NATIONAL QUALITY FORUM

+ + + + +

BEHAVIORAL HEALTH STEERING COMMITTEE

+ + + + +

WEDNESDAY
APRIL 18, 2012

+ + + + +

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

PRESENT:
PETER BRISS, MD, MPH, Co-Chair
HAROLD PINCUS, MD, Co-Chair
CAROLINE CARNEY-DOEBBELING, MD, MSc, Medical Officer, MDwise, Inc.
MADY CHALK, PhD, Director, Treatment Research Institute
DAVID EINZIG, MD, Children's Hospitals and Clinics of Minnesota
NANCY HANRAHAN, RN, PhD, University of Pennsylvania
DOLORES KELLEHER, MS, DMH, Principal, D. Kelleher Consulting
PARINDA KHATRI, PhD, Director, Cherokee Health Systems
TAMI MARK, MBA, PhD, Senior Director, Thomson Reuters Healthcare, Inc.
BERNADETTE MELNYK, RN, CPNP, PhD, Dean, The Ohio State University College of Nursing
MADELINE NAEGLE, APRN-BC, PhD, FAAN, Professor, College of Nursing, New York University
DAVID PATING, MD, Chief, Kaiser Permanente Medical Center

KARLENE PHILLIPS, BSN, RN, Director, Mayo Clinic Health System
VANITA PINDOLIA, PharmD, HFHS/HAP Vice-President Ambulatory Clinical Pharmacy Programs, Henry Ford Health System
JEFFREY SAMET, MA, MPH, MD, Chief, Department of Medicine, Boston University
LISA SHEA, MD, Associate Medical Director, Butler Hospital, Providence, RI
JEFFREY SUSMAN, MD, Dean, Northeast Ohio Medical University
LYNN WEGNER, MD, Clinical Associate Professor, UNC Department of Pediatrics
BONNIE ZIMA, MD, MPH, Professor-in-Residence, UCLA Department of Psychiatry and Bio Behavioral Sciences
LESLIE ZUN, MD, Chair, Mount Sinai Hospital

NQF STAFF:
HELEN BURSTIN, MD, MPH
SARAH FANTA
ANGELA FRANKLIN-HOLBERT, JD
SARAH LASH
EVAN WILLIAMSON, MPH, MS

ALSO PRESENT:
KYLE CAMPBELL, Florida Medical Quality Assurance, Inc.
MARCELA HORVITZ-LENNON, RAND Corporation
SARAH HUDSON SCHOLLE, National Committee for Quality Assurance

T-A-B-L-E O-F C-O-N-T-E-N-T-S

Welcome
Peter Briss . . . . . .6

1879: Adherence to Oral Antipsychotics for Individuals with Schizophrenia
Centers for Medicare and Medicaid (CMS) . . .7

1935: Use of Antipsychotic Medication
National Committee for
Quality Assurance . . . . . 54

1936: Continuity of Antipsychotic Medications for Treatment of Schizophrenia
National Committee for
Quality Assurance . . . . . 57

Related and Competing Measures
Discussion. . . . . 119

1926: Cervical Cancer Screening for Women with Schizophrenia
National Committee for
Quality Assurance . . . . . .139

1932: Diabetes Screening for People with Schizophrenia or Bipolar Disorder Who Are Prescribed Antipsychotic Medications
National Committee for
Quality Assurance . . . . . 154

1927: Cardiovascular Health Screening for People with Schizophrenia or Bipolar Disorder Who Are Prescribed Antipsychotic Medications
National Committee for
Quality Assurance . . . . . .210

1933: Cardiovascular Health Monitoring for People With Cardiovascular Disease and Schizophrenia
National Committee for
Quality Assurance . . . . . 267

T-A-B-L-E O-F C-O-N-T-E-N-T-S (Continued)

1934: Diabetes Monitoring for People with Diabetes and Schizophrenia
National Committee for Quality Assurance . . . . . 287

1938: Emergency Department Utilization for Mental Health Conditions by People with Schizophrenia
National Committee for Quality Assurance . . . . . 297

0576: Follow-Up After Hospitalization for Mental Illness
National Committee for Quality Assurance . . . . . 310

1937: Follow-Up After Hospitalization for Schizophrenia (7 and 30-day)
National Committee for Quality Assurance . . . . . 310

Next Steps/Timeline for Project
Sarah Fanta, Project Manager, Performance Measures. . . . . .340

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

8:33 a.m.

MS. FRANKLIN: Hello and welcome to day 2 of the behavioral health project meeting, our in-person meeting. And we have here with us our Co-Chairs Peter Briss and Harold Pincus, and I'll go ahead and hand it over to them.

Okay. So today we have a couple members of the Committee that we didn't have with us yesterday. And we'd like to go around and have them introduce themselves and also announce any conflicts that they may have. And first is Dr. Zun.

DR. ZUN: Good morning. Les Zun, professor and chair of Department of Emergency Medicine at Chicago Medical School, as well as Mt. Sinai Hospital. Conflicts. I sit on a number of boards. The American Academy for Emergency Medicine, the American Association for Emergency Psychiatry, the Illinois College of Emergency Physicians and American College

of Emergency Physicians, Practice Management Committee and a consultant for Alexza Pharamceuticals, which has no product on the market at this time. So I think -- did I get everything? All right. Thank you.

MS. FRANKLIN: And Madeline Naegle, I know you were on the line yesterday, but could you reintroduce yourself and disclosures or --

DR. NAEGLE: I'm Madeline Naegle. I'm a professor at the College of Nursing at New York University. I oversee our substance-related disorders educational tracts there and I'm co-investigator of Project SARET, Substance Abuse Education Research and Training, with our NYU Medical School. I have no conflicts.

CO-CHAIR BRISS: So good morning. The agenda says I'm supposed to do a recap of yesterday. I have no recap except thanks to everybody for a lot of hard work and I think we can get right to the first measure

evaluation.

DR. BURSTIN: So the first measure we have on the schedule for today is No. 1879: Adherence to Oral Antipsychotics for Individuals with Schizophrenia. Our lead discussant for this measure was Dr. David Einzig. Before we start discussion, however, we'll have the developer tee up the measure for us and then we'll begin discussion.

