmitted to any hospital just on the day of discharge, but what if they're -- why wouldn't that be within 30 days if you're looking at the 30 days? Why is it just they have discharged, they would be excluded from this? I have a couple of questions about your denominator and the exclusions and why the 180 days of continuous enrollment after discharge versus a shorter time period?

CO-CHAIR SUSMAN: Okay, good.

DR. KLEINMAN: Okay. The continuous enrollment, the typical justification for continuous enrollment relates to ability to control the outcome if something happens, so emergency department visits or other things. But if they're in the plan at the time of discharge, that discharge should be known by the plan, and the management of post-discharge things should be able to happen. So that's why.

We developed the measure set -- and there's a whole number -- there's a number of

measures that go up to 180 days. Otherwise, it really would have been reasonable to say for 30 days post discharge for this particular measure, but we didn't want -- again, we're trying to be harmonized for when we submit the others, so we said 180 days. The reason you exclude something on the day of discharge is -- because if someone's being transferred to another facility and it's done as not a direct transfer, but a discharge and a re-admission, we're not looking to -- we're looking to exclude those. Now the issue of excluding re-admissions came up a number of times. Actually, I was comfortable with either way.

I think our team was comfortable with either way. What we heard from the purchaser perspective -- and we had the Northeast Business Group on Health and Medicaid, from the purchaser perspective, on our team -- is that they were concerned that this would allow gaming for re-admission. They didn't want something that promoted re-admissions. In reality, when we

looked at the data, we found it made virtually no difference, in terms of the rates.

So it was inconsequential in the data. This was -- we are a collaboration, and this was what our collaborators urged us to do. I actually would have preferred to have done it the exclusions, personally.

CO-CHAIR SUSMAN: Okay, Jon.

MEMBER FINKELSTEIN: I think the and, requiring a follow up with both primary care and mental health professional puts us in a difficult box with respect to evidence, specifically. Because I'm willing to accept, both because of the process and the studies that are presented, that some follow up is related to outcome.

I think when I ask myself the question do I know that follow up with both is different than follow up with one, I understand the expert panel said so, but that, to me, puts me in the insufficient evidence, where we could make an exception. It puts me in that bottom pathway. We could make an exception if we believe the

process and the opinion is the opinion of the field. But they're very different. So on evidence, I just think we've done something very difficult by combining those.

Every time, if a patient is discharged, patients with psychiatric illness, sees their psychiatrist two weeks later and doesn't come to me -- every visit to any medical professional is half a day off of work for the family. It's a big deal. It's not a little deal. It may be very good care for me, as the primary care doc, not to see them.

CO-CHAIR SUSMAN: I just reflect on my own experience in a location where mental health and primary care are closely linked that I don't think that's the bleeding edge anymore. I think that's an increasingly prominent way to deliver high-quality care. The and, for me, is very problematic. Marlene.

MEMBER MILLER: I'm hoping this question is quick, Larry. You can help me. I see an increasing number of children that go from

an inpatient setting to a day hospital program, and that can be a month or two. How does that -- does that count as a provider? Does that count as an inpatient? They're not eligible? When does that clock start running when a day hospital program is the step down from an inpatient?

DR. KLEINMAN: A couple of things. One is it does count as an exception, and I honestly don't remember -- I'd have to look to see if it was in the codes for counting as a re-admission, which is, I believe, what it is, or if it counts as a visit. What's that?

(Off microphone comment.)

DR. KLEINMAN: I think -- okay, let me not say re-admission, under the re-admission exclusion. It's not that it counts as a re-admission, you're right. It is clearly very different. It's a lower level of care. But yes, we don't want those children who go into day hospitals to show up on the negative end of this. It is written to do that. I just would have to look at the specific codes to do that.

I also wanted to say, in response to what Jon said, that we did a focus group. We heard from parents that they felt the primary care visits were often much more helpful than the mental health visits, because the mental health clinicians did not have time, and they wanted primary care visits as a part of it in the focus group.

Also, our evidence showed, in New York State data, distinct and additive improvement of one versus the other. So mental health made a difference on its own, primary care made a difference on its own, and adding the other on made a difference, so both was better than either.

CO-CHAIR SUSMAN: Okay, thank you. I'm going to try to -- let's see, do we have another question down here? Okay, Amy.

MEMBER HOUTROW: To your point, I think day hospitalization is an outpatient -- the way that it's hospital administered is outpatient, so I think that is a question. I

have a concern about whether or not telehealth would count or not count for the mental health providers.

I also have a question related to Jon's point about or versus and. To be a little bit more specific, is there data, beyond what the expert panel says, that follow up at any time frame is beneficial from the primary care perspective, or from the mental health perspective? What is the data besides what the expert panel said for each of those, and how do those differ?

CO-CHAIR SUSMAN: If I gather correctly, there's re-admission data?

DR. KLEINMAN: Re-admission data in New York State Medicaid over the course of a year, population level data.

MEMBER HOUTROW: For primary care doctors or --

DR. KLEINMAN: Yes, for primary care only, mental health only, both is better than either, all three of those.

CO-CHAIR SUSMAN: With regard to that one outcome?

DR. KLEINMAN: Yes, with regard to the outcome of re-admissions, but that's -- yes. Telemedicine was specifically considered by the panel. Again, our process required us to follow the panel's recommendation, and they specifically did. And then, again, they have subsequently re-affirmed that in their endorsement of the measure as specified.

CO-CHAIR SUSMAN: I'm going to cut you off just because we do have some time limits. Jim.

MEMBER DUNCAN: I'm looking at Page 18 of the supplemental materials. It sounds like the expert panel said a telephone call would be appropriate, but the problem is, again, how would that be captured in this, especially if it's an and? Because I would imagine that might be a very common process that one sees, and they have a phone call, and it won't get scored as an and.

DR. KLEINMAN: Which -- I don't have

this numbered, so I'm -- I apologize --

(Simultaneous speaking.)

MEMBER DUNCAN: It's Page 18. It's when you're going through your Delphi panel.

DR. KLEINMAN: What's the top sentence? And I'll find it.

MEMBER DUNCAN: Continuity of care requires an actual timely visit to an appropriate mental health provider. That's scored an 8.89.

DR. KLEINMAN: It says, the timely visit may be on the telephone for a Tier 1 patient. Our entry criteria follow Tier 1 exactly. So that's -- tier 1 was a conversation that wasn't rated, but where there was unanimity where we defined it. It's 2.3, which is very low.

MEMBER DUNCAN: But it's nearly the same as follow up within 30 days will be considered be timely.

DR. KLEINMAN: That's right, 30 days is not considered timely.

MEMBER DUNCAN: But it's not that

different than 21 days.

DR. KLEINMAN: Seven days for a mental health professional, 21 days for a primary care professional is what they consider timely. Remember, we're measuring delay. So anything that's not timely is delayed, so that's how -- the telephone visit was 2.3, which puts it in the lowest tertile, 1 to 3, 4 to 6, 7 to 9 is how we think of these typically. I'm sorry that wasn't clear.

CO-CHAIR SUSMAN: All right. Ricardo, do you have a final comment here?

MEMBER QUINONEZ: Just let me point out another problem with the and that probably hasn't been discussed. Your inclusion in the denominator of 0 to 21 days, and your justification for that, which was some of these were kids with --

DR. KLEINMAN: That's one of the justifications.

MEMBER QUINONEZ: Units, right --

DR. KLEINMAN: Right, one of two --

(Simultaneous speaking.)

MEMBER QUINONEZ: -- the units with -- yes. A lot of pediatricians would challenge that those babies, which there a lot of nowadays, a growing number nowadays -- a lot of pediatricians would challenge that those babies need to be followed up by a mental health professional. Pediatric primary care follow up is more than sufficient for kids with that problem.

DR. SHEMESH: I will answer this because my group does this. We do allow a developmental pediatrician, so I guess that's the answer to your question. The --

(Off microphone comment.)

DR. SHEMESH: Right, but the question now is do we want to have a measure that is about good care, or do we have a measure about what's around? And that Larry can speak to.

CO-CHAIR SUSMAN: I think -- appreciate your response. I'll ask for any final questions before we start to the evidence. Marlene.

MEMBER MILLER: I'm still confused if the day hospital issue is solved or not? We don't want to ding people for doing a day hospital as the right kind of step down. It's very important.

DR. KLEINMAN: Could I stipulate that we can have a conversation afterwards and resolve, if there's a problem we'll clean that up?

(Simultaneous speaking.)

CO-CHAIR SUSMAN: I don't think that's going to fly. I'll look to our --

DR. KLEINMAN: Then I need a moment to look it up. I would need a moment to review it because I just don't remember how we handled it, but I remember we addressed it specifically. I just don't remember how.

DR. NISHIMI: You have seconds to look it up.

DR. KLEINMAN: What's that?

DR. NISHIMI: You have seconds to look it up.

CO-CHAIR SUSMAN: We're about to vote on evidence. Frankly, I don't know that the day hospital might be an important issue for a few of us, but there's probably --

DR. KLEINMAN: As I said, I can assure you they're not dinged for a day hospital. I just don't remember exactly how --

(Simultaneous speaking.)

CO-CHAIR SUSMAN: Okay.

DR. NISHIMI: Let's consider that under the specifications discussion for reliability and validity and vote on evidence now.

CO-CHAIR SUSMAN: I know it seems like we're pushing you. You're correct, we are because we want to go home at 3:00. We're going to vote on evidence. I'll turn it over.

MS. CHAVEZ: Now voting on evidence for Measure 2815, 1 high, 2 moderate, 3 low, 4 insufficient. Voting's open, and we are expecting 26 votes, 0 voted high, 8 voted moderate, 10 voted low, 8 voted insufficient, and

this measure does not pass evidence.

CO-CHAIR SUSMAN: Okay, feedback to the developer. I know that you've worked very hard and put a lot of effort into this. I think it's useful to sort of have a debrief. Jon?

MEMBER FINKELSTEIN: I wonder, because our process that I learned about yesterday was that changes can be made and these can be brought back quickly, even to our post-meeting call. I don't know if the chairs -- I wonder how much this is hinging on this and problem and if you want to take one of our straw votes or get the sense of the group if this wasn't the and, would this have been the result?

CO-CHAIR SUSMAN: No, I think that's a perfect --

DR. KLEINMAN: We wouldn't submit it that way, just to -- part of the real importance of this measure is that it addresses the issues both of continuity and of coordination of care.

CO-CHAIR BROOKEY: I think the committee members were giving input.

DR. KLEINMAN: Okay, I appreciate that, but --

CO-CHAIR BROOKEY: I think --

DR. KLEINMAN: Thank you.

CO-CHAIR BROOKEY: I think we can just say that would be some of the input of the committee. I think Jon's request was how many people feel like that is the sticking point with this particular measure? Is that what your question is, Jon?

MEMBER FINKELSTEIN: Yes.

CO-CHAIR SUSMAN: But if it isn't going to sway you one way or the other, then there's really no point.

(Simultaneous speaking.)

DR. KLEINMAN: -- information, but this is what we were told to do, and to follow the process requires that at this point.

CO-CHAIR SUSMAN: Okay. Any other feedback to the measure developer? Clearly just making sure about things like the specification, day hospital, other follow up? Jim.

MEMBER DUNCAN: I think this is a classic chicken and the egg that it's very difficult for you to develop a measure without data, and you don't get their data without having a measure. I think lots of folks fall into the same boat.

DR. KLEINMAN: This is what we were asked to do by AHRQ and CMS, so this is where I think the challenge is to the field, to figure out how to recognize and move measures forward with that.

CO-CHAIR SUSMAN: As you were here earlier today, you recognize that sometimes you're in a box by CMS or AHRQ or whoever is the person paying the bills, and we may or may not agree with how that is framed. It is what it is. David, final comment?

MEMBER KELLER: Sure. The thing I was wondering about -- because I was a little surprised how this played out. I thought we were going to go with insufficient and exception which, again, would have been a way for us to

move forward. I guess I'm wondering what criteria would have -- or whether there were any criteria that would have swayed committee members? One of the problems with our algorithm is that if you go with low -- I'm so sorry.

CO-CHAIR SUSMAN: We were having a sidebar. We can do that. If you believe, as a committee, that we want to vote insufficient, or an exception with insufficient evidence, we can do that, and we can have that vote now and see where it falls.

MEMBER KELLER: Oh, so --

(Simultaneous speaking.)

CO-CHAIR SUSMAN: I'm sorry?

MEMBER KELLER: That's what I was wondering. The criteria -- again, this is a procedural question for NQF is the criteria to be able to do the exception are what? What do we need to have up there on the board to be thinking about that path forward?

DR. BURSTIN: Somebody like you requests it, and your group agrees to consider

it.

MEMBER KELLER: Oh, okay, so it's not about having a majority of people say insufficient?

DR. BURSTIN: No, it's not. It's your choice --

(Simultaneous speaking.)

MEMBER KELLER: Okay, that's where --

DR. BURSTIN: It just has to fail.

MEMBER KELLER: I'm sorry?

DR. BURSTIN: It has to fail.

MEMBER KELLER: It has to fail --

(Simultaneous speaking.)

CO-CHAIR SUSMAN: So it's failed. Now I hear you asking the question of the committee.

MEMBER KELLER: My question would be whether the committee -- I do think -- I'll just put out there I do think that this is an important issue.

I think that we have some -- I think there are -- I think we, as a committee, have some issues with the methodologies chosen and

with the specifics -- with the details of the measure, but that conceptually, the notion that these children require follow up and that's a reasonable measure for this small, but really significant group of children is important to consider. I'm asking my fellow committee members whether we should consider the exception in this case? I'm open to -- I'm interested in doing that, but I'm open to discussion.

CO-CHAIR SUSMAN: Virginia.

MEMBER MOYER: I agree that this is a really important topic and a way needs to be found to do this. I don't think we will get to a pass on this. I certainly, having reviewed the measure, I think there are so many other issues with it that we would spend a lot of time getting to nowhere. I think that we'd be better off, having given feedback to the developers, to allow this to come back another time.

CO-CHAIR SUSMAN: I'm seeing a lot of heads nodding yes at the table. Let's just take a straw vote. How many want to reconsider this

for an exception, just raise your hands? We have two or three. It does not look like there's great appetite for that. I appreciate, though, your bringing it up, and remember that as an option for this group. Thank you very much, and we look forward to new measures. I think we're at the point where we can take our break. We need public comments and NQF member comments, and whoever has the keys to the operator, I'll turn it over to you.

DR. NISHIMI: Operator, can you see if there's any public comment?

OPERATOR: At this time, if you would like to make a public comment, please press Star 1.

DR. NISHIMI: Is there anyone --

OPERATOR: There are no public comments.

DR. NISHIMI: Anyone in the audience wish to make a public comment? Okay.

CO-CHAIR SUSMAN: We're about a half hour ahead, which is a good thing. There's no

clue about how long future measures will take, so we have 11:30 to noon for lunch. If you have any questions in the meantime, please come up. I think John's going to take over after lunch. Thank you all.

(Whereupon, the above-entitled meeting went off the record at 11:31 a.m. and resumed at 11:32 a.m.)

CO-CHAIR BROOKEY: We're going to make a transition to --

CO-CHAIR SUSMAN: Bring us home, John.

DR. BURSTIN: The developer has requested, John, that we do the tobacco measure first, and then the two antipsychotics together -- three antipsychotics together.

CO-CHAIR BROOKEY: So we're going to start from the end and work backwards. We're going to do 2803. We have one recusal. Any other recusals? Okay. Do we have the developer here or on the line? Oh, okay, step forward.

CO-CHAIR SUSMAN: Come on up, developers.

CO-CHAIR BROOKEY: Good morning. Can you introduce yourselves, and then give us a brief overview of your measure, please?

MS. BYRON: Yes. Hi, I'm Sepheen Byron. I'm an assistant vice president at the National Committee for Quality Assurance. NCQA is part of the National Collaborative for Innovation in Quality Measurement, or NCINQ Center of Excellence. I have with me here Mary Barton, and on the line, I think we should have Sarah Hudson Scholle, who is the PI of this center, but is actually at the CMS Medicaid/Medicare quality meeting right now. Sara, are you on the line?

DR. SCHOLLE: I am. Good morning, everyone.

MS. BYRON: Great, thanks. We're going to talk about this tobacco measure first. We have three antipsychotic measures coming up, but it felt it made sense to batch them together. This is the tobacco use and help with quitting among adolescents measure.

It is a clinician level measure, and it was tested at the clinician level in electronic healthcare records. With this measure, we are hoping to standardize the way tobacco use is documented. It looks to see if adolescents were screened for tobacco use and, if they are users, whether they received help with quitting.

We think tobacco is a very important topic. It's really something that I think many consider almost like a vital sign. Tobacco use is -- it's slightly decreasing among adolescents, which is good news, but it's still being used. The evidence has shown that a physician's advice to quit can be very effective. It's based on a guideline from the U.S. Public Health Service, as well as a more recent review by the U.S. Preventive Services Task Force.

CO-CHAIR BROOKEY: Before we ask for member comments or questions, can you just clarify, this is two measures, or is it sort of a composite measure? Describe just a little bit,

in more detail, the specs.

MS. BYRON: It is one measure. I know there was some discussion about the way it was structured in the workgroup, but the denominator is all adolescents with a visit, and then the numerator is those who were screened for tobacco use and if they're found a user, did they receive help with quitting, so it's all one numerator, so you get one rate.

Basically you're getting credit for having non-smokers, or treating your smokers. The structure of it really, I think, aligns with the way we have seen other measures done, such as depression screening and follow up. So screen for depression, if you find a problem, is a follow-up plan in place? That's an NQF measure. We were encouraged to align with that sort of structure, so that you have this population-based measure.

CO-CHAIR SUSMAN: I should know this, but are you counting counseling, V codes for counseling? Are you counting medication?

MS. BYRON: Yes.

CO-CHAIR BROOKEY: Tell me what counts for these.

MS. BYRON: Yes, it's actually pretty broad. You can look in the medical record, as well. We include counseling. We include referral to services, treatment services, medication. We've aligned it to the way we specify these sorts of counseling measures in other measures in HEDIS.

CO-CHAIR BROOKEY: Just one more question because I know health plans can game this. After visit summaries can include this for everybody. Is that going to count?

MS. BYRON: After visit summaries?

CO-CHAIR BROOKEY: Yes.

MS. BYRON: It should count.

CO-CHAIR BROOKEY: Okay. Because you can put that instruction, sort of 800 numbers, in everybody's after-visit summary, so I just want to make sure that you've considered that for the measure.

MS. BYRON: We did. It's at the clinician level. We wanted to be broad, in terms of what counted as treatment. Our expert panel felt that --

CO-CHAIR BROOKEY: That's been the issue with the Joint Commission measure because you can put the instructions in 100 percent of discharges and get credit for it, so I just want to make sure that you considered that for this measure, because that's how some people will manage it. I just want to call that out. With my biased comments -- yes, go ahead, Kevin.

MEMBER SLAVIN: Just for clarification, the discussion has been about tobacco use, but a lot of this says smokers, so are we talking about all tobacco use, or specifically just smoking?

MS. BYRON: All tobacco use. We specifically used the word tobacco use versus smoking in most of the specification, I believe.

CO-CHAIR BROOKEY: Good question.

MEMBER SLAVIN: Actually, no. The

specifications say specifically adolescents who are not smokers, or adolescents who are smokers but are receiving cessation counseling. That's why I wanted the clarification. So we are talking about all tobacco use, not just smokers?

MS. BYRON: Deeper in the specifications we do say tobacco use, but thank you for pointing that out.

MEMBER AGORATUS: This is Lauren. Are we also including the e-cigs?

MS. BYRON: So that's as good question. We developed this measure several years ago, when it was not as big of an issue. It doesn't specifically say e-cigarettes anywhere in here.

CO-CHAIR BROOKEY: That would be a no then. We have our experts here from the committee would be Maureen, Carol and Sue. Which of you want to kick off and give some comments? Can we first go to our members, and then we'll ask further questions? I want them to give a summary, if they could -- our members who were on

the subcommittee, that is. Don't all volunteer at once.

CO-CHAIR SUSMAN: You go. You go.

MEMBER STANLEY: Our discussion with our group brought up some of the issues you talked about, specifically the numerator statement including both non-smokers -- or non-tobacco users and tobacco users. The evidence was strong from the standpoint of obviously tobacco use is very harmful, and adolescents should quit.

The evidence of physician recommendation or intervention seems to focus on adults, so we wanted to hear a little bit more about the evidence for physician recommendation or intervention specifically with adolescents. We were also interested in hearing how things went with your HEDIS 2015 results in using this specific measure. Then there were some questions about the disparity between your results using the electronic record versus medical record abstraction.

CO-CHAIR SUSMAN: Would you like to respond?

MS. BYRON: First easy one. This measure's not actually a HEDIS health plan measure. It's specified at the clinician level, so it's not in HEDIS. In terms of the evidence, it is based on a U.S. Public Health Service recommendation that says clinicians should screen and if they find tobacco use, they should be treating it. Then there are also numerous studies out there that have found that a physician's advice to quit, even brief, has been effective.

Then also, the U.S. Preventive Services Task Force guideline, which is a little more recent, that actually came out, was in process as we were developing the measure, focused in more on the physician providing interventions to prevent tobacco use, which is another component. Taken together, we felt the evidence really supported this, and our expert panels and public comment results also felt that

the evidence was strong.

CO-CHAIR BROOKEY: For the age group, correct?

MS. BYRON: Yes, for adolescents.

CO-CHAIR SUSMAN: Because that was your question, right, Carol?

MS. BYRON: Yes, and those guidelines are specific to adolescents, as well. Was there one -- I think there may have been one more question.

CO-CHAIR SUSMAN: Just to clarify, I'm looking at the U.S. Preventive Services Task Force website. It's a B recommendation.

CO-CHAIR BROOKEY: I think David had his hand up.

MEMBER EINZIG: Is counseling defined? In other words, just saying smoking is bad for you, it causes cancer, does that count as -- as opposed to more gold standard motivational interviewing strategies, digging deeper, what do you like about smoking? I wonder if there's a definition of what is counseling?

MS. BYRON: Right. The specification actually lists advice given to quit smoking or tobacco use counts, counseling on the benefits of quitting smoking or tobacco use, we give some examples, assistance with or referral to an external program, or current enrollment in a tobacco treatment program.

I do recognize that it may be difficult to define counseling, but because the studies have shown that even brief interventions by doctors giving advice to quit, we did allow that to count. What we want to see is that if you find a smoker, you're at least doing something to try to get them to stop.

CO-CHAIR BROOKEY: Just to my earlier point, you can easily create your smart sets to have a counseling code for 100 percent of your patients, whether it be screen time, exercise, nutrition and smoking. I just want to call that out. This is very important.

I just want to call out the issue with V codes and how they may not really relate to

what actually happened in the office. I think it's getting to your point a little bit that it's a pretty broad definition of what counseling is. If you code it, you've done it, and it gets counted. Marlene.

MEMBER MILLER: I was just going to echo that same thing. There's so many measures like adult care and they're gamed with a pre-templated sentence in the discharge summary, and everyone's at 100 percent, not to mention the fact that tobacco screening is a meaningful use measure, so everyone's already building their EHR systems to automatically do this. I just really worry it doesn't get at what we want, and that it essentially will be, very soon, a measure at 100 percent. That may not be what we have achieved, but it's going to be a documentation, make your smart set do it type of thing.

CO-CHAIR BROOKEY: Jim.

MEMBER BOST: Does the denominator come from the claims data, but the numerator requires chart review? If there are multiple

visits in a given year, does the cessation need to be indicated in every visit?

MS. BYRON: This is a medical record measure. It was tested in electronic health records, but I wouldn't call it an e-measure with a capital E. It's probably like an e-measure with a lower case E, which is why we categorized it medical record.

You would look to see adolescents at the physician level, all of your adolescents who had a visit, and they would be included in your denominator at that point in time, in the measurement year. Then you would look back to see if there was documentation of screening and all the other numerator components. Once you do that measure in that measurement year, you're counting them once.

Adolescents with a visit, children. Think of it as children, right, but at the clinician level, in order to assign accountability, we also look for a visit.

CO-CHAIR BROOKEY: To clarify, is it

a hybrid measure, or is it a chart review measure?

MS. BYRON: It's really a chart review measure.

CO-CHAIR BROOKEY: So you're not basing it on codes at all?

MS. BYRON: You can look at codes in charts, but it's not a claims-based measure.

CO-CHAIR BROOKEY: Just to clarify Marlene and my concern about using V codes, is that going to count for the numerator? That's the reason we're bringing it up. Because what Tim is asking, are we opening up charts and looking for actual documentation, or if it's in the diagnosis code on an electronic chart, will that count?

MS. BYRON: Sara, do you know the answer to that question? I believe we are looking.

DR. SCHOLLE: I think the answer is yes. I think I understand the concern that what we want is an activity to document tobacco use

and if there is tobacco use, to provide some help or some response. This is aligned with an existing measure for adults that basically has the same construction.

The main difference is that this does not include prescribing of the patch, I think, because that's not recommended for adolescents. It's looking for some activity to happen at the visit where the tobacco use was identified. In an electronic record, that could be documented in the notes -- I mean in a medical record, whether it's electronic or paper, that could be documented in the notes.

It could be documented in the diagnosis code or procedure code if they're relevant, or it could be documented -- what's in the visit summary, as you say, that is part of the notes. We are trusting that clinicians who include something on their smart set are actually reviewing that information with the families and adolescents as they're handing out that visit summary.

CO-CHAIR BROOKEY: Yes.

DR. SCHOLLE: I understand there's a concern that it's just going to be automatically implemented.

CO-CHAIR BROOKEY: Right, you answered the question. We want to know whether or not it's filled out in a questionnaire -- the nurse fills out the questionnaire -- whether it's documented in a diagnosis code, whether counseling is a diagnosis code, and whether, in the after-visit summary instructions are included, whether it would count. From what you just said, the answer is yes. I just want to be clear about that.

DR. SCHOLLE: Yes, right.

CO-CHAIR BROOKEY: Okay, Jon?

MEMBER FINKELSTEIN: I was on the subgroup, too. I think the issue we're grappling with is the big one, but it's the big one with any counseling measure. We only have the chart. I can just lie in the chart. I can say I talked to you about X, Y, and Z, and if I didn't, I

didn't. We can do that with coding, as well. I would hate for us to not put forward quality measures because we don't have dash cams on the clinician's forehead yet to prove what was actually said or how good the counseling was. I think this is -- we're talking about evidence right now. The evidence, I think, for this is incredibly strong, especially in the domain of counseling interventions. Of all the counseling we do, there's a lot of evidence around this in an area where there often isn't.

I think what we're talking about that everybody could game it and get to 100 percent, that gets to whether there's currently a gap. I think if this was put out as a quality measure and five years from now, everybody had engineered things to be at 100 percent, it's no longer a useful measure, but I don't think we're there yet. I think we'll hear about that in gap.

CO-CHAIR BROOKEY: David, go ahead.

MEMBER KELLER: Echoing that just a little bit, but also pointing out that the fact

that the measure developers have mirrored this on an adult measure makes this an incredibly practical measure because large organizations that do both adults and children have experienced pulling this measure in their adult population, and it's relatively -- from a feasibility standpoint and a usability standpoint, it's wizard to be able to take what you're already doing and match it.

The other thing is that for -- in our state innovation model plan, the fact that this would become a measure would allow us to extend the measure that we put in for adults down into the adolescent population, which we currently can't do because we decided to use only NQF-approved measures for that.

Again, I think it takes work that's going on and is very well documented in the adult world and allows us to move that into the adolescent world in a really good way. I had exactly the same concern in the adult measure that there's no assessment of the quality of the

counseling that's given or the type of counseling or exactly what that is.

But again, we're taking the temperature. We're not doing an X-ray. We're just trying to see if it happens, and that's often useful for moving the mark. I'd just put out there that I'm very, very excited about this.

CO-CHAIR BROOKEY: Sounds like we're converging on our opinions. Kevin.

MEMBER SLAVIN: Just in terms of the practicalities of putting a blanket statement about tobacco use in an after-visit summary for an adolescent, is that something that you want in an after-visit summary for an adolescent? Isn't that something that we're supposed to sort of be doing in private, without the parents present? If that's the case, that kind of thing can't be the blanket way that an institution or a practice decides to solve this.

CO-CHAIR BROOKEY: I think we're getting close to agreeing that this is probably strong evidence for this. I just want to call

the issue that screening, in and of itself, is important, but you can do two at the same time with templated documentation.

If you screen and don't do the follow up, which may be very inadequate, you may still get credit for it if you automatically import it into your note. After-visit summaries, to your point, I think it depends on the specificity of it. If you put your child smokes on the after-visit summary, that's an issue. If you have a general statement about smoking, then I think it's not an issue. Same thing with STDs and pregnancy and everything else. There are certain things we would agree not to put in an after-visit summary. I think everybody around the room nodding their heads probably feels this is an important measure. Am I getting a sense? Are we getting close to voting on the evidence part of this measure?

MEMBER AGORATUS: This is Lauren. I have a question.

CO-CHAIR BROOKEY: Go ahead, Lauren.

MEMBER AGORATUS: Just looking at the evidence also, just doing a quick survey, CDC, HHS websites and so forth. The prevalence of e-cigarettes are actually surpassing smoking. I don't know if there's a way to capture that, in terms of adding e-cigs, because of the nicotine piece, to this. The percentage is much higher now, with the new generation not actually smoking.

CO-CHAIR BROOKEY: Your point is very well taken. We're going to vote on this measure the way it's specified, but we can provide that as input to the developer for future. Is that all right?

MEMBER AGORATUS: Sure.

CO-CHAIR BROOKEY: We don't want to lose your comment, for sure.

CO-CHAIR SUSMAN: I think that's important. I don't see the strength of evidence behind that that we have for actual tobacco use, as formally defined.

CO-CHAIR BROOKEY: Okay, are we ready

to vote?

MS. CHAVEZ: Okay, now voting on evidence for Measure 2803, 1 for high, 2 moderate, 3 low, 4 insufficient. Voting is open, we are expecting 24 votes, 23, 24, 11 voted high, 12 voted moderate, 0 voted low, 1 voted insufficient. This measure passes evidence.

CO-CHAIR BROOKEY: Okay, we have another decision now. I think we might be able to get through the rest of these fairly quickly, but I also have no problem breaking now and coming back to finish up. How many would like to try to finish before lunch?

PARTICIPANT: Just tobacco.

CO-CHAIR BROOKEY: Just tobacco. How many want to break, and then come back?

CO-CHAIR SUSMAN: Break and come back?

CO-CHAIR BROOKEY: So can we keep rolling then? Okay, so reliability.

PARTICIPANT: No, gap.

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

MEMBER KONEK: There was pretty good

evidence that there was a gap. It was looked at in a variety of ways. It was looked at the National Youth Tobacco Survey, which was in 2011, with only 32 percent of the adolescents reported being asked about tobacco use. Several other things were related to that. Looking at respondents with commercial insurance who received help was at 82 percent, while Medicaid was 60, and other insurances were 39, so it did seem to be a gap in a variety of ways.

CO-CHAIR BROOKEY: Is there any disagreement with that? Any other comments about gap? Could we vote for gap? Okay.

MS. CHAVEZ: Okay, now voting on gap, 1 high, 2 moderate, 3 low, 4 insufficient. Polling is open, 15, 23, we're looking for 24 votes, thank you, 12 voted high, 11 voted moderate, 0 for low, 1 insufficient. This measure passes gap.

CO-CHAIR BROOKEY: Okay, reliability, comments from our subcommittee? Jon, you have a comment?

MEMBER FINKELSTEIN: Just that the reliability -- the developers told us that the variability they found was based largely on the availability of data elements in different systems. I don't know if they want to comment on that. The issue was already raised, and the developers raised it, that it's not ready as an EHR automated data. This has to be digging through the chart because when you just look in an automated way, you don't find stuff. That impacts the reliability a little bit.

CO-CHAIR BROOKEY: Do you want to comment?

MS. BYRON: Yes, you're correct. It's not an e-measure. We still see that these things are not available in standardized fields that you can automatically extract from. You do need to go into notes and that sort of thing. We're hoping that EHRs will catch up with meaningful use, specifying how tobacco use should be defined, and we've aligned to that. We're hoping that eventually, we'll get there.

DR. SCHOLLE: Just to clarify, our testing was done several years ago, so that's before -- so first of all, meaningful use has progressed somewhat. As others on the panel noted, the adult version of this measure has been in reporting, so we really expect that if we did this today, we might be better uptake in the electronic health record setting. There can be measures -- our testing found issues also where the fields existed, but they weren't being used. Until the measure's really deployed, there's not really an incentive to use those structured fields.

CO-CHAIR BROOKEY: It might be a little bit more about feasibility than reliability. You're opening the record. There's a fair amount of reliability, but it's not that feasible, would you agree? So any other comments about reliability? There was testing done, right? Yes.

DR. SCHOLLE: Right, our inter-rater reliability and our ability to pull the data out

of the record was good between different raters.

CO-CHAIR BROOKEY: It's just high effort to get the data that's the issue. Reliability, are we ready to vote on reliability?

DR. NISHIMI: Just a reminder this is at the data element level, so not the performance scores. The eligible ratings are moderate, low, insufficient.

CO-CHAIR BROOKEY: Okay, ready to vote?

MS. CHAVEZ: Now voting on reliability, 1 high, 2 moderate, 3 low, 4 insufficient. Polling is open, 20, 23, 24, 0 voted high, 20 voted moderate, 4 voted low, 0 insufficient. This measure passes reliability.

CO-CHAIR BROOKEY: Very good. Let's go to validity. Any comments about validity from our subcommittee?

MEMBER KONEK: There was face validity discussed, assessed by the developers by convening a number of stakeholders, lots of people that you would expect on that group. They

supported this.

CO-CHAIR BROOKEY: So there was this just face validity, so that would be --

MEMBER KONEK: There was face validity. Validity testing involved the study of 400 patients, as well, who had well-care visits, who were at 58 percent documented tobacco use and received help to stop smoking. That was also reported. It looked like, by the algorithm, it would be moderate.