MR. CAMPBELL: Good morning. My name is Kyle Campbell and I'm project director for FMQI and we are a contractor with CMS for this particular measure.

Would you like me to give a description?

(No audible response.)

MR. CAMPBELL: Okay. So this measure is really looking at adherence to oral antipsychotics for beneficiaries with schizophrenia. The threshold that we use for the measure is 0.8 and the algorithm we use is a proportion of days covered methodology which

we find found from the literature to be the best approach for medication classes for which there's frequent switching and overlap. And we have also harmonized the methodology that is in this measure with the other adherence measures in the CMS portfolio, as well as the Pharmacy Quality Alliance.

DR. BURSTIN: Thanks. Dr. Einzig.

DR. EINZIG: Okay. So numerator folks are people with schizophrenia who have filled more than two prescriptions and have a proportion of days covered of greater than 0.8. Denominator, all adults with schizophrenia with at least two claims for antipsychotics.

This is a process study. Obviously in terms of impact I think this is fairly straightforward because we know in folks with schizophrenia compliance is often an issue and poor compliance often leads to hospitalizations.

Should we move forward to

opportunities for improvement?

CO-CHAIR BRISS: I think we can talk about it in both, right? So any comments about impact?

(No audible response.)

CO-CHAIR BRISS: Yes, hearing none.

CO-CHAIR PINCUS: Oh, just for the new members, the way you vote is with these things.

CO-CHAIR BRISS: And your voting options are on the back also. High, moderate, low, insufficient.

MR. WILLIAMSON: We will now be voting on impact. Begin voting now. We're waiting on two responses. If everybody could please vote again.

Okay. The measure was 16 high, 3 moderate.

DR. EINZIG: Okay. I move now for opportunity for improvement. I think there is a performance gap. Lots of studies document

poor medication compliance in folks with schizophrenia. This document alluded to lots of studies showing especially poor compliance in those 18 to 44.

CO-CHAIR BRISS: Comments or Discussion? Tami?

DR. MARK: If I look at the performance data that they present by state, it looks like the adherence is relatively high, almost close to the target that they're trying to get, 80 percent. So I wonder about the performance gap. I know there are other studies that show performance is low, but when I look at the state-level data, we're talking close to 80 percent.

CO-CHAIR BRISS: Anybody else, comments?

MR. CAMPBELL: I just want to point out the way the measure is reported, that's the portion of beneficiaries with a percentage of beneficiaries with a greater than 0.8 threshold. So it's not the threshold

of adherence itself. It's the percentage of beneficiaries that met that threshold. That make sense?

DR. MARK: So it's about 70 to 80 percent meeting the threshold of 80 percent. Thanks.

MR. CAMPBELL: Yes, from the lowest state, 67.5 to 84.7 percent meeting the threshold at the state level.

DR. MARK: And is there any -- 80 percent is just commonly used as the target, but there's not --

MR. CAMPBELL: Well, 80 percent is -- yes, it's the threshold on the measure, yes.

DR. MARK: I'm just saying the 80-percent target is just sort of what's used in the industry for adherence? There's not any particular scientific basis to say that 80 percent is the gold standard?

MR. CAMPBELL: Actually, there's -- in the form we've cited seven studies

related to outcomes and those outcomes all looked at, you know, and 80-percent threshold with regard to hospitalizations.

DR. MARK: Yes, I know adherence studies everyone uses 80 percent, but there's no study that actually shows that if you don't get to 80 percent -- if you get to 70 percent your outcomes are going to be worse in the population. Or if 90 percent is the key, I mean, 80 percent is what people use?

There's not a dose response study, and also those seven studies -- I don't know if I want to get into this now, but the seven studies for industry-funded, they're very poor design. They're all, you know, just retrospective correlational analyses.

CO-CHAIR BRISS: So, Tami, it seems to me that that's an evidence issue. So why don't we table that issue until we get to evidence? Is that okay? So then I have Harold and Caroline, I think. Yes.

CO-CHAIR PINCUS: So just I agree with Tami that the issue of it's surprising how high the performance is as reported. And in studies that we've done it's using, you know, sort of a 250-day, which is similar. I can't figure out the exact percentage, but it's not too far off from 80 percent. It was more at a 30 to 40-percent level. And in some of the studies that you cite, it's much lower. Do you have any idea in terms of your assessment why it's so much higher? Is it a difference in terms of how they measured, or is there some difference in terms of the populations that were being measured?

MR. CAMPBELL: I think potentially it is both. We do have a Medicare-age population and when we look at our stratification, adherence is clearly higher as you increase in age. We also a methodology in the proportion of days covered whereby if a patient refills early or has overlap, we actually adjust the prescription forward and

give them credit for that. And some algorithms that are published in the literature don't do that for say the same generic name. So if a patient's on olanzapine and they refill that early, then they get credit for the days covered moving forward the way we calculate.

CO-CHAIR PINCUS: And I guess the other issue is that, you know, we're going to be discussing another measure that's very similar. You require two prescriptions already to get into the denominator while some of the others only require one prescription.

MR. CAMPBELL: That's correct. And we did that for two reasons: One to harmonize with the existing adherence measures that are in the NQF portfolio. Under the Medication management Voluntary Consensus Project, NQF asked us as developers to establish a standard methodology for adherence. And so we worked with PQA on that and came up with the PDC methodology that we

use across all our measures.

And then the other point is with the two prescriptions we wanted to ensure that the physician's intent is to continue the medication. So we feel like with the evidence of two prescriptions in the denominator we think that that's sufficient.

CO-CHAIR PINCUS: Well, this actually is going to get into the specifications, but as we think about it, and also the way in which things get harmonized, we should think about which things make sense --

MR. CAMPBELL: Yes.

CO-CHAIR PINCUS: -- in terms of -- because essentially there's choices to be made.

MR. CAMPBELL: Absolutely.

CO-CHAIR BRISS: Yes?

DR. ZUN: I'm wondering if compliance to medication is related to what the states allow for medication, and has that

been looked at? And is that part of the data here? Meaning that is there a greater compliance with one class versus another class, and is that a consideration at all in how we're determining compliance?