(Simultaneous speaking.)

CO-CHAIR BROOKEY: Moderate or below.

DR. NISHIMI: They did do empirical testing at the performance score level, however, so it is eligible for high.

CO-CHAIR BROOKEY: Eligible for high. Other comments about validity? Ready to vote for validity? Okay.

MS. CHAVEZ: Now voting on validity for Measure 2803, 1 high, 2 moderate, 3 low, 4 insufficient. Polling is open, 10, 21, 23.

MS. ALLEN: Lauren, I'm waiting on

your vote.

MEMBER AGORATUS: Yes, my computer just crashed. I'm waiting for the screen to come back up. Sorry.

CO-CHAIR SUSMAN: You could just tell us.

PARTICIPANT: You don't have to.

MS. ALLEN: You can email us. No, I'm sorry, you can't.

PARTICIPANT: Her computer crashed.

MS. ALLEN: Sorry.

MEMBER AGORATUS: I can't believe this just went -- the whole thing just went down, and I can't get it back. Here we go. This is the second time, too. I don't know why it's doing this.

CO-CHAIR SUSMAN: You could text.

PARTICIPANT: We can give you a phone number to text.

MS. ALLEN: I can give you my phone number to text your response. It's 240-320-4744.

PARTICIPANT: Expect a lot of phone

calls tonight.

MEMBER AGORATUS: Okay, I'll see if I can do this. I've got to get on the other phone now.

PARTICIPANT: Do we need to wait for her to vote?

MEMBER AGORATUS: I'm sorry, got to love technology.

CO-CHAIR BROOKEY: We're going to go ahead and move forward --

PARTICIPANT: And we'll add your vote in.

CO-CHAIR BROOKEY: -- and we'll add the vote in later.

MEMBER AGORATUS: Thank you.

CO-CHAIR BROOKEY: We won't tell you how it all --

MS. CHAVEZ: Okay, for the 23 votes we received, 2 voted high, 19 voted moderate, 2 voted low, 0 insufficient, and this passes validity.

CO-CHAIR BROOKEY: I wasn't going to

tell Lauren that because we just influenced her vote. I'm sorry. Let's just move forward to feasibility. I think we just stated the issues with feasibility, in terms of it not being an e-measures, other issues with the use of smart text, smart sets, and the issue with being a chart review measure. Are there any other comments about feasibility?

Okay, Sue, go ahead.

MEMBER KONEK: It is noted that it is -- of course, the CMS meaningful use is now in place. That's been already discussed. But it's --

CO-CHAIR BROOKEY: I'm sorry, can you use your -- is your mic on?

MEMBER KONEK: Oh, yes.

CO-CHAIR BROOKEY: I just can't hear you.

MEMBER KONEK: Physician quality reporting system, as of 2/2015, the data elements can be generated through that.

CO-CHAIR BROOKEY: Any other comments

about feasibility? Can we vote on feasibility?

MS. CHAVEZ: Okay, Lauren will we be able to get your vote now?

MEMBER AGORATUS: No, I'm going to send comments, and I'll put what it's for, okay? I'm going to try a different browser.

MS. CHAVEZ: Okay, now voting on feasibility, 1 high, 2 moderate, 3 low, 4 insufficient. Polling is open, 23, okay, 1 voted high, 18 voted moderate, 4 voted low, 0 insufficient. This measure passes feasibility.

CO-CHAIR BROOKEY: Okay, the next is usability and use. Any comments about that? Subcommittee?

MEMBER KONEK: It's in the EHR incentive program, meaningful use, and also the PQRS, now it's already in use. There were no real comments about use from the group otherwise.

CO-CHAIR BROOKEY: Okay. Any other comments about usability? Can we vote?

MS. CHAVEZ: Okay, now voting on usability, 1 high, 2 moderate, 3 low, 4

insufficient. Polling is open, 18, 23.

MEMBER AGORATUS: I'm still trying to get on, so just add mine again later.

MS. CHAVEZ: Okay, thank you, Lauren.

PARTICIPANT: It could be the deciding vote.

MS. CHAVEZ: 5 voted high, 15 voted moderate, 3 voted low, 0 insufficient. This measure passes usability and use.

CO-CHAIR BROOKEY: Okay, so the last is overall. Any comments before we vote on overall suitability for endorsement? Can we vote?

MS. CHAVEZ: Okay, now voting on Measure 2803, overall suitability for endorsement, 1 yes, 2 no. Open, 23.

MEMBER AGORATUS: Again --

(Simultaneous speaking.)

MS. CHAVEZ: Lauren, you'll send us your vote later?

MEMBER AGORATUS: I'm writing them all down, and I'll send it all in the comments. I'm

still trying -- Mozilla's not working. I'm trying Internet Explorer now.

MS. CHAVEZ: Okay, so if we had 24 votes, we needed 15 votes for this to pass. We have 20 yeses, 3 nos. This Measure, 2803, is recommended for endorsement.

CO-CHAIR BROOKEY: Thank you developers. You've heard the input. There was input about e-smoke, which you can follow up on, may or may not belong here, but at least you can follow up on it, questions just about the sources of information. I think the after-visit summary was one of the concerns. Were there any other concerns people wanted to make sure the developers took back with them? Go ahead, Carol.

MEMBER STANLEY: I think this is a really important measure. We have a lot of spit tobacco users in Virginia, so please be sure -- and across the country, I'm sure -- to be sure it's tobacco use and not just smoking.

CO-CHAIR BROOKEY: Sue? And David?

(Simultaneous speaking.)

CO-CHAIR SUSMAN: I'm not sure that you had it, but it would be really interesting to look at disparities as you accrue more data. My understanding is you didn't quite have enough sample to really look at some of those issues, but I think it's going to be real important.

CO-CHAIR BROOKEY: Including what payer type, yes. We want to be mindful of people's time, so I would recommend that people grab your lunch, and when I see that everybody's sort of back at their seats, in maybe 15 minutes, we can resume. Would that work for people? Okay.

(Whereupon, the above-entitled meeting went off the record at 12:08 p.m. and resumed at 12:23 p.m.)

CO-CHAIR BROOKEY: Looks like the majority of people are back at their seats, so we're going to start going. We were just talking about these psychiatry measures. We're going to go on to -- I believe the first one is the -- I believe it's the multiple concurrent

antipsychotics, is that correct? I'm sorry; I'm mixing up my notes here. Who is the subcommittee for this particular measure? Okay.

MS. BYRON: Hello, everyone, it's me again. The next three measures, actually, are part of a set, so I'm going to tell you about all of them. These are a set of measures that we developed to assess the safe and judicious use of antipsychotic medications in children and adolescents. They're based on the problem that we're seeing that antipsychotic medications are being used increasingly more frequently, particularly among children in Medicaid, children in foster care. While they have been approved for a limited, narrow amount of conditions, such as bipolar disorder, we do see a lot of prescribing patterns that are problematic, such as too many medications, too long of duration, poor management, off-label use. There was actually a very high state need for these measures.

There has even been a recent Office of

the Inspector General report from Health and Human Services that highlighted problematic prescribing for Medicaid kids. Given the medications' effect on long-term health, it leads to hyperlipidemia, diabetes, weight gain, metabolic disturbances. These are all long-term health consequences.

We developed these measures to really encourage a think before you prescribe, and if you are using these, proper management of these kids who are on these powerful medications. They're developed and tested at both the state and the health plan level. We actually started with seven measure concepts and put them through the whole process. Three rose to the top as being prioritized for the HEDIS health plan set, and so those are part of the HEDIS health plan set. Consequently, they've gone through multiple multi-stakeholder advisory committees, two public comment periods, and I think that the common theme here is that many people prioritize these as really high-need measures. They tested well,

were found to be reliable and valid.

In addition to what we submitted in our form, we had some late-breaking data from the first year of HEDIS reporting that we sent in after submission because those were not available during the time of submission, but those results also confirmed that the measures are valid and reliable.

CO-CHAIR BROOKEY: If we could have our subcommittee members speak about the measures. We're going to do one at a time. The first one will be the multiple antipsychotics, right?

MEMBER THACKERAY: Similar to some of the discussions we've had already, I think where the workgroup had some concern in discussing the evidence was the fact that the use of multiple antipsychotics really hasn't been studied rigorously. It makes sense to all of us that if the side effects of one are bad that multiple use is potentially worse. But obviously, those studies aren't out there. A good point was

raised about the lack of clarity with the risks of different combinations. If you have a combination of two or more, is the risk of that combination less or the same or the greater than another combination of different medications?

The point was raised that in certain individual cases, it may be reasonable to have two antipsychotics on board. The measure's, instead, really based on clinical practice guidelines. There's four guidelines from three different organizations referenced, probably the most prominent and largest area of focus being the American Academy of Child and Adolescent Psychiatry, the AACAP. That'll be referenced throughout these three measures.

They're very clear in saying the simultaneous use of multiple concurrent meds has not been studied rigorously and generally should be avoided. Their guidelines, their opinions are based on a lit review of about 150 publications. By their acknowledgment, and to our knowledge, I think there's a couple case reports, but none of

those publications are rigorous studies that directly address concomitant use of antipsychotics. The developers had initially indicated the evidence would be graded as high.

I think our workgroup decided that in the absence of a systematic review and empirical evidence, I think we felt like this might lead us down the pathway to a decision that the evidence is insufficient, but certainly such an important topic that we would consider an exception to that, but I'll open it up to the rest of the subcommittee and the larger group.

CO-CHAIR BROOKEY: Did your group talk about the -- in my experience, it's not just two medications. It's more than two. Did the group or did the developers talk about the cutoff of two versus more than two? I just wonder if the evidence shows that there is a difference between -- I think in your remarks, you have low doses of Risperdal or something else versus higher doses of two or more medications. Was that addressed during the development or the testing?

MS. BYRON: We focused mostly on two or more. We didn't necessarily distinguish between whether you were only on two, and whether you were on five.

CO-CHAIR BROOKEY: I guess the question is why two and not three, I guess, is really what I'm asking.

MS. BYRON: Why two and not three?

CO-CHAIR BROOKEY: Or four, yes.

MS. BYRON: There is evidence -- the evidence is what it is. Really, there's a lack of evidence about long-term safety of antipsychotic use in general. When you add in, too, you have the risk of drug-drug interactions, really felt that the safety evidence was lacking.

In light of that, given those case studies on polypharmacy, two or more, we felt that two -- and it's not actually just being on two. It's being on at least two at the same time for a sustained period of time. That's another component of the measure. The expert panels, and based on our review of the evidence, we felt that

two was the right threshold.

CO-CHAIR BROOKEY: Right, but is that based on consensus, really, the cutoff between two and three, or is it based on clear evidence is my question?

MS. BYRON: Yes, it is in the clinical practice guideline. I'm going to see if, actually, Molly Finnerty has joined us yet. Molly, are on the line?

DR. FINNERTY: Yes, hi.

MS. BYRON: Do you want to introduce yourself and maybe address some of the issues around events with polypharmacy?

DR. FINNERTY: Yes, hi. I'm Dr. Molly Finnerty. I'm a psychiatrist as a research associate professor at NYU, in their department of child and adolescent psychiatry, and director for the New York State Office of Mental Health, Bureau for Evidence-Based Services and Implementation Science.

CO-CHAIR BROOKEY: Molly, I don't mean to interrupt you, but the question we'd like you

to address is the cutoff between two and anything above two. Where do you draw -- is the evidence supportive of two versus three versus four? I'm talking about the threshold for the measure. That's really what we're looking for, in terms of evidence.

DR. FINNERTY: Right. The types of evidence that we considered were one, the reviews and clinical consensus that has come out of clinical practice guidelines from experts in the field. There, they focus on two or more as being a practice to be avoided in children, with a strong recommendation across all of the guidelines available. The other type of evidence is through studies. There has been a review published that found a couple of studies.

There was an analysis of some retrospective and prospective clinical reports that suggest that a patient's risk of weight gain can be influenced by the synergistic interaction between age and polypharmacy, but there was other studies that looked at the relationship between

age and polypharmacy that demonstrated strong positive associations with increased risk of obesity with odds ratios of, say, 2.28 in one study, and cardiovascular and cerebrovascular hypertensive adverse events in children of odds ratio of 1.72. We don't have a wealth of studies, but to the extent that the issue has been examined, it does suggest that there is an increased risk here.

CO-CHAIR BROOKEY: Right. So there's nothing wrong with it being consensus based, but the decision for two is really based on consensus, is that correct, primarily?

MS. BYRON: I'd say it's evidence-informed consensus.

CO-CHAIR BROOKEY: Okay. Jon?

MEMBER FINKELSTEIN: I was part of the small group, as well. I really do see this as resting on the consensus guidelines. To their credit, that's all it can rest on. It's not like there's a randomized trial of two versus a high dose of one. I think we can't have an

unreasonable threshold. That said, the guideline statements, which the developers included, aren't quite black and white. They say these haven't been studied and generally should be avoided. Another one says the prescriber needs a clear rationale for using medication combinations.

It doesn't say never, ever. It says you better be thoughtful about it and not jump to polypharmacy right off the bat. I think that's where we're going to have to really do some thinking. I have some more comments about the reliability/validity stuff that I'll make later, but that's a challenge.

CO-CHAIR BROOKEY: Again, there's nothing wrong with it being consensus based. We do have to think about the unintended consequences to an individual prescriber who may be doing the right thing, and for the kids who are being treated. Amy.

MEMBER HOUTROW: My question is exactly based on that. The way I understood previous measures, such as the tobacco one, is

that we knew that using counseling is effective in reducing smoking. More counseling, therefore, on every patient who is a smoker, then, is beneficial. The goal would be 100 percent. But there is no goal here that I can see from the literature.

There's no epidemiologic data that says how many kids are on polypharmacy, and there's no epidemiological data on which kids have the most severe conditions for which polypharmacy may be appropriate, so there's no possible way that this represents to me a quality measure. It represents a practice measure of what people are actually doing. But because there's no data to say what percentage of kids would be appropriately treated with more than one medication because there's not the evidence for that, I can't see this -- or maybe I'm just missing something. I don't see this as a quality measure because there's no appropriate threshold or goal set. There's just we're going to measure it, and we want it to be low. We aren't going to

know whether or not that was appropriate or not.

CO-CHAIR BROOKEY: I think that, as Jon mentioned, it's not very tight, in terms of where a threshold should be drawn, but I think generally, people are in agreement that it's a red flag if anybody's on multiple psych meds, and we know that psych meds do have problems, especially with drug-drug interactions and weight gain and everything else. I think the issue that you're describing is a threshold issue, is what I'm hearing. Let's go around the room. Go ahead. Who's got their -- is that David? They both have blue shirts on.

MEMBER EINZIG: Just to take a step back, and I apologize if I rehash some of the comments that were already made, but talking about semantics because semantics are important to me, we're calling these medications antipsychotics, even though majority of the time, they aren't used for psychosis, per se. What these medications are, by definition, are they're dopamine modulators. Majority of them are

dopamine blockers. Some of them are dopamine modulators, like Abilify. What we're doing is we're targeting specific brain chemistry to help improve quality of life or quality of functioning.

The concerns about weight gain and maybe two might be more risky than one, I don't know if that really floats with me. I think 15 milligrams of Zyprexa, compared to low doses of two other medications in the same class that are less likely to cause weight gain, I don't think there's strong evidence to back that up. I'll stop.

CO-CHAIR BROOKEY: Okay. David, Number Two? Oh, I'm sorry, Jill?

MEMBER MORROW-GORTON: So, I just wanted to kind of get back to the question, I think again, it may be just semantics. We're talking about polypharmacy, polypsych-pharmacy.

Are we talking about antipsychotics?

CO-CHAIR BROOKEY: Antipsychotics.

MEMBER MORROW-GORTON: Are we talking

about other -- well, just because when people talk about polypharmacy, sometimes, you know, these kids are probably on six or seven drugs. But, they may not all be antipsychotics.

CO-CHAIR BROOKEY: Yes. I just gave an example to Dave in Number One there that I recently had a kid that was on Risperdal, Adderall, Valium, Methadone, and some sort of an opiate.

I'm not kidding, a six year old autistic kid. So, it may have been one antipsychotic, but so it wouldn't fall into this measure.

But your question I think is really about the class of drugs. And I'm interpreting it to be an antipsychotic class. Is that correct? Okay.

So, David?

MEMBER KELLER: And I topped him with the kid who was on seven different meds and was coming to us for Viagra because he was impotent because of all the medications he was on.

CO-CHAIR BROOKEY: Yes, that did top my story.

MEMBER KELLER: So, we all have a story. And by the way, you all can refer to me as The Doctor.

(Laughter.)

MEMBER KELLER: So, -- no, the point I was going to make though was I agree that this is not a measure where the goal is to get the number down to zero.

I think this is a -- though I think it is a useful measure to get -- start getting a handle on what is I believe a problem.

Which is that very often these psychiatric medications are being prescribed in multiple ways. Some of which may be appropriate.

But I think that at a population level, particularly as this measure is being proposed, it gives us a way of looking at the population and trying to recognize outliers in treatment.

And the most important thing that

we're going to be able to do with this is to, I think, is to be able to recognize when we have places where more kids than usual are getting treated with multiple medications. Which should prompt investigation, what's going on here.