MS. HORVITZ-LENNON: So let me just make sure that I understand the question. The question is whether there is evidence in the literature about varying compliance with antipsychotic classes?

DR. ZUN: So the measure talks about two antipsychotics. And the question that I have is what antipsychotics may be allowed could vary state by state, correct? That Medicaid formulary may restrict which meds they get? Is there any allowance made for that issue?

MS. HORVITZ-LENNON: The specification --

DR. ZUN: Or is this Medicaid and Medicare? Yes.

MS. HORVITZ-LENNON: -- calls for

two prescriptions of -- two consecutive prescriptions of antipsychotics regardless of class or type of medication.

DR. ZUN: I'm not sure that I'm being clear. States limit what medications can be prescribed. And so my question is is that consideration in the data or in the measure, meaning that they have to give two different ones? Well, what if they want to give one that's not on the formulary and each state varies? So there's some issue about the ability to comply.

MR. CAMPBELL: I hope I'm understanding your question correctly, but in the measure specifications we are including all antipsychotics. So if there was an antipsychotic that wasn't covered under the formulary, we're calculating adherence across the whole class. And so in that way, you know, we would pick up those claims for the prescription. I'm not sure if I'm answering your question.

CO-CHAIR BRISS: So, Caroline?

DR. CARNEY-DOEBBELING: That would only be possible if there was a claim. And if it's not on the formulary, then the person has to cash pay. So there will be no claim. So I think that's your point?

DR. ZUN: Well, that's part of it. But the answer --

CO-CHAIR PINCUS: But they wouldn't have gotten into the denominator in the first place because they wouldn't --

DR. CARNEY-DOEBBELING: They may have if they had --

CO-CHAIR PINCUS: -- have had a first prescription, unless they had a different prescription that was covered.

DR. CARNEY-DOEBBELING: Right, if they had had a different first prescription, they would be in the formulary. And say the first failed, so then they moved to --

DR. ZUN: A non-formulary.

DR. CARNEY-DOEBBELING: -- a non-

covered agent.

DR. ZUN: Right.

CO-CHAIR BRISS: I mean, it sounds to me like the answer may be fairly simple to your question; which is it sounds like there isn't an allowance for essentially the number of available meds in the formulary, right? So right now we're talking about a performance gap, right, you know? And so I think we've gotten -- or we're supposed to be talking about a performance gap where we may have gotten a little afield from the topic that we're supposed to be on.

So does anybody have anything else specific to the existing performance gap? Caroline? Tami?

DR. MARK: This is a little off, but it might be relevant to the discussion.

CO-CHAIR BRISS: Okay.

DR. MARK: In terms of the specification I thought I read that they had to have psychiatric hospitalization to get in

the denominator. Can you clarify that?

MR. CAMPBELL: No, they don't have to have. Just a diagnosis of schizophrenia in either the inpatient and/or outpatient study.

DR. MARK: Okay. I guess I read the spec wrong. Thank you.

CO-CHAIR BRISS: I think I'd like to suggest that we go ahead and vote on the performance gap issue.

MR. WILLIAMSON: We will now vote on the performance gap. Begin voting now.

For the performance gap we have 6 high, 11 moderate, 1 low and 1 insufficient.

CO-CHAIR BRISS: So moving to evidence.

DR. EINZIG: Okay. So looking at quantity, quality and consistency of the studies, the document cited there were 138 studies cited from the 2009 PORT Psychopharmacological Treatment Recommendations documenting effectiveness of antipsychotic medications. Of those 138

studies, 13 were cited in support of maintenance of antipsychotic medications. Six of those studies were randomized controlled studies. And 7 of those 138 were associated with treatment and outcome showing decreased hospital rates, although none of those were RCTs.

Looking at consistency, compared the 2009 PORTs with the 2003 PORTs showing similar results. Documenting maintenance with medications reduces relapse. Just balance the good with the bad. Balancing the good of medications to decrease schizophrenia, balancing that with the bad of the side effects of the antipsychotics. Keeping that in mind.

CO-CHAIR BRISS: So the floor is open for questions and comments. Jeff?

DR. SUSMAN: Could you talk about the number of individuals excluded, at least in rough terms, when you use the exclusion of the injectable antipsychotics? Was that a

larger number, small number, percentage wise? And maybe talk a little bit to the rationale based on that information.

MR. CAMPBELL: Sure. It's approximately 14.7 percent of our denominator. And the rationale for the exclusion was -- from the literature it indicated that it was difficult to reliably calculate adherence to specifically depot injections because of the variable day supply that are in the data for those medications. And so our technical expert panel, we looked at data with and without the exclusion, evaluated it and came to the conclusion the most conservative approach was to exclude those individuals.

CO-CHAIR BRISS: Tami?

DR. MARK: It's hard to get into the nuances of the evidence, but when I looked closely at the PORT, it looked like there were really three studies that looked at discontinuation and its effect on hospitalization, three randomized trials. And

the results were somewhat conflicting. And the conclusion that the PORT made was that the medication should be used to reduce the risk of symptom relapse during the first and second year following an acute symptom episode.

I read that as a little narrower than, you know, everybody who has a diagnosis of schizophrenia, you know, should be taking their medication for, you know, a year. And the issue of the risks in the application, it's asserted that if you see two prescriptions, it's assumed that the physician thought that the benefits outweighed the risks. But, you know, it may be that they prescribed it twice, saw some weight gain, you know, EPS, decided not a good decision.

MS. HORVITZ-LENNON: So in terms of the first comment, the PORT guidelines over the three versions that they've put out have been pretty consistent about recommending adherence to antipsychotic medication for people who have several episodes of

schizophrenia, have established schizophrenia and have responded to the medicine. There's actually a number of additional studies that have looked at the association between adherence and risk for hospitalization, which is a poor outcome.

In terms of the important concern that you raise about risks and safety concerns, you know, this is something that we recognize as an issue. But the specification actually calls for medication, antipsychotic medication adherence that is not necessarily to the one antipsychotic that they initially prescribed. So it allows for doctors to tailor their treatment to the particulars of the patient they are treating. And there's enough variability in terms of metabolic and other health problems for antipsychotics that the doctors should be able to select/identify an antipsychotic that is right for that patient.