Very often, now I would hope we would discover that these are people who are taking care of folks with serious emotional disturbance and require that.

But, it's a -- again, it's a measure that prompts us to look at things. It's not necessarily a judgement that the thing we're looking at is bad.

And I think with that in mind, I actually like this measure. Picking two or three, I see the rationale for picking two.

And so I don't have a problem with it. But again, I think we have to keep in mind that it's not a measure that's to be run into zero.

CO-CHAIR BROOKEY: We may be headed towards a situation with insufficient evidence with exception. I just want to call that out

when people are thinking of formulating their ideas here.

Because I think that we all agree this is an important problem. Ricardo, you're next.

MEMBER QUINONEZ: Yes, I just want a clarification maybe from the developers. Where in -- is there a specific population of kids where this is more of a problem?

So, for example, ADHD, where I could get my head around? That's probably if you're on two or more medications, it shouldn't be a high percent.

Versus kids who actually have psychosis. What was this -- was there a thought of narrowing the population of kids?

DR. FINNERTY: We wouldn't recommend that approach. In children, most of the children receiving antipsychotics do not have a psychotic disorder or any primary indication for the medication.

It's often used for sedation. To try to manage maladaptive behaviors. And

particularly when clinicians are struggling, as they can, that sometimes medications just get added.

It may be of interest to the Committee to note that at the outset, an overall psychotropic polypharmacy measure was there. But that doesn't have the same level of either clinical consensus nationally or evidence.

And even though the evidence is limited here for antipsychotics, it's fairly clear about risks and new onset of diabetes and hyperlipidemias and things like that in children. It's also fairly clear that children are at an increased risk.

And there's just a bit of evidence suggesting that polypharmacy further exacerbates that risk. So, I guess the restriction of diagnosis would maybe not sort of be pointing the flashlight.

That these measures are a powerful tool for doing that. And might not point it in the right direction.

There are some populations that not diagnostically, but really for psychosocial reasons, like children in foster care, are very much at increased risk for receiving polypharmacy.

And also geographic regions. You know, things that don't always make clinical sense. Sort of regional variation. You know poverty, foster care, things like that put you at increased risk.

Children in Medicaid increased risk. So, I think restricting to diagnosis might -- we might miss an opportunity to help increase awareness that this is a concern.

CO-CHAIR BROOKEY: Yes, thank you.

MEMBER AGORATUS: This is Lauren. I have a clarification question.

I actually was just going to ask about the foster care and Medicaid subgroups. And also, a third subgroup, children of color who are more likely to be prescribed psychotropic medications.

Are they also at increased risk for polypharmacy?

CO-CHAIR BROOKEY: Are you looking at disparities I guess is the question.

MEMBER AGORATUS: Yes.

DR. FINNERTY: We have looked at disparities. I think that once you control for these other statuses, we did not see that. So, the regional variation is stronger.

And foster care status is stronger. So, once you're controlling for things like that, you wouldn't observe. And you know, this isn't something that, you know, I could just say in the study that I have fresh in my mind because we did them for example in New York State.

But, I think it's more Medicaid versus not. And foster care versus general population, where you really see it.

So, to the extent that children of color are more represented in those populations, yes. They would be more impacted.

But it may not be due to race or

ethnicity alone.

CO-CHAIR BROOKEY: Okay. Thank you.

DR. FINNERTY: It's more their status.

CO-CHAIR BROOKEY: Kerri, you have a comment?

MEMBER FEI: Sure. I know we've talked a little bit about population level. But at a health plan level, talking with some of our member plans, they see this as an important safety measure.

As a red flag for them. Not that they would necessarily score or consent or reimburse. But, if there is something awry, it would be something for them to trigger them to look into it.

CO-CHAIR BROOKEY: Yes. And just remember, the health plan measure can also drill down to an individual prescriber. So it doesn't mean that it's at a higher level.

It just depends on the health plan, right. Okay. Let's see, go ahead, Kevin.

MEMBER SLAVIN: So, and I was -- I've

been looking at the recommendations. And certainly recommendation eight, which is probably more strongly worded than principle 12.

But it specifically talks about atypical antipsychotic agents. And not using multiple atypical antipsychotic agents as opposed to, you know, multiple antipsychotic agents in general.

So I'm wondering if there's some comment on that? And mixing classes or mixing atypicals with more traditional?

And the second thought is, you know, a lot of what we do in medicine is balancing the -- sometimes very substantial risks of treatment versus the potential benefits. And I see this as one of those situations where there is an unknown at the moment about which patient might benefit from polypharmacy.

And that at the time that the prescribing is done, hopefully there is a discussion with the parents. And if the child is old enough to assent to this, an understanding of

what the risks might be.

And I think if that discussion happens, I don't necessarily know that I see this -- that there's evidence that that -- that this is a problem as long as that discussion of risk benefits has occurred.

And as long as there's a plan if the success is not being seen that we remove part of that risk.

MS. BYRON: So, I agree. The recent Office of Inspector General Report that I just referenced actually came out in March 2015 did a study across on -- across six states that account for about 40 percent of the antipsychotic prescribing.

And they found that at least I think 40 percent of them, of the claims that they looked at, it was the same with both claims, had at least two quality concerns. And of the highest ones, it was poor management.

And it was, you know, too long duration. And too many measure -- I mean, too

many medications.

Recognizing that clinicians need to weigh whether or not the benefits of the medication would outweigh the risk, they also looked to see how many claims were associated with conditions that are included in the FDA boxed warning.

The boxed warning states that if you have a certain condition such as depression, you would have an increase in suicidal ideation and suicidal behavior. And found that a third of the claims were associated -- were prescribed in light of those conditions.

And so, I think that what you're describing is the ideal. And unfortunately, it particularly in Medicaid, I think it's not necessarily the norm.

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

CO-CHAIR SUSMAN: So, again, in an integrated primary care psychiatric setting that was one of the only places where children with serious chronic mental illness could get service.

And having some I think very topnotch folks working.

There were times when multiple antipsychotics seemed at least the appropriate thing to try. If not continue.

So, I agree with the conversation around, this probably shouldn't be a zero percent. But, in looking at the data of people both referred in and then across our community when I was down in Cincinnati, it's pretty clear that there is overuse and not regular reassessment that would justify the ongoing use.

Versus what I think are some tangible harms of ongoing therapy. Therefore, this is one of those measures that I see as, you know, there's not that much evidence.

It's mostly expert opinion. Consensus based. But it's probably the right thing to do.

CO-CHAIR BROOKEY: So, I'm going to ask the final commenters to reflect on the evidence so that we can move forward to a vote for evidence. Because I think we've heard a lot

about how much there is or isn't about evidence irrespective of the importance of the measure. Which is a different question.

So, I'll ask those comments to go quickly. We'll go with Jeff I believe has a comment.

MEMBER SCHIFF: So, just to -- and I'll just say evidence that we have in our Medicaid program is that this is a particular issue just for -- especially for foster care kids.

And I think that I'm in support of this measure. But part of it is because when people look at polypharmacy, they often look at ADHD drugs and SSRIs. And we're really just looking here at the antipsychotics and the issues around those.

I have other comments, but I think I'll wait until further down the discussion.

CO-CHAIR BROOKEY: Okay. Amy?

MEMBER HOUTROW: I appreciate that the number shouldn't be zero. Right, there are kids

that it is appropriate for them to be on more than one medicine.

But what I'm missing here is the evidence to support this. This isn't an overuse measure. This isn't just a measure of the clinical behavior.

And while I think that the issue is important, overuse. This isn't -- there's no evidence that we know of where the appropriate number should be to look at overuse.

And we don't have any ability using this measure to do anything other than red flag potential overuse. And so, I've heard that red flag statement a couple of times. It's totally an important point.

But, the measure is a quality measure, not a red flag measure. And I just don't think that there's evidence that this is giving us enough to use it as a quality measure as it stands.

CO-CHAIR BROOKEY: Okay. Dave?

CO-CHAIR SUSMAN: Would you find it

important if one plan had 20 percent of individuals who had more than two? Or two or more? And then three percent in another plan?

Would that be something worth looking into?

MEMBER HOUTROW: Right. Variations in care are important to evaluate. But we just don't have whether the 3 percent number is right or the 7 percent number is right. Or the 2.1 percent number is right.

We just don't have that information.

CO-CHAIR SUSMAN: Yes. I think all you can say about it is that this calls for further review. If you have an outlier, it calls for further review.

It doesn't necessarily mean that they're a bad outlier, right? These kids could be totally, appropriately treated.

I think we would agree with that Amy. David?

CO-CHAIR BROOKEY: I think Marlene was next. I can't see very far down there.

MEMBER MILLER: I wasn't next. But I just have a quick comment I wanted -- because it builds off of there.

I think the problem we all have, or at least what I feel, is that I know we sometimes think this is a measure just for internal quality review. But then it gets put out for accountability.

And some arbitrary measure of this should be zero gets out there. And it's the wrong way.

And we -- every measure group I've ever been on, we've always had this discussion is this for QI or accountability? And every time the usual answer is, it's going to be -- end up used for any of those.

And so we don't control whether this will be put out there for accountability and then some arbitrary number gets put on the right level.

CO-CHAIR BROOKEY: Okay. I think we may be back down to the last comment by David.

MEMBER EINZIG: As usual.

CO-CHAIR BROOKEY: David One.

MEMBER KELLER: No, I'm The Doctor. You're David. My Tardis is outside.

MEMBER EINZIG: I appreciate the differences. And it's a legitimate issue.

But my concern is, so I wouldn't rate the evidence as insufficient. I would rate the evidence as low.

Because there is -- there just is no evidence. It's not studied. There's no studies of looking at this long term.

People on one medication versus on two medications. And so I don't see a way around that.

And then with the expert consensus, you know, stated as polypharmacy should be avoided. It's not saying it shouldn't be done.

Granted, there are doctors out there that do wacky things with medications. But, I just don't think this measure really flies with me.

CO-CHAIR BROOKEY: So, I'm going to see if maybe we can get close to a vote. I think that most of us are thinking this is probably not going to be a high or a moderate in terms of evidence.

But I think if people want to move it forward, then they -- and if they don't think it's high or moderate, they're going to have to vote it insufficient.

And then we can go for exception. So, to David's point, if you think it's low and you vote low, that means you're not really wanting it to move forward.

So, I just want to be clear about how we're voting. And you know, it's a really good discussion. I guess the question is about the unintended consequences of a measure like this. How it's going to be used? You know, what does it mean? And so forth.

Go ahead Helen.

DR. BURSTIN: I just want to respond to your comment and Marlene's as well. It's

really important to remember, this is a health plan level of analysis measure.

NQF measures are only intended at the level of which they're tested, submitted, and endorsed. So, that is what it is.

So, to invoke how it might be used at the clinician level, if that measure is intended ultimately at the clinician level, it has to return and demonstrate testing. And look at all those other issues you're raising at the clinician level.

It's truly, only at the health plan level.

CO-CHAIR BROOKEY: Health plans use data though to get down to individual practice. And so I think we have to be clear about how health plans may use data.

It's not necessarily just going to be aggregated data. It can go down the practice. It can go down --

MEMBER MILLER: Accountability of the health plan level, right?

CO-CHAIR BROOKEY: Yes.

MEMBER MILLER: Of someone judging a health plan does or doesn't provide good care based on some arbitrary -- or lack of evidence as we were saying of what is the right number? Or right level or acceptable performance on this measure?

So it would be at the clinician.

CO-CHAIR BROOKEY: If the health plan isn't doing well on a measure that pressure can trickle right down to the individual practice of a doctor. There's no question about that.

Right Kerri?

MEMBER FEI: Absolutely.

CO-CHAIR BROOKEY: So, I just want to be clear, this is maybe a health plan level. But that doesn't mean that it's not going to have implications for individual practitioners.

MEMBER FEI: The other thing is that it really as far as holding health plans accountable, it would be up to their accrediting body to do that.

To my knowledge, it is not included in the health plan accreditation standards as a HEDIS measure to which plans are evaluated at this present time.

And I don't see where this would -- this one specifically, I don't see where they would consider adding that at any time. Maybe some of the others.

But, I don't think this would be one that would rise to the top for that purpose.

MS. BYRON: I just wanted to say, it's health plan and State level. So, it's both.

And to Kerri's point, you're right. Not all HEDIS measures are used in accreditation.

We have another process where an external multi-stakeholder body would look to say does this measure make sense in an accreditation context? And they would consider all of the things that have been actually very, you know, well thought out and raised in this Committee.

And you know, some of the issues I will just say, these are issues that do apply to

all over use appropriateness measures, whatever you may call them. Where lower is better. We don't know that we necessarily want to get to zero.

I would submit that there are other measures where higher is better. And we don't necessarily think we can get to 100. And it may not make sense there either.

And so, you know, in the context of thinking about appropriateness and overuse, which is where I think we do want measurement to be headed towards, where you can be looking at the appropriateness of clinical care. And not advocating for more, is better.

We do have the challenge of where do you set a threshold? But, we do hope that the measure can shine a light on issues and flag problems. And allow for comparison.

CO-CHAIR BROOKEY: So, I don't -- unless we have anything new to say about evidence, I mean, I think we all have heard.

I think we all understand the

implications that having a measure that's based on consensus and not evidence. And how that our first charge is to really evaluate the evidence of a measure before it goes to the second question.

So again, I'm going to repeat that what you decide on will determine whether or not this could move forward either as a medium or insufficient evidence with exception.

If you vote low, that means you're not wanting it to move forward at this time as specified. It doesn't mean that it's forever dead.

It just means that as specified, you're voting it low. And will not move forward.

Are we ready to vote on the evidence aspect of this? Okay. So, let's vote.

MS. CHAVEZ: Okay. We're getting ready to vote for Measure 2799. One for high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Ten. Twenty-one. Twenty-three. We are expecting 23 votes. Okay. Zero voted high. Zero voted moderate. Six voted low. Seventeen voted insufficient.

CO-CHAIR SUSMAN: I think it would be worth asking if we want to have an exception on the evidence for this one.

CO-CHAIR BROOKEY: Can we just get a straw vote of how many would like to open this up for discussion and vote for exception?

CO-CHAIR SUSMAN: A vote, no discussion.

CO-CHAIR BROOKEY: Why don't we just vote then. We will put this up for a vote for insufficient evidence with exception.

Which means it can move forward to the full review. Okay?

MS. CHAVEZ: Okay. One for insufficient evidence with exception. Two, no exception.

(Pause.)

MS. CHAVEZ: Sixteen. Twenty-two.

We're expecting 24 votes. Twenty-three. Twenty-four.

Okay. Seventeen voted insufficient evidence with exception. Seven for no exception.

CO-CHAIR BROOKEY: So I think that means that we think it's an important issue. Whether or not we agree with the specification. So that's -- well, that's the way it will be presented.

The next question has to do with performance gap. And do our Subcommittee Members have some comments about performance gap?

MEMBER THACKERAY: Yes, sorry. My computer is now updating Windows and restarting. But, I think --

(Laughter.)

MEMBER THACKERAY: It couldn't be worse timing. But I think we touched on it in the previous conversation.

The developers cite several -- thank you -- systematic review that found one in ten youth being prescribed any antipsychotic were on

multiple, concurrent antipsychotics.

They also cite data from their own analysis of Medicaid data showing that there were average rates of 6 percent in the general Medicaid population. A slightly higher average rate for children in foster care.

And then the disparities we already touched on in previous discussions. So, I think it's pretty clear about that.

CO-CHAIR BROOKEY: I think we talked about this quite a bit earlier. Is there any other discussions about performance gap? If not, could we -- go ahead Jeff.

MEMBER SCHIFF: I have a question probably for NQF Staff. You know the -- my understanding is our measure goal here is around accountability.

And so this seems like a quality improvement measure that we're applying at a health plan level. And I just -- I guess I just want to -- it seems like we're a little off.

And I mean, I think -- I don't think

this is an unimportant issue. I could talk for a long time about why it's important.

I just -- I'm just trying to figure out, you know, it seems to me like this is a quality improvement topic that we're applying to an accountability standard where we don't have a threshold.

And I just -- I don't know if any -- I think along the same lines that Amy maybe had said. So, I think we have to -- you know, I think it's -- I would love for this to be measured in a lot of places to look at the quality of it.

But, I'm not sure that at the measurement level it's the right thing.

MS. BURSTIN: Again, I think it's a judgement call. I mean, it is certainly -- we say measures can be used for accountability and quality improvement.

So, I don't think we're making a distinction saying it's one or the other. Ideally it would be intended for accountability.

And certainly I think the way NCQA would use it would have some element of bench marking, et cetera. Which would fit under accountability as well.

CO-CHAIR SUSMAN: And if I'm correct, NQF has developed a process to consider this particular issue. Is that?

MS. BURSTIN: Yes.

CO-CHAIR BROOKEY: Any other comments about performance gap? Otherwise can we go to a vote?

(No audible response.)

CO-CHAIR BROOKEY: Okay.

MS. CHAVEZ: Okay. We're now voting on gap. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Thirteen. Twenty-two. Twenty-three. Twenty-four. Five voted high. Seventeen voted moderate. Two voted low. Zero for insufficient.