DR. MARK: But how would you allow

for patients who are at the point where they should discontinue?

MS. HORVITZ-LENNON: Say that again?

DR. MARK: I mean, what do you do with patients who are at the point where they just should discontinue? You know, they're not having an acute episode. They're stable. It's time to discontinue the medication. I mean, would those patients be recommended in this to -- I mean, it seems like they would have a diagnosis of schizophrenia, they'd have two medications and they would have poor adherence. And so you would get a negative performance score for that.

MS. HORVITZ-LENNON: So there are a few issues here: One is that schizophrenia is a chronic psychiatric disorder that, you know, most experts would agree requires medications, antipsychotic medication for the duration of the episode, of the illness, sorry, which for most people is lifelong.

If the patient is responding very well and is, you know, all better with the medication, then obviously there might be some room there for the doctor to discontinue the medication. However, that is not necessarily a recommendation given that it is unclear whether that patient will relapse upon discontinuation.

If the patient is not responding, then there are options within the treatment, the antipsychotic armamentarium, which we allow for. So, you know, I think consistent with recommendations we expect that people will be on antipsychotics, but we also call attention to the fact that we're not expecting necessarily 100 percent performance.

CO-CHAIR BRISS: Harold?

CO-CHAIR PINCUS: You know, I think Tami brings up an important point that I think applies to almost all medication treatment for all chronic conditions, where it's essentially a lifelong but not always a

lifelong kind of situation, where there's some people that don't respond or the benefits are not outweighed by the risks or problems encountered. And so you don't expect it to be 100 percent. But it's really no different than thinking about diabetes or hypertension or, you know, other kinds of illnesses, you know, in terms of how one thinks about it.

Now, I think one issue for all of these is, you know, it's not clear what the appropriate threshold is. You know, if everybody achieved 100 percent, I would worry a lot.

CO-CHAIR BRISS: In fact, it would have to be a data mistake, right?

CO-CHAIR PINCUS: Yes. Well, in fact it's a data mistake. But I also wonder about what's the nature of the interaction with the patients.

CO-CHAIR BRISS: So, David?

DR. EINZIG: Just a quick question for clarification. So is the purpose of this

to alert the prescribing physician that the patient is not adhering just as a red flag, or what's the overall goal?

MR. CAMPBELL: Correct, yes. Yes, the purpose at this point would be for quality improvement purposes at the physician group level.

CO-CHAIR BRISS: I'm sorry, I have a jurisprudence question, I think. I think when we approve measures it's for either internal quality improvement or for public -- and for -

MR. CAMPBELL: Just in terms of the public reporting, this measure is specified in the rule for the adult Medicaid core set, so it will be publicly reported at the state level. But at this time we don't have any definitive plans for public reporting at the physician group level.

CO-CHAIR BRISS: So, Nancy? Oh, Vanita first.

DR. PINDOLIA: Well, when you're

reviewing the evidence looking at over the last few years what we've seen in the Medicare population, in our own health plan and in talking to the statewide in Michigan, each year we see more and more people using the $4 and $10 programs and using fewer and fewer of the cards, especially through the doughnut hole. In our plan I've calculated at least 14 percent now, and it might just be the socioeconomic of Detroit itself, unfortunately. Has that been found and taken into account through the evidence that's been gathered? Because it just keeps growing each year.

MR. CAMPBELL: We did some limited sensitivity analysis with internal data related to cash prescriptions. For antipsychotics at the time of our data, 2007 and 2008, there were very few, only the older antipsychotics obviously on the formularies for the cash discount programs. And the impact from our analysis suggested that there

wasn't a large impact due to cash prescriptions, and that's the only evaluation we did.

DR. PINDOLIA: Right, and now, you know, looking at the last year the drugs that have lost their patent protection; we've had three now antipsychotics that are available generically, really the only one left is -- I just drew a blank. It's the one that's the number one right now. But so I think your data might not really reflect what's really going on in today's world because of all the drugs that went generic.

MR. CAMPBELL: So for the ones that went generic though, those aren't on discount formularies where patients would paid cash, right? They would just be generic under the plans tier?

DR. PINDOLIA: For the first six months they aren't, and then afterwards they are. So the one that went in October is about to go into the $4 programs. And the two that

are going through in March, by the end of this year they'll be in the $4 programs.

MR. CAMPBELL: Okay.

CO-CHAIR BRISS: So I think at this point we're still talking the evidence of adherence and outcomes. And so we may be getting a bit afield again. And so I'd like to take Nancy and then maybe let's try to vote this one.

DR. HANRAHAN: This is a process measure and the level of analysis is at both the clinician and the state level. I think what troubles me about this measure is the relationship between adherence and what we're measuring. How is it that measuring an individual's taking a medication, an antipsychotic is going to be improved or changed in some way because we are going to be giving feedback at the state level about the numbers of people with schizophrenia that are adhering to medication?

MS. HORVITZ-LENNON: So if I

understand the question, you're wondering if our measure captures perhaps those patient-driven factors that might be associated with adherence that would not be actionable from a physician or health plan standpoint?

DR. HANRAHAN: The level of analysis you described as both at the group practice and the clinician level and the state level. Those are two different levels. And so the percent adherence of these individuals, how is it that giving that feedback or collecting that kind of data is going to change how somebody is adhering?

And I guess the other concern is that will Medicaid/Medicare; and I'm speaking from the field -- if they have patients that are not adherent is that going to be held hostage to any payment?

CO-CHAIR BRISS: So I'm pretty sure that this one's not an issue of evidence. So this is a usability issue or feasibility or something.

DR. MARK: Nancy, not to put words if your mouth, but I think you're asking what's the evidence that using this kind of performance measure will improve outcomes at the state level?

DR. SUSMAN: But if I understand the process; and perhaps this is more of a Helen question, once the script's approved, it will be used by all sorts of organizations, potentially. And we're just looking at the specifications and performance characteristics. This happens to be a state by state analysis, but one could use this at a group practice level. One could use it within an ACO. One could use it at a state to look at overall performance as a public health issue, if you will.

So, I mean, I understand what you're saying. And if that were the only way this measure were ever used, perhaps I'd feel a little bit concerned. But given the fact that we're actually, you know, looking at

generic measures that will be used in many different ways, I don't have that concern or qualm.

DR. BURSTIN: The one thing I'll put on is the measure as specified is written for a level of analysis clinician group and then up to state, I assume to get at state Medicaid. I think what is fair game; and I think this is what Nancy's asking, is actually in terms of the evidence question. If you're assigning the level of evidence to those different levels, is there any evidence in fact to suggest that there is a relationship between the process and the outcomes of greater adherence through this you measure? And that's, I think, an open question. And that, I think, is fair game. Other than that, I think they've put the measure forward. I believe it's tested. But Jeffrey's actually right. Any measure that's endorsed, whatever level of analysis for which it's been approved can be used in a variety of accountability and

QI uses.

CO-CHAIR PINCUS: But, you know, from my point of view the bottom line is that every single practice guideline in every country has recommended that antipsychotics be maintained consistently for people with schizophrenia.

CO-CHAIR BRISS: I'm sorry, I'm going to take off my chair hat for a second and just comment. As the public health guy around the table, there are all sorts of sort of educational and policy approaches that I could imagine that could be triggered at state or other geographic levels to try to promote adherence, many of which are at least as well documented as individual clinician level approaches. And so, I'm not at all troubled by measurement at a variety of levels. And we won't know. Nobody will know, at the point at which you approve a measure, all of the uses to which it could be put.

So, anybody else want to comment

on this topic before we vote? So, now chair hat back on.

(No audible response.)

CO-CHAIR BRISS: So let's try voting the evidence?

MR. WILLIAMSON: We will now vote on the evidence. This is a one, two, three. Begin voting now.

The measure passes with 11 yes, 3 no, and 5 insufficient evidence.

CO-CHAIR BRISS: So next -- so let's re-vote quickly.

MR. WILLIAMSON: I apologize. I will be better about that.

We will now vote on the evidence. This is a yes, no, insufficient question. One is yes, two is no, and three is insufficient. Begin voting now.

And we are waiting on one more vote. If everybody could please vote one more time. There we go.

And now we now have 14 yes, 0 no,

and 5 insufficient evidence.

CO-CHAIR BRISS: It's still early in the morning. So let's move to reliability and validity, please.

DR. EINZIG: Okay. Reliability. So using the terms comparing signal to the noise trying to filter out the noise. Looking at the variance between the groups and trying to filter out the variance of physicians within the one group. The study looked at reliability on a state level and received very good scores. Greater than 0.9, with good defined as greater than 0.7.

And looking at physician group reliability, became a little bit more interesting there. For groups with greater than 45 patients, they received a higher reliability score compared to those with less than 45.

I'm not sure if other folks have comments about that.

DR. CARNEY-DOEBBELING: I have a

question. For the identification of schizophrenia, are you using a single claim? That does get to the issue of the downstream ultimate reliability of this measure.

MR. CAMPBELL: No, it's at least two claims, outpatient face to face visits and one inpatient clinic.

DR. CARNEY-DOEBBELING: A combination or an and/or?

MR. CAMPBELL: Or.

DR. CARNEY-DOEBBELING: Okay.

DR. ZUN: I have a question about 2-A-1.1. I'm a little confused. During the intake period, I thought we're talking about a 10-month period or something. I mean, I'm confused. The intake period?

MR. CAMPBELL: We didn't specify an intake period in our measure. I think that is the NCQA measure.

DR. ZUN: Am I in the wrong place? Okay. Maybe.

CO-CHAIR BRISS: So maybe while

we're -- Leslie, maybe while you're looking --

DR. PATING: I just have a clarification question about the construction of the measure. So is this only for Medicare populations, or is it everyone?

And secondly, if it's claims data, which I believe it is, is it the doctor's office visit claim or is it the pharmacy claim? Because when you get these carve-outs of pharmacy benefits -- I was just trying to figure out how all the data gets collected and follows the patient around. So just in terms of the construction of the measure.

MR. CAMPBELL: Sure. So the measure's based on integrated claims data. So these are fee-for-service beneficiaries, and we use Part A, B and D claims to construct the measure; A being the inpatient, B being outpatient, and Part D being the prescription drug benefit from Medicare. So the measure is tested in the Medicare fee-for-service Part D eligible population.

DR. PATING: And then I guess part of that individual quality improvement cycle -- so then if it's at the sort of meta-level that you gather it and you're a fee-for-service doctor, how do you get that information back to your system? Because you may not have the A and B portions of that data in your office.

MR. CAMPBELL: Right. So for the physician group in the CMS reporting programs that provide physician group data back, they do that. They provide it back to the individual provider groups.

DR. MARK: To follow up on that point, so you're excluding the Medicaid dual-eligibles. So what do you do with the fact that you don't get the Medicaid outpatient claims?

MR. CAMPBELL: No, we do have Medical dual-eligibles in our population.

DR. MARK: But you only get their drug data, right? You don't get their

Medicaid claim data, which would have all the outpatient and inpatient services.

MR. CAMPBELL: No, we get their A and B data as well when they're dual-eligible.

DR. MARK: Yes, but Medicaid dual-eligibles would be covered under Medicaid. And if you're getting Medicare claims, you're not getting their outpatient claims data. You're only getting their drug data, because --

DR. CARNEY-DOEBBELING: Well, the duals, are -- the inpatient and outpatient are first primarily covered by Medicare. And if Medicare doesn't cover the claim, it then passes to the state Medicaid agency. So if it's a covered benefit under Medicare, like an inpatient stay would be, Medicare will pick it up and pay for it.

DR. MARK: But if it's a psychiatric rehab benefit, it's --

DR. CARNEY-DOEBBELING: Only things like Medicaid rehab option would --

DR. MARK: Right, which is where most of the mental health outpatient services are covered, is under the Medicare rehab option.

DR. CARNEY-DOEBBELING: Not medication management.

DR. MARK: Right, but the idea is you have to get a schizophrenia diagnosis in an outpatient or inpatient setting. And if you're getting treatment in an outpatient setting covered under the Medicaid rehab option, that's not going to be picked up in this data. So you're going to be missing a lot of the picture of the services received by the Medicaid dual-eligibles.