This measure passes performance gap.

CO-CHAIR BROOKEY: Thank you. Let's move onto reliability. Does our Subcommittee have any comments about reliability?

MEMBER EINZIG: So, reliability testing at the State level was deemed as good. But at the Medicaid level, I'm not a statistician, so I can't really speak strongly about that.

But, it was not seen as reliable at the Medicaid level.

CO-CHAIR BROOKEY: Okay. Jon, did you have a comment?

MEMBER FINKELSTEIN: Yes. So, I'm part of the Subgroup as well.

CO-CHAIR BROOKEY: Um-hum.

MEMBER FINKELSTEIN: So, the reliability issue is that these are relatively rare events. So that you need very large populations to meet any reliability standards.

So, it came out okay at the State level. At the Medicaid health plan level, if they're big enough and have enough of these kids

in the denominator, it was eh, okay.

And that at the commercial plan level, they had to eliminate 24 of 72 commercial plans because they had less than 30 patients in the denominator. So, it's all about only in big populations.

I understand NCQA's approach to say this is only reasonable at the State level or very large Medicaid health plan levels. How it will be used in practice we can't control.

I just -- I'll do validity here as well. Which is that they used psychometric validity methods which are good.

And they came out again with coefficients and rankings that are okay. But, what we don't have, and I think this measure is supremely important.

What we don't have is taking people that would fail the measure based on claims. And looking in the chart for some sample and saying would three other psychiatrists agree that this was terrible care? Or 90 percent of the time,

would they be acceptable?

So, we don't have any of that validity. We do have face validity because of the process they used, the stakeholders they got, and the guidelines that it rests on.

So that's -- it's a very -- for me it's a very complicated reliability and validity picture.

CO-CHAIR BROOKEY: So, this is a plan level measure. But it could become a Medicaid plan measure, right?

So I mean, theoretically, this could just be -- it doesn't have to become a commercial measure. It could be --

MS. BYRON: Right. So -- and actually so, -- so, you're right. I mean, those points are well taken.

In the more recent HEDIS data that we used to assess, you know, based on first year results from HEDIS, the Medicaid results were actually reliable at .95.

So, you know, in our field tests, they

were moderate. But then when we went to the larger HEDIS reporting, it was high.

So, it was lower for commercial though. And it is for the reasons that you stated about the denominator.

And right now it's in HEDIS. It applies to Medicaid and commercial plans.

CO-CHAIR BROOKEY: Um-hum.

MS. BYRON: But, we will be publically reporting for Medicaid plans only because of the issue with the commercial plans and the denominators.

CO-CHAIR BROOKEY: So that gets to John's point. So, we're not voting right now as to whether or not it's going to be a Medicaid plan or -- a measure, or a commercial measure.

It sounds like the reliability is high in the appropriate population. So, keep that in mind when you're voting for reliability.

That it's going to be more reliable in a higher risk group that is more -- that has more exposure. I think that's all intuitive, right?

So, do we have enough information to vote on reliability? Go ahead Jill.

MEMBER MORROW-GORTON: So, if we -- if this is reliable at only a big plan, so you know, if you have a State that has five Medicaid plans and they're all really small, then you can only use this at the State level, right?

CO-CHAIR BROOKEY: Possibly.

MEMBER MORROW-GORTON: But, States aren't going to know that. I mean, States aren't going to know that.

And they're going to try to use it at a plan level that's not going to be reliable.

CO-CHAIR BROOKEY: So that's a limitation of the measure because of the small volume. So, that's what John just stated.

Anything else new about this particular issue? Karen?

MEMBER DORSEY: Yes, I mean, I'll just that that can be pretty readily dealt with. Because if the cases are too small to report a reliable number, you can report that the cases

are too small to report a reliable number.

And that in and of itself in this context is useful information. Because it shows comparatively the population of kids who were even considered for mono-therapy in private plans versus Medicaid plans. Right?

So, in the context that we're talking about this, it's still useful information to be reporting.

CO-CHAIR BROOKEY: Okay. Just go up the line. John?

MEMBER THACKERAY: So I think in our small group discussion, we had one or two questions regarding specifics of the numerator for the developers. One was, if you have a child who's on a scheduled medication and one that's prescribed on a PRN basis, would that count towards the numerator?

And the other question was, if you have a child who's on one medication throughout the period of measurement, is on a second measurement for 45 days, gets switched from that

to a different second medication for 45 days, would those kids be included in the numerator as well?

MS. BYRON: Okay. You're having me do math in my head.

(Laughter.)

MS. BYRON: Let me see if I can explain and answer the question. And Molly, feel free to jump in as well.

So, if you are on one -- it's any two. So, if you're on one for 40 -- I think we specified 90 days actually. So, it's 90 days of sustained use.

We do allow for a gap in case of titration. So, we do want to account for that. And that's where the -- we place that.

And I -- Molly, do you want to respond to the questions about --

DR. FINNERTY: Yes, sure. You know, we look at a day's supply. So, if you say that you have a second antipsychotic on as a PRN, you know, one of -- the challenges with PRNs is we

don't, you know, when folks go home, we don't know exactly how they're using them.

This would say that if they're using it enough that they have a day's supply that covers the majority of the 90-day period, then it really is a concomitant use.

If on the other hand they're using it so rarely that they don't need to fill that second PRN, you know that they had very few days' supply, then they won't meet criteria for polypharmacy.

MEMBER THACKERAY: Okay. Thanks.

CO-CHAIR BROOKEY: Okay. David?

MEMBER EINZIG: So, from a clinical perspective, you know, are we hitting the right target on this in terms of equating single medication as better quality than polypharmacy?

And I still don't think that I'm there in terms -- if that's accurate to say if that's a validity question.

If, you know, when the examples that we used in our telephone discussion was, so

Risperdal is six milligrams, is that safer than one milligram of Risperdal with 25 milligrams of Seroquel at night? Not that you would do that at first.

But if ultimately you try different things and they get elevated prolactin or elevated dystonias. Other problems with Risperdal, but it works great. Every other medicine fails.

You know, so there's going to be scenarios where it does make more sense and is clinically indicated to go with the polypharmacy. And it's maybe safer.

CO-CHAIR BROOKEY: Yes. So that's a validity question I think. So, we want to talk about reliability.

And I think the major issues that I'm hearing about reliability have to do with the size of the plan. And the mix of the plan in terms of payer source.

Because Medicaid will probably yield more. But irrespective of that, a small plan

will be challenged to produce a report that's statistically significant and comparable.

So, having -- understanding all those limitations of the measure, can we vote on reliability?

MS. CHAVEZ: Okay. Now voting on reliability. One for high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Twelve. Twenty. Twenty-three. Twenty-four. Zero voted high. Ten voted moderate. Twelve voted low. Two voted insufficient.

This measure does not pass. It's in the gray zone. This brings us in the gray zone.

CO-CHAIR BROOKEY: What's the percentage? You had three and four, right?

MS. CHAVEZ: For 24 votes, between 14 and 10, it takes us to the gray zone.

CO-CHAIR BROOKEY: Okay. So validity, we heard there is some face -- I'm sorry, can we move on? Yes.

So, validity we had a little bit of conversation about face validity. Any other comments about validity from the Subcommittee?

MEMBER HOUTROW: I'm reading under specifications, the section to determine if a measure of specifications are consistent with the evidence. The statement is that the specifications are consistent with the evidence.

The goal of the measure is to assess inappropriate prescribing of antipsychotic medication to children and adolescents. And what I would say is this does not measure the inappropriate prescribing.

It only measures the prescribing of antipsychotic medications. So, therefore it doesn't meet validity specifications.

Because it doesn't do what they intended. Which is to assess the inappropriate prescribing. The numerator is two or more. The denominator is any.

So therefore they're only measuring the percentage of prescribing with no

relationship to whether or not that's appropriate or not.

CO-CHAIR BROOKEY: Other comments?

(No audible response.)

CO-CHAIR BROOKEY: So, I think we're challenged with the evidence. Amy's point is that just because somebody is prescribing more doesn't necessarily comment on the inappropriate prescribing.

David's made his comments about his concerns. Any other concerns about validity or comments before we vote on validity?

(No audible response.)

CO-CHAIR BROOKEY: No? Okay.

MS. CHAVEZ: Okay. Now voting on validity. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Sixteen. Twenty-one. Twenty-two. Twenty-four. Zero voted high. Six voted moderate. Fifteen voted low. Three voted insufficient.

This measure does not pass validity.

CO-CHAIR BROOKEY: So we stop here. And we'll pause to make any comments that we have not already made.

I don't want to repeat comments that we've already made. I think the developers have taken notes.

I think we all think it's a very important issue. We're just not thinking that the measure is getting to the specificity of the individual practitioner's problem with prescribing.

It's just a question of how much is prescribed and not whether it's appropriate or not. And really, the evidence does not support a specific threshold.

And I think those are the reasons why we have concerns about this. Jeff?

CO-CHAIR SUSMAN: So, I personally think this was a good enough measure. It wasn't a perfect measure.

But if we're going to get to the

issues of appropriateness more granularly, then one approach could be to define those exceptions where using more than one antipsychotic would be appropriate. Or excluded.

That would be challenging given the multiplicity of reasons why that might be, you know, needed.

But it also if you had two or three exceptions documenting trials with other medications, what have you, documenting improvement over a sustained period of time, one could at least imagine a new measure that is more granularly looking in appropriateness.

CO-CHAIR BROOKEY: Really good discussion. And I feel bad that we can't move this measure forward. A better measure forward I guess is my point.

Let's go to the next discussion unless there's any other --

CO-CHAIR SUSMAN: You do have cards. I don't know if --

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

just kind of went blind with the cards that are -- if you have new comments that have not already been said, let's go ahead and make those.

So, Ricardo?

MEMBER QUINONEZ: So, just a point of -- a challenge to the developers. I think this one suffered a little bit, not as badly.

But suffered a little bit from the same problem as the overuse of imaging. Where it is definitely a problem.

But the bigger problem is kids who are on antipsychotics inappropriately versus this probably smaller population of kids who are on multi -- multiple antipsychotics.

So, a great numerator would be kids who have been diagnosed appropriately with the right DSM-5 criteria, with psychosis. And then the number of children on antipsychotics or whoever.

But just getting at the bigger problem rather than a little bit more -- such a specific one.

CO-CHAIR BROOKEY: David?

MEMBER KELLER: So, I'll respectfully disagree with my colleague. And in that I think the other approach you might think about and it might be worth thinking about on our phone call coming up, is to accept that we are not measuring inappropriate use.

That we are measuring use. And that there is value in measuring and tracking use. And in that you can look for variation.

Please keep in mind that I said both, crossing thresholds and finding practice variation. And I think that is something that's worth doing.

So, I'd think about that.

CO-CHAIR SUSMAN: You weren't disagreeing at all. I think --

MEMBER KELLER: Yes. No, I was disagreeing with the last. I don't -- I was disagreeing with him.

(Laughter.)

CO-CHAIR SUSMAN: Well, then disagree

with Ricardo.

MEMBER KELLER: I'm disagreeing with Ricardo. And it's just another approach.

But, I do think, and again, I fully believe that if we don't measure it, we don't fix it.

And by having no measures, even of the -- of use of antipsychotic medications right now, it gets swept under the rug except when the GAO does a big expose about it.

And it has been identified by our -- by the GAO and by HHS as a problem that we need to deal with.

So, while we're building that better measure, which I think will take time and is worth doing, I think we need to have something to be able to track this.

CO-CHAIR BROOKEY: I agree. Jeff?

MEMBER SCHIFF: So, I think part of the Agency for Families and Children, which does overseas foster care in the country is one of the groups that has really flagged this.

I just want to say really quickly, 9,000 kids in foster care in Minnesota for 30 days or more. A little over a third of them on psychotropic meds.

Only about 350 to 400 on four or more meds. And we sort of had a freebie on -- we figured that a lot of them were on ADHD drugs or SSRIs first.

So, that's the case. And then we looked at that 350 and saw whether or not they had adequate mental health interventions in terms of therapy.

Because I think the goal of a measure like this is to make sure these kids are treated appropriately for their mental health disorders. And 80 percent of those in our population had 10 or more mental health visits during that year.

So, I guess what I'm really saying to the developers is that I don't think a measure -- we're never going to get to a measure that is going to perfectly look at the correlation of diagnosis with the number of psychotropic meds.

And I would rather put our chips on looking at whether these kids in health plans or not in health plans are actually getting some sort of trauma informed therapy as part of their care.

CO-CHAIR BROOKEY: Okay. Carol, last comment?

MEMBER STANLEY: Yes. Actually we're using this measure currently with our foster care population in a special study. And we're approaching it as being able to gain a snapshot and a potential baseline.

So, we're going to see what we end up with. We're going to look at regional variations within our State. Racial disparities and that type of thing.

So, we're approaching it as being able to have a snapshot.

CO-CHAIR BROOKEY: That's great. So, we're going to move on. I'll use the Chair's prerogative to make the final final comment.

Which is, we can't figure out who our

foster care children are. And I wish the plans could. Because I think that would be a whole other level of reporting.

So I'm glad that you can do it. So, thank you very much. You have the comments.

And we're going to move onto the metabolic monitoring. For which a lot of the earlier discussion reflects on this measure as well.

So, anything new that you want to present -- than what you presented earlier about this particular measure.

MS. BYRON: Well, actually I think that, you know, I'm glad that you -- the Committee has identified those concerns. Because I think both this measure and the one on psychosocial care that's coming up, does get at some of the things that you've highlighted here.

So, nice segue.

CO-CHAIR BROOKEY: Thank you. To those in the Subcommittee, and I don't have the list in front of me, so who are you?

(Laughter.)

CO-CHAIR BROOKEY: So, who wants to start out on making some comments about this particular measure? Jon, do you want to start? Or Maureen?

MEMBER FINKELSTEIN: So, this measure addresses -- it's -- I think we have to decide whether it's a processor or an intermediate outcome. It could be either.

But, metabolic monitoring, specifically glucose monitoring and lipid monitoring for kids on antipsychotics. I think it rests on a fair amount of evidence.

And in this case, real clear recommendations by the professional societies. That even though safety of these drugs hasn't been fully evaluated in kids, there's concern for these metabolic derangements. And that there should be blood testing for metabolic monitoring.

If you ask whether we have evidence for the long term outcomes, we don't. But that doesn't bother me in this particular case.

I think -- I'll stop there. Because that's the evidence piece.

CO-CHAIR BROOKEY: Maureen? Anything to add? David? So, you're saying the evidence for this one is much stronger.

And there was a question about age group, but irrespective of age, we all agree that kids should have metabolic screening? Okay.

So, any other question about evidence? Carol, you have your flag up? Marlene?

MEMBER MILLER: Well, I guess I -- oh, I would just like a little information about the two or more. Does it really have to be two or more?

CO-CHAIR BROOKEY: Well, where are --

MEMBER MILLER: In the measure. It doesn't say if you have two or more -- two more prescriptions had metabolic testings. Is that two more drugs?

MEMBER FINKELSTEIN: No, no, no. That's two or more dispensings. I read that as having two or more dispensings.

CO-CHAIR BROOKEY: Right. It could be one medication.

MEMBER MILLER: Oh, I read that as two or more medications.

CO-CHAIR BROOKEY: Yes. So just one medication would put you in. Yes, yes.

MEMBER MILLER: So, two prescriptions ever? Or --

CO-CHAIR BROOKEY: During the measurement year.

MEMBER MILLER: You know, serial --

MS. BYRON: During the measurement year.

CO-CHAIR BROOKEY: During the measurement year.

MS. BYRON: It can be the same.

CO-CHAIR BROOKEY: Two or more dispensings. And irrespective of the number of days.

MS. BYRON: It can be -- we wanted to operationalize some sort of ongoing use. So, more than one.

It could be two prescriptions of the same drug. Or it could be two different.

CO-CHAIR BROOKEY: Okay. Ricardo?

MEMBER QUINONEZ: So, this is not going to be a popular opinion. But, I always have difficulty with screening measures.

Because this really is a screening measure. It's screening for complications of a drug.

And every time there's a screening measure that has very little outcome data to validate it, many of them are later reversed. Because the screening is not necessarily tied to the outcome.

And so it does bother me that there is no outcome. Because of previous experiences with screening measures.

CO-CHAIR SUSMAN: I mean, I think there's a pretty clear causal pathway here from metabolic outcomes to patient oriented outcomes that we really would think are important. Pardon me?

MEMBER QUINONEZ: To get to from an antipsychotic?

CO-CHAIR SUSMAN: No, from the metabolic issues that occur with the use of antipsychotics.

So, while I agree, it is not A Level evidence, I think there is a believable sort of nice connection that puts the dots. At least for me.

CO-CHAIR BROOKEY: Yes, in my mind, if you call it screening, I'd call it almost like secondary screening.

I think the term monitoring is probably more appropriate. Because you're really monitoring for side effects of a drug.

So, I don't see this as a screening measure per se. I see it more as -- more like any other medication for which it requires monitoring.

So, I think it's different than the screening of a low risk population. You're doing monitoring of this very targeted population of

kids on a particular class of medication.

So, I'm just seeing it differently. Any other? Okay.