DR. CARNEY-DOEBBELING: I actually don't think so, having studied this quite intensively, because Medicaid rehab option A isn't even used in all 50 states very extensively. Every state has a variable number of MRO-type services that they cover. And for anyone who is on medication for

ongoing medication treatment, there will have to be a med management outpatient visit.

CO-CHAIR BRISS: So could --

DR. MARK: But I'm just -- let me just --

DR. CARNEY-DOEBBELING: For a traditional psychotherapy --

DR. MARK: Let me just say, where I'm speaking from is we have a large Medicaid claims database with 10 million covered lives. And we have data from dual-eligibles, and I can say that there's a lot of mental health outpatient and inpatient claims in the Medicaid dual-eligible claims database. So, you know, based on my experience you'd be missing a lot of Medicaid services if you're not getting the Medicaid claims. You'd be missing a lot of mental health services if you're not getting the Medicaid claims.

CO-CHAIR BRISS: So the concern as I understand it is that you might be leaving a lot of people out of the denominator who

would otherwise be appropriate to be there. Is that right?

So, and we ought to let the developer comment on and to answer the question. So, do you have comments on that?

MR. CAMPBELL: Right. So, but if the services weren't identified, then the patients would be in the denominator. So that is true. The only way we capture the patients is if they show up with a Medicare A or B claim.

DR. MARK: Yes, and you're using this for performance measured for a state, and so you're missing a large part of the their Medicaid population, potentially.

MR. CAMPBELL: Yes, I can't comment on the actual implementation for how the measure will be calculated related to the rule, just that this is how we tested the measure with the integrated claims data.

DR. SHEA: I, too, have just a question in terms of how you account for

patients that are in hospital. They will have an interruption in pharmacotherapy prescriptions being filled, but yet they're getting treatment and how that gets accounted for.

MR. CAMPBELL: Right, so that's an issue that our technical expert panel was concerned about as well. And so one of the things we did in our early alpha formative testing was to evaluate the exclusion of hospitalizations. One of the things we did was look at what would happen if we excluded those that had hospitalizations and we ended up losing a lot of the people in our denominator, which was not a good thing that we wanted to happen.

And then the other option we looked at was potentially crediting hospital stays as actually days covered. And when we did that, at least in our limited sample; because we only did this with two states, we did not evaluate at the physician group level,

but the state level. It was a very small difference. I want to say about a 1-percent difference between what we saw before and after the exclusion. So the expert panel felt that, you know, based on that it wasn't appropriate to do the exclusion for hospitalizations, or the credit.

CO-CHAIR BRISS: David? David?

DR. EINZIG: So sorry. Sorry. I don't want to talk.

CO-CHAIR BRISS: That's all right. Vanita?

DR. PINDOLIA: This is for the developer. If you can help me understand the 2-A-2.3 testing results. So, and this kind of goes back to my question of trying to take into account the $4 drug and $10 drug programs, or group homes where they give a lot of samples, where a lot of the young adults do end up going with this because the parents can't have them in their home anymore.

I understand the reliability score

is very high each state, but there's such a variance from state to state of 67 percent in Arizona being the lowest all the way up to mid-80s. And maybe I'm misinterpreting, and that's why I'm asking. If that interpretation is correct, that there's this variance from state to state, was there any look into why it was so low in certain states? And was there more use of group homes, was there more use of these free drug programs, or something like that?

MR. CAMPBELL: Yes, the answer to that question is, no, we did not evaluate further the variance that we saw within the individual states. We have not done that to date.

CO-CHAIR BRISS: So I see no further cards up. Are we ready to vote reliability?

MR. WILLIAMSON: We'll now vote on the reliability. This is a high, moderate, low, insufficient rating. If you would begin

voting now.

All right. We have high 2, 14 moderate, 1 low, and 2 insufficient.

CO-CHAIR BRISS: So validity?

DR. EINZIG: Okay. Validity. They looked at face validity. They had 12 individuals on the expert panel and they were asked the statement does the measure appear to -- does -- the measure appears to measure what is intended. All the folks, all 12, 12 out of 12, either strongly agreed or agreed.

Threats to validity included cash for prescriptions and missing data. They felt that this was low numbers.

Are there comments on that?

CO-CHAIR BRISS: Floor is open.

(No audible response.)

CO-CHAIR BRISS: Hearing none, are we ready to vote?

MR. WILLIAMSON: We will now vote on validity. Again this is a high, moderate, low, insufficient rating. Begin voting now.

Okay. We have 2 high, 14 moderate, and 3 insufficient evidence.

CO-CHAIR BRISS: So moving to usability.

DR. EINZIG: Okay. In terms of usability, one basic premise is this should be useful because an adherence measure will help providers recognize patients that are not adherent. I think there might have been a question that was alluding to: does adherence equal compliance? So that might be a question for discussion. But for those with low adherence it could be useful to help develop interventions for the groups and the patients.

Now when the technical expert panel were asked, 12 out of 12 all agreed or strongly agreed on usability.

CO-CHAIR BRISS: So the floor is open. We may have already some of this discussion.

CO-CHAIR PINCUS: Question: You said in terms of reporting, sort of a limit on

the least number of patients that a group or individual physician provided valid data.

MR. CAMPBELL: Yes, that's correct. Based on our case volume analysis, we set that at 45 patients in a physician group practice.

CO-CHAIR BRISS: Questions? Comments? Ye?

DR. PATING: Just the level of data. So it can state and I guess systems of care, clinical levels, this is also available at county level data, do you know?

MR. CAMPBELL: Yes, we don't have it specified for county level. We just have it specified for state population and physician group at this point.

CO-CHAIR BRISS: Any other questions or comments?

(No audible response.)

CO-CHAIR BRISS: Let's vote.

MR. WILLIAMSON: We will now vote on the usability. Again, as a high, moderate,

low, insufficient rating. Begin voting now.

Okay. We have 7 high, 9 moderate, 2 low, and 1 insufficient.

CO-CHAIR BRISS: And feasibility.

DR. EINZIG: In terms of feasibility, much of the data is already out there. It's coded by somebody else. There's use of electronic claims. Susceptibility to inaccuracies were not identified. And data required is readily available.

CO-CHAIR BRISS: Questions? Comments? Concerns?

(No audible response.)

CO-CHAIR BRISS: None? Let's try voting.

DR. PINDOLIA: I know I've said it before, but -- just that is going to be the main issue that we're going to have in accountability, even though it says here susceptibility and accuracy errors are unintended consequences, but we know we have a large percentage that we will not be able to

account for because of the free drug programs.

CO-CHAIR BRISS: Yes?

DR. SUSMAN: I mean, just to that issue, have you who are CMSes thought about doing some additional work to try to quantify how big of a issue this really is?

MS. HORVITZ-LENNON: Yes, we have. We've talked with the team about this. Based on our initial limited analysis we thought that there wouldn't be a major issue, but I appreciate the comments of members that times are changing. And I think to support that we should take a closer look at this. Appreciate the input.

CO-CHAIR BRISS: So let's try voting feasibility, please.

MR. WILLIAMSON: We will now vote on the feasibility. This is high, moderate, low and insufficient. Begin voting now.

We have 2 high, 13 moderate, 3 low, and 1 insufficient.

CO-CHAIR BRISS: Yes, we were so

hard on the measures yesterday, I'd forgotten what happens when you get to the end and you want to approve one. So it's time to vote overall approval. One is yes and two is --

DR. BURSTIN: Just one comment on that. This is a measure that is directly competing to at least one other that we're going to talk about today. So just evaluate this as is, suitability for endorsement. It would not move forward until we've run through the issues of combining the other measures, etcetera. So we need to say suitability for endorsement before we can get into the harmonization competing discussion.

MR. WILLIAMSON: We will now vote on the overall suitability for endorsement. This is a yes/no question. Begin voting now.

We are still waiting on one response. If everybody could please -- there we go.

All right. The measure is 16 yes, 3 no.

CO-CHAIR BRISS: So are we ready for 1935? So would NCQA like to tee up measure 1935, please?

DR. SCHOLLE: Okay. Great. Good morning, everyone. I'm Sarah Hudson Scholle. I'm Vice-President for Research Analysis at NCQA, and I wanted to tee up this whole suite of measures about care for schizophrenia that you'll be looking at today.

NCQA worked with Mathematica Policy Research under a contract from ASPE to develop this suite of measures. These measures are intended for use at the state Medicaid program level, and they were tested using fee-for-service Medicaid claims data and as well as other kinds of testing.

Our goal in developing this suite of measures was to look at the physical health needs, the pharmacological health needs and the psychosocial needs of people with schizophrenia. And so, we started off with a number of measure concepts. As you'll notice,

there are no measures that relate to psychosocial treatment because, despite the evidence base for those measures, we could not find reliable ways to measure those constructs in the Medicaid claims data.

We reviewed all the measure concepts with a multi-stakeholder panel that involved consumers, researchers and experts in schizophrenia treatment, as well as representatives from state Medicaid and mental health programs. We conducted an evidence review for each measure. We presented that to our advisory panel. We prepared the specifications. Our testing, as I mentioned, occurred in the claims data, and because it's fee-for-service claims and we used the Medicaid extract file, there were some challenges in that testing. And so that's why some of the measures didn't make it past our specification and testing phase.

We included in our testing feasibility testing with state Medicaid

medical directors, state mental health directors, and representatives from management behavioral health care organizations. So you'll see the results of that related to the feasibility and usability of the specifications as well.

This work began before the development of the Medicaid core set, and it was our hope that the measures would be something that could be suitable for consideration for the Medicaid core set, although these measures were not ready in time to be presented for the initial round of evaluation of potential measures for that set.

NCQA is the owner of the measures, and we recommended that the measures be included in the HEDIS data set for Medicaid health plans. They've been out for public comment for that use. They were presented for state Medicaid programs in the specifications that you see because the measures have not been approved for HEDIS by NCQA's Committee on

Performance Measurement, but we do intend them to be useful for health plans as well.

So that's just a background to this process, so that you understand sort of how we got to the set of measures that you're seeing today.

The first two measures we really view as paired measures. One looks at the use of antipsychotic medications, and the second one looks at continuity of antipsychotic medications. And it's that continuity measure that is very similar to the one that you discussed, and I have to thank our colleagues on the measure developer side for answering many of the questions that I think you'd probably raise about our measure as well.

I would point out that one of the things that I'm hearing in this discussion has to do with the definition of this population who should be on antipsychotic medications. And one issue that came up in our discussions frequently was whether there is a way from the

claims data to identify people with schizophrenia in a reliable way. That's why you'll see that in our denominator definitions for these measures we started at age 25.

Our multi-stakeholder expert panel felt that, starting at age 25, we'd have more confidence in the diagnosis of schizophrenia than if we looked in younger age groups and -- so that's why we have that age difference. We've heard that this is challenging, because most other measures for adults start at age 18. And certainly that's one of the differences between our measure and the CMS measure.

The other issue that came up was trying to understand when do people voluntarily take themselves off of an antipsychotic. And while we considered that to be -- it's something that came up in our expert group and other places, a claims-based measure is not a place where you can find that information about that. And I would just

encourage you to think about these measures as being measures that would allow us to evaluate and compare state by state, rather than saying that 100 percent is always the right number that you're aiming towards.

We want it to be high, and our expert group felt that the evidence supported this measure for people who have a diagnosis and who have been placed on a medication -- that that's a sense of: this is the treatment. If you're on it, the benefit will come from staying on it. However, we realize that there are some cases, but that's the kind of measurement issue that with a claims-based measure you can't get into the: where did they voluntarily come off or not. We likewise saw a lot of variation across states in the performance rates, and the Medicaid data was much lower than it was in the Medicare claims data that our colleagues presented.

CO-CHAIR BRISS: Thank you. And, Dr. Pincus?

CO-CHAIR PINCUS: So what we're going to be discussing are two measures initially that are paired measures, as Sarah said, that together really represent a lot of what the information we were just talking about. So I guess I'll present for the first measure first, but in some ways a lot of it we've already discussed. And once we go through the first measure, we may not need to discuss very much about the second measure. So --

And the reason they're paired as I understand it -- and, Sarah, correct me -- is that basically the first measure, 1935, that we're going to be discussing, is whether this denominator population of people with schizophrenia received any antipsychotic medication, at least one prescription a year. So in principle it gives you a sense of what percentage of the broader population the second measure, which is the maintenance measure, represents. So you get a picture of

context and see how many people are actually under care with some intent to prescribe.