MEMBER QUINONEZ: Is there any evidence that any interventions that you would do after this monitoring, actually are beneficial?

CO-CHAIR BROOKEY: We have a psychiatrist who can answer that questions.

MEMBER EINZIG: Yes. Is a --

MEMBER QUINONEZ: Especially on long term outcomes?

MEMBER EINZIG: Well, the -- if it's a direct effect of the mediation. You stop the medication. You go with an alternative approach.

Or you do dietary or nutritional or omega-3. Lots of things you can do for it.

MEMBER HOUTROW: So, I have a question that goes back to the original numerator question. Does that also include children who are on PRN medication that they might have only taken two or three times in a given year?

If you have Risperdal, and you got a

prescription in January, and then in December.

CO-CHAIR BROOKEY: Is there a day supply? Or is it just a dispensing?

MS. BYRON: It's just a dispensing. Molly, do you want to elaborate?

DR. FINNERTY: Yes. I think the thinking here is that the guidelines recommend that you should test children at baseline. And then on follow up.

And this is sort of more -- it allows that you could have done it at baseline, or you might have done it at follow up. But, it's only looking for one lab at any time.

So, it's a generous measure in that way.

CO-CHAIR BROOKEY: Well, the question though is, if you just got five pills of Risperdal, I'm just making something up. And then you got a second prescription for two pills six months later, you'd still be in the denominator, right?

DR. FINNERTY: Yes.

CO-CHAIR BROOKEY: But that probably doesn't happen that often. I mean, I'm not a psychiatrist.

But, just -- so, I'm thinking people get at least maybe 30 pills or something. Okay. So, I think your question is a good one Amy, but I think it's answered. Right?

Okay. And Sue?

MEMBER KONEK: The fact that it does have the baseline screening. Which is really critical for this to work.

CO-CHAIR BROOKEY: Yes, yes. To know where you started. Are there any other questions or comments about -- what are we on? Evidence.

Can we vote on -- oh, I'm sorry, do you have a comment? David? David Two?

MEMBER KELLER: Yes. The -- no, the only -- I actually was following up on the question of screening versus -- this to me is actually a patient safety measure.

And so I think about it as -- I think about it a bit differently. But I think that

makes it equally important.

And the interventions are about that we are doing something and we need to be able to monitor that. So, to me the other piece of this measure is that we're monitoring for specific conditions.

But we're also tracking to make sure that a patient on a potentially dangerous medication is being monitored by someone. Because someone's ordering tests around it.

CO-CHAIR BROOKEY: Right. Any other comments? Otherwise we'll move towards a vote on evidence.

(No audible response.)

CO-CHAIR BROOKEY: Okay. Good question. Is this -- do we have sufficient evidence for a high?

There was a --

MEMBER EDIGER: It was graded.

CO-CHAIR BROOKEY: It was graded.

MEMBER EDIGER: And the consistency, quality, and quantity, yes.

CO-CHAIR BROOKEY: Yes, I think there -- you can vote high if you choose so. Okay. Ready?

MS. CHAVEZ: Okay. We're now voting on Measure 2800 on evidence. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Sixteen. Twenty-two. Twenty-three. Okay. Four voted high. Eighteen voted moderate. Zero low. One for insufficient.

This measure passes evidence.

CO-CHAIR BROOKEY: So let's move to performance gap. Subcommittee Members want to comment on performance gap?

MEMBER FINKELSTEIN: Well, I'll start. I think there is some, there's studies that they quote that 31 percent had a glucose test. And 14 percent had a lipid test in their field tests.

The rates were very low. I hope -- I wonder if the developers can comment on their confidence in the capture in claims data that all

of these tests end up as a separate claim?

I'm particularly thinking about kids who are in inpatient settings where I'm not -- they know more than I do. But I'm not sure that every blood test is billed separately to the insurer given how places are now paid.

And I wonder if we have information.

MS. BYRON: So, this -- the specification for this is actually modeled on the way we look at metabolic testing for a couple other HEDIS measures. We have a HEDIS measure that looks at schizophrenia for adults.

And so, we -- so based on, you know, the experience on those measures as well as our field testing, we do feel pretty confident that we captured it.

Our field test findings were really low. The first year HEDIS results were a little higher actually. Probably closer to what we saw in the literature.

I think it was between 25 and 30 percent in terms of our first year means. So, we

-- we're pretty confident.

I'd have to check the spec to your question about the inpatient. They may be excluded. And I don't know, Molly, if you remember off the top of your head?

DR. FINNERTY: Labs usually drawn in an inpatient setting are not included. But, it's, you know, a minority of children who are in an inpatient setting.

And you might wonder whether the outpatient psychiatrist who's prescribing the meds, you also don't for example see the medication. So, whoever is prescribing the meds should also be monitoring.

So, that, you know, that -- a medication was prescribed. Or that a monitoring was done in the inpatient wouldn't necessarily cover the child for monitoring in the outpatient setting.

CO-CHAIR BROOKEY: I think as well, lipids may not often be a priority when they're at the inpatient. So, I think that unless we're

measuring the inpatient, we -- you may not be that much -- there may not be that many more hits that we would get from getting the inpatient data.

That's my experience. Any other comments about whether there's a gap? It sounds like there's a relatively low rate of performance.

So, that would indicate if we believe this is an important measure that there is a gap. Is that right?

So, unless there's any other comments, could we vote on gap?

MS. CHAVEZ: Now voting on gap for Measure 2800. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: We're expecting 23 votes. Twenty-two. Twenty-two.

MS. ALLEN: Lauren, please submit your vote.

MEMBER AGORATUS: I did. Did it go

through?

MS. ALLEN: Yes. Thank you.

MS. CHAVEZ: Okay. Ten voted high. Thirteen voted moderate. Zero low. Zero insufficient. This measure passes performance gap.

CO-CHAIR BROOKEY: Okay. Let's move onto reliability. Do our Subcommittee Members have any comments about reliability?

(No audible response.)

CO-CHAIR BROOKEY: This is a health plan level -- is this -- remind me, is this administrative? Or is this a hybrid? Or is it a chart review?

MS. BYRON: Administrative.

CO-CHAIR BROOKEY: It's an administrative measure. So it's based on claims data or coded data.

So reliability should be high. Testing was high. Any other comments about reliability of the measure?

MEMBER FINKELSTEIN: I'm waiting for

my other subgroup. So, they did testing in the max data sets.

The States had a meeting about denominators were about 12,000. So, it doesn't -- it's not as small as the previous measure.

Still a problem in some commercial plans not having enough to get very high reliability. But, I thought it was pretty well documented.

CO-CHAIR BROOKEY: Yes, so the question is, as specified, is it a reliable measure. Are you going to be able to collect what you're trying to collect?

Any other comments? Yes, go ahead Kevin.

MEMBER SLAVIN: Just a quick question. Glucose testing I know can be done at the point of care. So, a child with an early morning doctor visit might have a fasting glucose done at the office.

Would this measure capture that as something that would be billed?

MS. BYRON: I believe it would. Yes.

CO-CHAIR BROOKEY: We would code it as a POCT. And if you're -- yes, I would think we could.

Any other comments about reliability?

(No audible response.)

CO-CHAIR BROOKEY: Okay.

MS. CHAVEZ: Okay. We're ready to vote on reliability. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Again, we're expecting 23 votes. Nineteen. Twenty-two. Twenty-two.

CO-CHAIR BROOKEY: Vote again.

MS. CHAVEZ: Twenty-three, thank you. Five voted high. Eighteen voted moderate. Zero low. Zero insufficient. This measure passes reliability.

CO-CHAIR BROOKEY: Okay. I think we've talked a lot about validity. Other comments about validity? Ricardo?

MEMBER QUINONEZ: This does not answer

my question about what consequences are involved.

CO-CHAIR BROOKEY: Is your microphone on?

MEMBER QUINONEZ: Oh, I'm sorry. No, it's not.

So, one of the questions that was put forward, and it says what evidence is there to demonstrate that a change in antipsychotics, which I assume would be what you would do, as a result of metabolic tests, is less harmful from a behavioral health perspective to the affected patient then remaining on the antipsychotic that is causing these issues.

CO-CHAIR BROOKEY: David, do you want to take a stab at that?

MEMBER EINZIG: So you could change -- if it's a bipolar diagnosis, you could change to a Depakote, Lithium, other meds that aren't going to affect at least metabolic syndrome.

In rare cases you might want to go with typical antipsychotics. Which, you know, would be lower risk.

But within the atypical class. Certain medications are clearly higher risk then other meds within the class.

MEMBER QUINONEZ: Right. But is there any evidence that changing to those offers less harm? Or -- then, you know, then being on the appropriate antipsychotic?

CO-CHAIR SUSMAN: Well, there's different risks by medication and by general class. And even within the class.

Whether that then is associated with a reduction in ultimate, say cardiac events or issues related to the abnormal metabolic events, I don't know of any data for that. I mean, is that an accurate?

MEMBER EINZIG: Yes. I don't know of any specific studies looking specifically at that questions.

But, intuitively, if you want to go with face validity, it makes sense.

CO-CHAIR BROOKEY: Okay. I think these patients are on medications for a long,

long time. And so I think that's the issue is you have to adjust them to where they're potentially going to cause the least harm, right?

But oftentimes you're trading off one medication that has risk for another medication that has a different risk. And that's clear.

Any -- so, any other -- oh, I'm sorry, Kevin, go ahead.

MEMBER SLAVIN: I think this is one of those situations where if you were talking about a non-antipsychotic, like if you were talking about Monacolin, which can cause a life threatening lactic acidosis. If the patient started showing signs of that, you would stop and find an alternative therapy.

I think having a metabolic derangement on a medication that somebody's going to be on for a long time, I think that the risks of that would certainly outweigh any concerns about trying to find an alternative regimen.

CO-CHAIR BROOKEY: Jill?

MEMBER MORROW-GORTON: And I think the

other thing we have to think about is talking about appropriate medication. I mean, all of the antipsychotics sort of do the same thing in little different ways.

And they all have minor differences in terms of their actions and their side effects. So, I'm not sure that you could say that just because drug A gave somebody metabolic syndrome, that drug B would be an inappropriate behavioral health choice to treat their disorder.

Does that make sense?

CO-CHAIR BROOKEY: Um-hum. Yes.

MEMBER SLAVIN: And likewise, someone who's been really difficult to control. You may despite the metabolic effects, just elect to continue to monitor.

I mean, you know, there's always a benefit-to-harms.

MEMBER FINKELSTEIN: So can I just -- that's not what this measure is about though. This measure is only about whether you screened or not.

MEMBER SLAVIN: Correct.

MEMBER FINKELSTEIN: So, if the patient is obese because of their antipsychotic, but you say look, it's worth it because it's the only way to help them. You may -- you still pass the measure.

Unless you say I wouldn't want to know that his glucose --

(Laughter.)

MEMBER FINKELSTEIN: Was enough, hyperglycemic range. So, I don't think that matters.

CO-CHAIR BROOKEY: So, Jon has brought us back to where I think we can vote now on validity. Okay?

So, we're talking about validity. Any other comments before we vote?

MS. CHAVEZ: Okay. Now voting on validity. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Nineteen. Twenty-four.

CO-CHAIR BROOKEY: Twenty-four? Did somebody come back in the room? Oh, Jeff is back.

(Laughter.)

MS. CHAVEZ: Five voted high. Eighteen voted moderate. One low. Zero insufficient. This measure passes validity.

CO-CHAIR BROOKEY: The matters to reflect if you're in the bathroom during the vote. So, I just said.

Okay. So, we're up to feasibility. And I believe that this is going to be similar to the previous measure.

It's going to be -- is that correct? It's an administrative measure. We are on feasibility, right? I'm getting lost.

So, from a health plan perspective, it would be based on coded data. And it should be easy for a health plan.

Any other comments about feasibility for a health plan?

(No audible response.)

CO-CHAIR BROOKEY: Okay. Can we vote?

MS. CHAVEZ: Okay. Now voting on feasibility. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Twenty-one. Twenty-three. Twenty-four. Fifteen voted high. Nine voted moderate. Zero low. Zero insufficient. This measure passes feasibility.

CO-CHAIR BROOKEY: Thank you. Next is usability and use. Any comments that we haven't already made about usability?

(No audible response.)

CO-CHAIR BROOKEY: It kind of relates to everything else. Shall we vote? Okay.

MS. CHAVEZ: Okay. Now voting on usability and use. One high, two moderate, three low, four insufficient. The voting is open.

(Pause.)

MS. CHAVEZ: Fifteen. Twenty-one. Twenty-four. Ten voted high. Fourteen voted moderate. Zero low. Zero insufficient. This

measure passes usability and use.

CO-CHAIR BROOKEY: Okay. So, finally overall suitability for endorsement. Any other comments?

(No audible response.)

CO-CHAIR BROOKEY: Okay. Can we vote?

MS. CHAVEZ: Okay. We are now voting for overall suitability for endorsement for Measure 2800. One for yes, two for no. Polling is open.

(Pause.)

MS. CHAVEZ: Thank you. Twenty-four voted yes. Zero no. Measure 2800 has been recommended for endorsement.

CO-CHAIR BROOKEY: This is your last measure? Great. So, we're up to our last measure. And it's another NCQA measure.

And I don't want the clock to prevent us from making good decisions. So, what I'm going to say is, I think it's feasible to get this done in an hour.

Would you agree? Because I don't

really want to say we're going to finish at 3:00. I think we need to like -- some people need to get to the airport.

I'd like to say that we're going to finish by a quarter to. Maybe even earlier. But if we can't, if we're getting crunched on time, then we have the option of finishing this on a phone call later.

So, I think we can do it. But I don't want us to feel pressured by the clock. Is that reasonable?

So why don't we try to be concise and clear. And we'll move forward. And I'll let you go with a summary if you would.

MS. BYRON: All right. So, some folks have actually raised this issue of off-label use. And we explored a measure like that.

But for various reasons it was difficult to specify given there aren't many FDA prescribed uses for antipsychotics. So, our attempt to address this issue is with this measure, which is the use of first-line

psychosocial care.

It looks at kids with a new prescription for antipsychotic who do not have a primary indication as prescribed by the FDA for an antipsychotic. And looks to see that they had psychosocial care either before or immediately after the prescription for an antipsychotic.

So, where you see most cases of these drugs being described for ADHD or disruptive behaviors for which they are not indicated. For those conditions, psychosocial therapy is really recommended as first-line care.

And so we look to see of that was given.

CO-CHAIR BROOKEY: Thank you. Who's our Subcommittee for this particular measure? Okay. Would one of you like to make some comments?

MEMBER EINZIG: Yes, so the -- so it's a process measure. It's claims data.

Numerator children who have psychosocial care as the first-line treatments

prior to or immediately after a new prescription for an antipsychotic. And the denominator is all children who receive antipsychotic for a condition that does not have an FDA indication.

Looking at it from an evidence perspective is another one of those measures where it was consensus I believe. That was used for evidence.

No other hard core studies that I reviewed in there. The -- most of my other comments will have to do with things other than evidence. So, I'll just hold there.

CO-CHAIR BROOKEY: So, let's limit the conversation to evidence right now. And it sounds like that was a very strong consensus and based on a lot of guidelines.

So, is that correct? Okay. Other comments about evidence? Go ahead Jon.

MEMBER THACKERAY: It was based on a lot of consensus statements. But I think similar to a couple of measures ago, there's a little bit of general wording in them.

So, recommendation one, prior to the initiation of and during treatment with an atypical antipsychotic, the general guidelines that pertain to the prescription of psychotropic medications should be followed. Including education and psychotherapeutic interventions for the treatment and monitoring of improvement.

And the other one is physicians should consider other medication or psychosocial treatments before initiating antipsychotic treatment.

So it's -- I mean, the recommendation is there. But it's not --

CO-CHAIR BROOKEY: Specific.

MEMBER THACKERAY: It's not black and white.

CO-CHAIR BROOKEY: Yes.

MEMBER THACKERAY: And I'm a little worried we're going to get into the same conversation we had with the previous measure about what is the threshold for this? Like, how often, you know, in our Subcommittee we talked

about there are probably scenarios where it's appropriate to initiate pharmacotherapy without waiting for your psychosocial intervention.

And where is that threshold? So --

CO-CHAIR BROOKEY: Okay. Amy?

MEMBER HOUTROW: My question relates to the evidence in relationship to the numerator and denominator. So, a lot of children who are in health plans are receiving psychosocial care outside of that system.

Whether it be in their school or at a community health center. And many plans won't even cover some types of psychosocial.

So, I don't know if this belongs in our evidence discussion. But, do we know what percentage of children would receive all of their mental health care within a health plan system?

Or how many kids are getting it outside of that? Because that could really change the population since in my own practice, a majority of children are not getting their mental healthcare within their health plan.

CO-CHAIR BROOKEY: It's a good point Amy. And of course in California, we have a regional center. I'm sure you all have something comparable in your own States.

And we don't necessarily know if children are getting services from the regional center. And with Medicaid, we actually have carve outs.

And so we don't actually always know what's happening in the county with the health plan carve out. I don't know if -- I think that's true in more of the States other than California.

But, plans do not always know. And so I'm going to ask the developers to comment on that.

DR. BARTON: We are aware that there are a lot of State variables such as you mentioned. And NCQA's point of view here is that if you are a State, you're responsible for the children in your State.

And carving out is not an excuse. And

if you're a health plan, in all honesty, there should be pads in place for bi-directional communication.

And we realize that we're not -- we're not God. And so just because we say it, it's not going to be so.

But on the other hand, in the absence of a measure that's incentivizing and encouraging plans to start behaving that way, we don't think that it will ever happen. So, it's an interesting situation to be in.

And we can appreciate your concern.

CO-CHAIR BROOKEY: I hear what you're saying exactly. But, I have to ask you a question.

This is an administrative measure. So, it may be true that the psychiatrist knows that the child is getting therapy from here. And they may include that in their --

(Simultaneous speaking.)

CO-CHAIR BROOKEY: But you wouldn't know --

(Simultaneous speaking.)

DR. BARTON: With all due respect, that's a validity question I believe, not an evidence question.

CO-CHAIR BROOKEY: No, but I'm just trying to get to the issue that's been raised about, can we actually pull the data?

So, the question on validity we'll put aside for now. The question on evidence, I'm hearing that it's largely consensus based.

And so it's --

DR. SCHOLLE: So this is Sarah. I just wanted to weigh in. This does require that the health plans be responsible for the mental health benefit.

So the carve out situation you mentioned would not fall in that. The health plans who do not have responsibility for the benefits, mental health benefit, wouldn't be responsible.

MEMBER AGORATUS: This is Lauren. I have a clarification question.

Children can also get counseling on their individual education plan as a related service under the Individual Disability Education Act. Would that count?

CO-CHAIR BROOKEY: If you had a B code, I suppose it might. Right? If we were in the --

MEMBER THACKERAY: I just don't believe that's properly been in as administrative claims though.

CO-CHAIR BROOKEY: Yes. If the health plan knew about it and somebody has coded it. But, that's a big if, right?

So, let's -- I agree. Let's go back to the evidence question. And Jill, you have a comment?

MEMBER MORROW-GORTON: I actually had a couple of questions. One was when -- with respect to Jonathan's comment about medication, I think we're only talking about antipsychotics, not other medications.

And antipsychotics for kids that don't

have one of the diagnosis that there's an indication for it.

CO-CHAIR BROOKEY: Right.

MEMBER MORROW-GORTON: So, I think that it would be highly unlikely that you would have a kid that you would just start on an antipsychotic de novo.

MEMBER THACKERAY: I'll defer to that.

MEMBER MORROW-GORTON: Yes. I just think that would be unlikely. My other question was kind of the definition of psychosocial intervention.

You know, what -- sort of what does that include? And my third is a comment about kind of as, I mean, there are all these States do things in all different ways.

And so does commercial insurance. So, behavioral health is often carved out of the medical benefit. And then you come up with DD services and waiver services and what not.

But, I think as you -- as we look at where things are going, things are all moving to

having everything, you know, either ACO or behavioral health integration, or that sort of thing.

Which may strengthen your ability to get this actual data in a more reliable way.

CO-CHAIR BROOKEY: Good comment. Kevin?

MEMBER SLAVIN: My question really had to do with the recommendation two. Where there's also the carve out of substantial evidence for effectiveness in addition to FDA indications.

Not being familiar with these medications and how they might be used outside of an FDA indication, I don't know how -- how does that get captured?

What would be considered substantial evidence that might allow for it to be used in this capacity? And sort of the indications that are specifically listed, are those all FDA approved indications?

And if there are others, how would they or would they not be included in this?

MS. BYRON: So if the -- so, I'll just say, the FDA conditions -- the conditions listed in the measure are the FDA approved conditions.

Molly, did you? Or Sarah, were you trying to go?

DR. FINNERTY: For children, for the kids who are receiving antipsychotics who do not have an FDA indication, it's generally for behavioral control. For agitation, for behavioral control, for sedation, for things like that that can be manifest in any diagnostic condition.

So, we're saying if there is an indication, those kids are carved out. But for those who don't have an indication where it is being used for behavioral control, you want to see that there's some evidence that first-line psychosocial interventions were tried.

And I think it's important to talk about two different kinds of evidence here. And one is the evidence for a treatment sequence.

You know, so the first piece of

evidence is who has a first-line indication? And that we're relying on the FDA where evidence is substantial.

And those who do not have that, the evidence is less. Right, so those are the kids we're talking about.

The second is the evidence for a sequence of treatments where you try psychosocial before you try a medication that does not have a primary indication that carries with it some risks for children.

And these are the kinds of things that come out of the guidelines and treatment recommendations. And you won't see one stronger than that.

Like if you see a treatment algorithm, it's not going to be any stronger then you should try this before you try that. You know, or you should consider all these other alternatives before you go to that.

That's kind of as strong as those treatment sequence recommendations will get.

But, the other piece of evidence is, what is the evidence for the effectiveness of psychosocial treatments to manage behavioral, maladaptive behaviors?

And there, there is good evidence. And those guidelines cite the evidence for psychosocial interventions across an array of diagnosis.

And those are recommended treatments. So, the point is, if we have evidence for recommended treatments, why would we jump to a non-recommended treatment?

CO-CHAIR BROOKEY: Yes. Thanks for your comments. Although it is about sequencing and whether the evidence is to require it before.

And that's what this is measuring. So Jeff, you have a comment?

CO-CHAIR SUSMAN: It was just a question. It was on the lines you were following.

Is this a blanket exclusion? So, if any of the antipsychotics are approved say for

bipolar disorder, then all of them are approved?

Or is it a specific drug by drug evaluation of what the FDA approval is? And if it's that, I have a lot of problems with it.

If it's the former, then I'm probably not so.

MEMBER EINZIG: It's drug by drug.

CO-CHAIR SUSMAN: It's drug by drug. Okay. Well, you know, the willy-nilly approach to approving children's medications and children's mental health medications gives me some pause.

I mean, I think it challenges the validity at some levels. And while I guess it's a stricter standard just because Zyprexa's been approved for A and another drug has been approved for B, I think is parsing this rather fine, just based on how a pharmaceutical company believes its profits can be maximized.

DR. FINNERTY: And I think that's an excellent, excellent point. And that is why this measure actually takes antipsychotics as a class.

And it says if even one antipsychotic has an FDA indication, all of them have a pass. So they would not --

CO-CHAIR SUSMAN: Okay. So, I'm hearing some discrepancy from -- in the Committee and -- but is it true then, so if one of the antipsychotics has an indication say for a tic disorder, then all -- and David's shaking his head no.

DR. FINNERTY: That's, I mean, let's see, I don't have the measure specifications in front of me. But that's my understanding of how it was specified.

That for any tic disorder, schizophrenia, bipolar disorder, any psychotic disorder, those children are just taken out of the denominator.

CO-CHAIR SUSMAN: Okay.

CO-CHAIR BROOKEY: All right. That makes sense. So, let's limit our comments to -- right now to the evidence. And we'll get into validity and everything else later.

So, David?

MEMBER EINZIG: I don't even remember anymore.

(Laughter.)

CO-CHAIR BROOKEY: Ricardo?

MEMBER QUINONEZ: Yes. So I just -- I think this is like a really good measure. Because it takes -- it's using a non-pharmacological approach to treat something that has been shown to be -- that you can actually improve with a non-psych -- you know, pharmacological approach.

And so I don't think, speaking to the evidence, I don't think we're going to find that this is supported by strong evidence. But I think this is clearly one of those where the evidence might be insufficient.

But, it's a very good indicator.

CO-CHAIR BROOKEY: So you're saying it's a valid measure.

MEMBER QUINONEZ: Very valid.

CO-CHAIR BROOKEY: But, we -- you're

suggesting that if we don't think the evidence is strong enough, it might be strong enough to vote for insufficient with --

MEMBER QUINONEZ: Correct.

CO-CHAIR BROOKEY: With exception. Is that what you're suggesting?

MEMBER QUINONEZ: Yes.

CO-CHAIR BROOKEY: Are there any other comments about evidence before we make our decision? So, the people have their suitcases behind their chairs I see. So, go ahead Dave.

MEMBER EINZIG: So, you know, just to throw this out there. So, I would agree that it's insufficient evidence with exception for me.

But the, you know, the bigger question is, is there evidence to say that it's inappropriate to move forward with a medication for an unapproved condition? Say we label that mood disorder NOS instead of bipolar, which is probably the majority of the kids out there.

Because bipolar is very difficult to diagnose. But the medication still would be

appropriate.

Is it, you know, is it an unfair measure in that way to imply that it's inappropriate to move forward with medication without the interventions?

CO-CHAIR BROOKEY: Okay. So, I'm going to have more questions about validity. But in terms of evidence, do we think we have enough information to vote on evidence? Okay.

MS. CHAVEZ: Okay. Voting for evidence for Measure 2801 is now open. One high, two moderate, three low, four insufficient.

(Pause.)

CO-CHAIR BROOKEY: You cannot vote high I think. Right?

MS. CHAVEZ: Fourteen. We're expecting 24 votes. Eighteen. Twenty-three. Twenty-four.

Zero voted high. Seven voted moderate. Four voted low. Thirteen voted insufficient. Consensus was not reached.

CO-CHAIR BROOKEY: So I would propose

that we vote for insufficient with exception. Should we vote for that? Or should we decide to vote on that?

We should just vote for it.

MS. BURSTIN: You could decide to vote. It will be cleaner in the report.

CO-CHAIR BROOKEY: We're going to vote for insufficient with exception or not.

MS. CHAVEZ: Okay. All right, now voting on evidence. Yes for insufficient with exception. Two no for no exception. Voting is open.

(Pause.)

MS. CHAVEZ: Twenty-one. Twenty-four. Twenty-one voted insufficient evidence with exception. Three voted for no exception.

CO-CHAIR BROOKEY: Okay. So, we move forward to performance gap. Any of the Committee members want to comment on performance gap?

MEMBER THACKERAY: Yes. So the developers cite a study of Medicaid in rural children showing that a third of youth starting

an antipsychotic were not receiving concurrent psychosocial therapy.

It showed some disparities in adolescents compared to younger children. They also did their own analysis of Medicaid data from 11 States and found that a little less than half of children on an antipsychotic were -- had documented psychosocial care.

CO-CHAIR BROOKEY: Any other comments on gap?

(No audible response.)

CO-CHAIR BROOKEY: Should we vote on gap? Okay.

MS. CHAVEZ: Okay. Now voting on gap for Measure 2801. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Fourteen. Twenty-one. Twenty-three.

CO-CHAIR BROOKEY: We lost another couple of people, didn't we?

MS. CHAVEZ: Okay.

CO-CHAIR BROOKEY: We lost Jeff and Jim.

MS. CHAVEZ: Twenty-three votes. Okay, we received 23 votes. Five voted high. Sixteen voted moderate. Two voted low. Zero for insufficient.

This measure passes performance gap.

CO-CHAIR BROOKEY: Okay. We started to talk about reliability and validity. I think we made some comments about data collection. It's an administrative measure.

And I think Amy raised a concern about where the care might be given. Can you -- can the health plan capture that data?

So, to someone else's point, if it's a complete carve out, then the prescriber of the medication may be the same provider as the therapy.

If it's a partial carve out/carve in situation, you may still have that issue with therapy being given over here and medication over here.

So, other comments about reliability of this particular administrative measure? Carol?

MEMBER STANLEY: Yes. In seeing health plans in action and calculating some measures, they want to do well in these measures.

And so, they get very diligent and persistent with States and other contractors, other providers to identify supplemental data sets that their HEDIS auditors will approve of them using.

So, I really don't have many concerns about the need for supplemental data and identifying those opportunities.

CO-CHAIR BROOKEY: Okay. David?

MEMBER KELLER: So, I'm just cognizant that when people have looked at who is prescribing the antipsychotic medications in Medicaid, actually it's primary care providers more the psychiatrists who are prescribing them.

So, the issue of the separate systems is actually quite -- is more prevalent then I

think you implied in your previous statements. It's not the psychiatrists who are part of the behavioral health carve out who are often prescribing these things.

So, to your point though, it would require people to be pretty aggressive about being able to cross those data lines. I do think that's an issue.

I guess my question for the developer is did you find that to be an issue when you were testing these measures?

MS. BYRON: Well, because of some of those concerned, we do require the mental health benefit for the health plans. You know, hoping that that will help.

So, -- right.

CO-CHAIR BROOKEY: And I will clarify that for California, we cannot get the data from the State or the county or the regional center at a granular enough level for supplemental data. It's just --

MEMBER KELLER: Yes. And nor can we

in Colorado actually. We can't access behavioral health utilization data easily because it's all carved out at the -- it's actually at the BHO level.

It's not even at the county level.

CO-CHAIR BROOKEY: Yes.

MEMBER KELLER: And then subcontracted to the counties, making it even more complicated.

CO-CHAIR BROOKEY: David?

MEMBER EINZIG: So, I think this is an appropriate place for me to voice a few concerns. Hopefully it is. Cut me off if it's not.

First, -- my first concern is looking at, you know, are we improving quality of care by recommending therapy first before medications? And you know, as far as validity goes, I think we might be missing the mark on that one again.

You know, for an example, if a person can't get in to see the psychologist or counselor or tie into those services for several months, and you're not moving forward with medication, into September. You know, that whole school year

is gone basically.

And so you're altering that developmental trajectory which can worsen quality of care. So, the risk for not treating are there.

I think this is a bigger issue then looking at are we over-prescribing. I think the bigger issue is, the reason why we're prescribing these medications in some cases for, you know, foster care, reactive attachment, those type of things, you know, I think the common thread is they aren't getting the early intervention at a younger age.

And so we're trying to improve quality of life through the use of medications to target these instant reactions and whatnot. When the bigger picture is, the role for early intervention services, which this measure doesn't really address.

CO-CHAIR BROOKEY: Good comment. Kerri?

MEMBER FEI: So from the health plan

perspective, I think a couples -- others have already said it.

The measure does require you to have medical, mental health and pharmacy benefit included within the product. So, if you don't have those, it's kind of a non-starter.

Secondly, in my discussion with a lot of plans, they had become way more diligent if they do have a carve out, about being able to get their data. And have that data sharing worked into their contracts.

I know it's harder for those that do have Medicaid plans. Because individual by State, like you've highlighted, privacy issues are getting in the way and other things. Obviously they can't get the data.

But, for those that do have the control, they become much more proactive in getting their data.

CO-CHAIR BROOKEY: Yes. If you have a carve out to another behavioral health plan, that's one thing. If you have a carve out to the

county, a.k.a. black hole, that's a difference.

(Laughter.)

CO-CHAIR BROOKEY: But your point is well taken. There are some plans that do have carve outs to another behavioral health plan.

And that's a different situation all together. I agree.

MEMBER FEI: It's the complete difference --

CO-CHAIR BROOKEY: Yes, exactly. Carol?

MEMBER STANLEY: So, along the same lines, this is another measure we're using in a foster care study. And we do carve out behavioral health.

So, we have an external organization that's taking our behavioral health data along with the managed care data and doing the study. But, with the way this measure is, our health plans would not be reporting it directly to NCQA.

CO-CHAIR BROOKEY: Yes. It's going to vary State by State and so health plan by health

plan.

Amy? Oh, to Jon. I'm sorry.

MEMBER FINKELSTEIN: Yes, so look, I just --

MEMBER HOUTROW: I can be quite for once.

CO-CHAIR BROOKEY: Amy Finkelstein.

MEMBER FINKELSTEIN: It's hard to always just disentangle reliability from validity. But just to focus on reliability for a minute.

Someone should say that the reliability testing using statistical approaches was quite reasonable at the State level and for the large health plans. Six of the 19 commercial plans had insufficient denominators.

So it's again, this issue of -- at least by the numbers, the reliability was okay. I've comments too on validity. But I'll hold them.

CO-CHAIR BROOKEY: Okay. Well hold on validity. Let's limit our comments to

reliability. And then we'll go for a vote.

So, Jill? Or is that David?

CO-CHAIR SUSMAN: That's David. But he's already -- he just has it turned up already.

CO-CHAIR BROOKEY: Ricardo?

MEMBER QUINONEZ: So, I think this is a case of, you know, the perfect being the enemy of the good. Because you know, I hear what people are saying around the room about this being hard probably to measure from a reliability standpoint.

But, the question I think we should ask for, because I feel this is such a beneficial measure for children's health, is it impossible? And I don't think we can answer that question yet.

I think this is one of those instances in which a measure may drive improvement.

CO-CHAIR BROOKEY: Are we ready to vote for reliability? Oh, --

MEMBER HOUTROW: I mean, I --

CO-CHAIR BROOKEY: You knew Amy was

going to go next.

MEMBER HOUTROW: One of the things that I've been feeling, is that we all feel very passionate about the issues. And while we might feel passionate about the issues, what we're supposed to judge is not the importance of the issue.

But the how what is presented to us meets specifications that were put forth by NQF. And so, I mean I think that I agree with Ricardo that this is an important issue.

And we shouldn't let perfect be the enemy of the good. But we still have an obligation to evaluate based on a certain subset of data.

CO-CHAIR BROOKEY: But I kind of want to vote for it just because Ricardo likes it. But I won't do it or base it on that.

(Laughter.)

CO-CHAIR BROOKEY: But I agree. I think that we have called out the barriers. It doesn't mean that they're insurmountable, right?

All that we're doing is calling out the barriers. It doesn't mean that we don't think it's important.

It doesn't mean that we don't think that we can overcome the barriers. And there is some testing that would indicate that in some areas it can be reliably obtained.

So any other comments before we vote on reliability?

(No audible response.)

CO-CHAIR BROOKEY: Okay.

MS. CHAVEZ: Okay. Now voting on reliability for Measure 2801. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Twenty-four. We're expecting 24 votes. Okay.

CO-CHAIR BROOKEY: It's probably moderate. Yes.

MS. CHAVEZ: Four voted high. Eighteen voted moderate. Two voted low. Zero

insufficient. This measure passes reliability.

CO-CHAIR BROOKEY: Okay. So let's go to validity. And Jon, you have some comments?

MEMBER FINKELSTEIN: Yes, so I think you can think of validity for this in two ways. And it's where I really get hung up.

So, the measure developer did solid validity testing within the body of claims data that they had. And in that they did correlations between measures which weren't that strong.

But, I'd love the developers to correct me on that. They were okay. And they did correlations by ranking to see if the plans that ranked high in one ranked high in other things you'd expect them to rank high in.

If they were good at one, they should be good at the other. That kind of internal validity testing.

I have to say here's my major concern. There are some things I'm willing to take on face value. I love claims-based measures. I really do.

That I'm willing to take on face value that claims based do well almost all the time. Right?

You have -- you're a woman of a certain age. You either get a mammogram, or you don't. I think -- I wouldn't look for any validation besides that for whether you got a mammogram.

These are complex kids with complex stories. They didn't just wake up on Thursday afternoon and need an antipsychotic.

They have histories --

(Laughter.)

MEMBER FINKELSTEIN: They have histories. I'm not making light of this Dave.

Histories of difficulties and counseling at school, at home. Different mental health providers. And at some point they get put on a medicine.

The kind of validity I would have liked to see, is to see whether the claims, the snapshot, this biopsy of time that we do in

claims, actually reflects the story of these children. I would have appreciated it in 20 charts.

I would have appreciated it in 100 charts to see whether the claims-based information got it right. Or when you really read the story, whether it just didn't line up.

And I wonder if the developers thought about that? Or did any of that that wasn't included?

DR. SCHOLLE: This is Sarah. I do want to say, we did consider that. So, as alluded to those were just huge barriers in getting access to all of the records that we would have needed to really answer your question.

Did they get the services at school? That would have required a HIPAA, you know, authorization.

So, it's a lot harder to do. And it was something that we considered. Of course, there is various ways to look at validity that we've presented you data on validity that we

could present.

But, your point is well taken. Molly, in the work that you've done in New York, do you have any insight on this?

And actually, a point that I wanted to make, but I'm not sure it's clear, is that this set allowed for sacrificial intervention to occur 90 days before and 30 days after. So it does allow for, you know, urgent prescriptions with an expectation of getting kids into care quickly.

So, I'll defer to Molly to see if she has any insight on the issue about the claims and validity.

DR. FINNERTY: Yes. I think to the heart of this question is, you know, is this really measuring something that there's a problem with?

Are we moving to medications before exhausting therapy? Especially when the therapy has evidence to support it and the medication does not.

And the answer to that is yes. Things

that look at trends in access to services over time have documented increasing use of antipsychotics even in very young children.

And simultaneously decreasing trends in therapy. You know, so these are trends over time. They were long tracks of time.

Another thing that I can share with you is that at least in our State, there was a study of children in foster care. And you know, we can say well children of foster care by definition have experienced a trauma in their life.

It took them away from their parents. And placed them in foster care in the first place. So, I think it was something like a third of those kids had not had any counseling of any kind.

And that was really a study done internally trying to improve the quality of care and access to care, appropriate of care given to those most vulnerable of children here in the State.

And we have some other studies have suggested children in foster care get more access to psychosocial services then kids not in foster care. So, the question, do we have a problem?

I think the, you know, the answer there is yes. We do. And the nice thing about a measure like this is, I think others have said, is we really need to focus on the heart of the problem.

The heart of the problem is perhaps overuse of these medications without first exhausting recommended treatments. And that's really what the point of this measure is.

CO-CHAIR BROOKEY: So we voted on this to move forward with insufficient evidence. So we have to kind of put that aside a little bit.

And then talk about whether or not the clinicians believe this is a -- there is validity. And it relates to reliability.

So, let's focus on the question at hand in terms of validity. Other comments about whether we can move this forward on validity.

So, David?

MEMBER EINZIG: Yes. So, I think it's really complex. Because there's going to be those individuals where you do what to do therapy first.

But, there's also going to be those individuals where you send them to a therapist's office without a medication onboard, without their moods being regulated, it would be an absolute joke. That the kids are just going to be too emotionally off to let that therapy to be effective.

And if that's the case, unintended consequence of that therapist sucks. And I'm not going back.

And then the, you know, the school issue. Of if you hold off on something that can be potentially helpful that you try everything else first, therapy, psychosocial interventions.

You move them into a smaller classroom setting that's EBD based with a bunch of oppositional kids, when their issue may not be

that they're primarily oppositional. But rather that they just have this difficulty with severe, impulsive, reactive behaviors.

But you get that better regulated, all of a sudden they can stay in their mainstream classroom setting. And go along a different trajectory.

So I think it's a very complex issue. And it's not straightforward at all.

CO-CHAIR BROOKEY: Can I get a clarification? If you do get a kid that comes into the ER. Let's just say full-blown mania.

Is that an exclusion if they start somebody on meds in the ER during an acute episode?

DR. FINNERTY: They have a month to get the child into services.

CO-CHAIR BROOKEY: Okay.

DR. FINNERTY: And I think, you know, some of the examples that we're talking about, the need for medications in acute situations or for kids to properly engage in therapy, you know,

this says you can start -- you should have the therapy either before or within 30 days of starting.

You know, and the idea here is that there is a way that kids, if they need it urgently, can get the medication.

But making this a State measure or a plan measure would incentivize States and plans to ensure that there's access. Because we're talking about well, what if there's no access to that therapy?

It would ensure that there was access. It would promote access to those services.

CO-CHAIR BROOKEY: It actually ties into the earlier measure about 30 day access for services. And so that makes me feel a whole lot better.

I missed that in my reading of the measure. I think Jill had a comment.

MEMBER MORROW-GORTON: So I'm going to play devil's advocate to David here. I know he's sitting right beside me.

But, as I was listening to what you were saying, one, if you get a kid who's manic, I think that's probably possible bipolar disorder, it doesn't count. Two, the kid that you're describing in terms of going into school, behavior being out of control, et cetera, et cetera, as a clinician, I would do both at once.

So you would count -- it would count if you did both at once. I wouldn't do medication and then later do the other.

I would -- I mean, when kids come to see me I recommend both.

CO-CHAIR BROOKEY: Yes. Ricardo?

MEMBER QUINONEZ: I think the point that I was making has been made. But just to say it another way, this wouldn't be a zero sum game. Right?

You wouldn't expect 100 percent compliance with this I assume? Because there's no benchmarks for -- only for this. So, they would be established as we go on.

So, those instances in which

medications are used in the acute setting may end up being asked is this operationalized? Being an exclusion, et cetera.

I don't think this will be one where we expect 100 percent compliance.

CO-CHAIR BROOKEY: Yes. And we know we have a performance gap. And we know as well that reassures me that we have a 30-day window there.

It makes it a little bit of a difference. Let see, are we getting close to reaching our conclusion? Jeff?

MEMBER SCHIFF: Just a very quick point that I worry. Which I think relates a little bit to the validity.

Is I worry about the diagnostic accuracy and the variability. And we're layering a measure on top of a great amount of diagnostic variability.

CO-CHAIR BROOKEY: Amy?

MEMBER HOUTROW: My question was to I agree clinically you would try to do both. But

it's not that at the same time you made a referral.

It's that they actually achieved the care within a month? Is that right? Because that is somewhat outside of the practitioner's realm.

It's a health care measure. But let's say I work in a health system for which access takes months. And I refer that child to their school counselor.

And that won't be -- it's the same issue for me about how you actually capture the data that you need to capture.

CO-CHAIR BROOKEY: Yes. Can I respond to that from a health plan -- this is a health plan measure. It is the responsibility of the health plan to provide access.

So, it's the accountability is at the health plan level. And I have to say that that is my accountability as a health plan leader.

So, is that a correct statement?

MEMBER FINKELSTEIN: Yes.

CO-CHAIR BROOKEY: Okay. David?

MEMBER EINZIG: Real quick. I promise. So, I agree with Gerald that it has to be both at the same time, Medications plus therapy.

And yes, that is and should be and is the standard of care to do both. But this measure is saying exhaust therapy before medications.

When I think it would be a better measure is if it were medications in conjunction with therapy. Or therapy in conjunct -- the two together. Not necessarily one exhausted before the other.

MS. BYRON: I -- okay.

CO-CHAIR BROOKEY: Yes, but just to clarify, it sounds to me that the medication can proceed the therapy.

MS. BYRON: It -- yes. Yes.

CO-CHAIR BROOKEY: There's a 30-day window.

MS. BYRON: There is a 30-day window.

And I think the -- you know, the concern with some -- with a measure that says use both, you may inadvertently encourage medication use.

When we really don't want to see medication use in these kids who do not have an indication. So, ADHD, disruptive behaviors, we're really not looking for medication use.

But the measure's bar is well, if you're going to have them on there, I want to see that you've tried psychotherapy first or immediately after.

CO-CHAIR BROOKEY: So, any other comments about validity? Jon?

MEMBER FINKELSTEIN: So, this will be really quick. Because I feel that it's easy for the conversation to get away from validity.

So, just to tick them off, I don't think anyone questions that this is a big problem. That there's evidence that it's a big problem.

I don't think anybody questions -- well, I don't question that you should try

psychotherapeutic approaches before or with medication approaches.

So, there's no argument on the thing. The validity question is whether the measure accurately measures the thing.

MS. BYRON: Okay.

MEMBER FINKELSTEIN: And I just think we have to keep those two things separate. And I'm okay. If we think claims is close enough and feasible and it measures the thing, then fine.

But that's the question.

CO-CHAIR BROOKEY: Yes. So, I think when we vote, we have to consider all the barriers that we've raised. We know that this is not perfect.

This is not a perfect measure. Is that a reason for it not to move forward?

So you have to consider all those things when you decide to vote on validity. Because we have -- this is a must pass vote.

So, can we vote on validity? Okay.

MS. CHAVEZ: Okay. Now voting on

validity for Measure 2801. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: We're looking for 24 votes. I see 23. Twenty-four. Zero voted high. Thirteen voted moderate. Seven voted low. Four voted insufficient.

And this measure passes validity. Sorry, it puts us in the gray zone.

CO-CHAIR BROOKEY: With 13 out of --

MS. CHAVEZ: Thirteen out of 24.

CO-CHAIR BROOKEY: Oh, that's below 60 percent. Yes, okay. So, it's gray zone. But that's --

MS. CHAVEZ: Gray zone.

CO-CHAIR BROOKEY: That's still okay. All right. So the next one is feasibility. And I think we've actually already talked about this.

Because we've talked about this as a claims-based measure with a caveat that it may be difficult for the plans to collect supplemental

data. I don't -- I'm not sure, is there anything else we should say about that?

(No audible response.)

CO-CHAIR BROOKEY: A recommendation to vote? Okay. We're too quick for you. I know.

MS. CHAVEZ: Okay. We're now voting on feasibility. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Twenty-one. Twenty-three. Twenty-four. Okay. Six voted high. Twelve voted moderate. Five voted low. One voted insufficient.

This measure passes feasibility.

CO-CHAIR BROOKEY: Okay. Usability and use. So we've agreed to pass this with insufficient evidence with exception.

We believe there's a care gap. We believe that it's reasonably reliable, reasonably valid. That we believe that we can measure it.

So, is there much more discussion about usability is my question?

(No audible response.)

CO-CHAIR BROOKEY: No? If not, we can vote.

MS. CHAVEZ: Okay. We're not voting on usability and use. One high, two moderate, three low, four insufficient. Polling is open.

(Pause.)

MS. CHAVEZ: Twenty-three. Twenty-four. Four voted high. Thirteen voted moderate. Six voted low. One voted insufficient.

This measure passes usability and use.

CO-CHAIR BROOKEY: Okay. So we're one vote away from the group hug.

(Laughter.)

CO-CHAIR BROOKEY: Overall suitability for endorsement. And then we'll do public. Don't leave until we have public comment.

So, all right.

MS. CHAVEZ: Okay. We're now voting for Measure 2801 and its overall suitability for endorsement. One for yes, two for no. Polling is open.

(Pause.)

MS. CHAVEZ: Twenty-three. Twenty-four. Seventeen voted yes. Seven voted no. We needed 15 votes to pass.

So, this measure has been recommended for endorsement.

CO-CHAIR BROOKEY: So I think the developer has heard our comments. Are there any other comments that have not been made for the developer before we wrap up?

(No audible response.)

CO-CHAIR BROOKEY: Okay. We're going to move to public comment. But I would just like to say that it's been a pleasure working with all of you.

I had no training for the job. And it's been on the job training. And --

(Applause.)

CO-CHAIR BROOKEY: And it's been a pleasure to work with Jeff.

CO-CHAIR SUSMAN: Thank you, sir.

CO-CHAIR BROOKEY: And I have to say

--

CO-CHAIR SUSMAN: I'll shake your hand.

CO-CHAIR BROOKEY: I'll shake your hand. Really, you know, I think we all recognize that the NQF Staff does do a tremendous job.

And they're not through. We're not through yet. We have more to go. More phone calls to go.

But, I'm going to turn to the public comment. And then turn it to Robyn to wrap up.

DR. NISHIMI: Operator?

OPERATOR: Yes, ma'am. At this time if you'd like to make a public comment, please press star one.

And there are no public comments at this time.

CO-CHAIR BROOKEY: In the room?

DR. NISHIMI: In the room? I don't think so. I just wanted to show you one last slide before we go over the -- what the schedule is.

This the result of really what has been absolutely stellar work. It's always an interesting prospect when a Committee meets for the first time.

And we have some Committees that meet multiple times. And this is really excellent. And I've seen a lot of Committees in my time at NQF. So, thank you very, very much.

You approved one, two, three, four, five, six, seven measures. One of which had the multiple components of fact managers.

There was one measure for which consensus was not reached. It will go forward in the draft report that goes out for comment with that designation.

And you all will revisit it. And then there were eight measures that weren't recommended and the two FECC components that were not recommended.

The developer does have the opportunity to bring back information that may sway you for those. And that would be part of

the post draft report comment call.

So, thank you very much. Nadine, do you want to go through the dates for the next -- we won't have to have the post-meeting call because you finished your work.

But there are some other key dates coming up.

MS. ALLEN: So, we're not having our post-meeting conference call, which was scheduled for December 10. However, we go out for public and Member comment from January 14 to February 12.

And then we'll have our post draft report comment call where we'll bring back all the comments to you. As well as any additional information we might receive from the developers for you too probably revote on the measures or not.

We'll have Member voting starting from March 15 to March 29. CSAC review on April 12. Board review on May 12. And appeals from May 16 to June 14.

DR. NISHIMI: You will receive a copy through the PEDS mailbox of the draft report. You know, so you have that as a marker of what's happened.

Is there anything else the Committee would like to give us feedback on?

CO-CHAIR SUSMAN: Yes. The one thing I would note is clarifying the issues in the green boxes and down below.

And then secondly, making it explicit when votes should be held for an exception with insufficient evidence. And how you count low and insufficient evidence together.

That wasn't entirely clear to us I think as we went through the process.

DR. NISHIMI: All right. Thank you. Ricardo?

MEMBER QUINONEZ: So, yes. Thank you for this. This was a great process. And it makes me so much more comfortable now accepting quality metrics that are endorsed by NQF.

And I would like to challenge NQF

though that because, you know, my -- one of my hobbies is to beat a dead horse.

(Laughter.)

MEMBER QUINONEZ: It would be to -- going back to the measure of where we were testing, and adequately so everybody stated this is for monitoring, not for et cetera, et cetera.

I would challenge NQF that when somebody presents a test that they require things like number needed to test. Number needed to harm.

Because unintended consequences from testing occur. And it's everywhere in the literature.

And I think -- I have no problem with getting tests. Because you think you're being safe. Because you think you're monitoring something.

But I do have a problem endorsing tests for what is supposed to be a quality measure that is supposed to improve children's health. I think the bar should be higher.

DR. NISHIMI: Thank you. Anyone else? Safe travels.

CO-CHAIR BROOKEY: Okay. Thank you.

(Whereupon, the above-entitled matter went off the record at 2:38 p.m.)

G

C E R T I F I C A T E

This is to certify that the foregoing transcript

In the matter of: Pediatric Measures
Steering Committee

Before: NQF

Date: 12-02-15

Place: Washington, DC

was duly recorded and accurately transcribed under my direction; further, that said transcript is a true and accurate record of the proceedings.

Neal R Gross

Court Reporter