So in terms of impact, basically it's the same information we just went through in terms of the extent of the population with schizophrenia, their clinical needs and the extent to which they're costly to the population.

I don't know if there's more that we need to discuss about that.

CO-CHAIR BRISS: Yes, so anybody have additional comments that we haven't already said about importance to measure? Yes?

DR. MARK: So the consensus is there is a public health problem with adherence to antipsychotics?

CO-CHAIR PINCUS: Yes, there's a problem in that, you know, people with schizophrenia -- you know, the care could be improved. Well, that adherence could be improved, and this meets the priorities in

terms of chronic illness care, in terms of impact at a population level.

CO-CHAIR BRISS: All right. So let's vote. This is a high, moderate, low, insufficient. Oh, sorry.

DR. HANRAHAN: One observation that I have, too, is that this measure was vetted among consumers, and the first measure was not. And also, that this measure seems to be more looking at the prescribing of antipsychotic versus the adherence to antipsychotics. Yes, they're related but, you know, it's more direct and factual to measure the prescribing, I think, of antipsychotics, given all the problems that we've already mentioned.

CO-CHAIR PINCUS: I mean, your point is well taken, but it's got its pros and cons to it, because the prescription of a single antipsychotic prescription is probably not going to be that significant, but it does provide that picture that is not the picture

when you look at just the maintenance one.

DR. CARNEY-DOEBBELING: I'm not sure if this is the right part to ask, but I have heard a lot of information or opinion from colleagues across the country about the 25-year-old cutoff, and I was curious if our Committee assessed that. The evidence for schizophrenia is that treatment -- a first break early in the course of that disease -- the earlier in the course of the disease, the better. So I've been confounded about the 25-year-old cutoff. That would imply at that point more of a chronic persistent schizophrenia for most folks.

CO-CHAIR BRISS: So this is probably not the right place, but this issue's going to come up. So do you guys want to comment for --

DR. SCHOLLE: So actually those are the things that our group was balancing. And our advisory group felt that trying to make sure that if we're going to have measures

that are saying: are antipsychotics used and are they used continuously -- because that's the goal of these measures -- they wanted to be really comfortable that in the denominator we had people who really had schizophrenia and not people who got into the denominator because they had bipolar, and then they had schizophrenia, then they had bipolar and schizophrenia. Our denominator definition of the diagnosis is similar to the CMS measure, requiring either an inpatient or two outpatient diagnoses.

So, from the claims data we didn't have a way to go back and say, well, really is this really a schizophrenia, or it could be some other condition? But and that's where our panel came down in trying to say, okay, we want to make sure that these are folks who have schizophrenia. But we've also heard a lot of concerns about that age limit and request to drop it --

DR. CARNEY-DOEBBELING: So how was

the 25-year cutoff chosen?

DR. SCHOLLE: By the consensus from our advisory groups saying, you know, that would be the best age --

DR. CARNEY-DOEBBELING: I guess based on what evidence was 25 the age?

DR. SCHOLLE: I believe that their recommendation was based on their sense that the epidemiology suggested that by the time you're 25 that the changes in diagnoses would have lessened the kind of jumping back across diagnoses.

DR. MARK: And you include schizo-affective disorder within schizophrenia, because it's within 295?

DR. SCHOLLE: Yes.

CO-CHAIR BRISS: And, Bernadette?

DR. MELYNK: I really agree with the earlier comment. I think the evidence out there would support that the earlier again these folks get into medication adherence, the better prognosis. So I would really encourage

to look at an earlier age for this particular measure.

CO-CHAIR BRISS: Okay. So let's try to vote evidence of impact, and we'll circle back to 18 or 25 or something else.

MR. WILLIAMSON: We will now vote on impact. This is a high, moderate, low, insufficient vote. Begin voting now.

All right. We're waiting on one more response.

Okay. For impact we have 12 high, 6 moderate, 1 low, and 1 insufficient.

CO-CHAIR BRISS: Sorry, we're trying to shoehorn age into the rigid NQF process up here, and it's hard actually.

CO-CHAIR PINCUS: Yes, we're trying to figure out under which category that we would discuss the 25 or 18.

CO-CHAIR BRISS: So let's --

CO-CHAIR PINCUS: Maybe when we get into reliability, we would discuss the actual measure specifications.

CO-CHAIR BRISS: Or maybe -- let me try putting it into evidence, because I mean, in some sense it might be an evidentiary question. So let's try to quickly move through opportunity, and then we'll talk about the age stuff under evidence, or at least we'll start the discussion under evidence.

CO-CHAIR PINCUS: So in terms of the performance gap, it's basically the same type of information that was conveyed as under the previous discussion. I think as Sarah mentioned, in looking at the performance gap under the Medicaid population, it's larger than it is under the Medicare population.

CO-CHAIR BRISS: So anybody want to make additional comments? Vanita?

DR. PINDOLIA: But looking at the performance gap, it's very small. It's 89 percent to a mean of 93 percent. And I understand that this is basically developing your denominator for the next measure. So is that why -- I'm troubled at this one being

needed as a separate entity, or it should be just combined because if the performance gap is so small --

CO-CHAIR PINCUS: Well, I think that's one of the issues that we're kind of dealing with, because these are paired measures. So, you know, one could argue that this measure by itself wouldn't stand. But on the other hand, having this measure as a kind of benchmark against which you can look at the -- you know, contextualize the second measure -- improves the second measure. So I don't know how NCQA sort of deals with that issue. NCQA, how you thought about in terms of pairing the two. And, I mean, is there an option to have it so that it's considered together if you --

DR. BURSTIN: I mean, essentially -- and I think this is as it's proposed, this measure would not be a stand-alone. It would only be as paired with the second measure. So we're fine with that. Otherwise, the

denominator specification would be built into the measure. But if there's a usefulness to having them both, as long as they're paired, that's okay.

CO-CHAIR BRISS: So performance gap.

MR. WILLIAMSON: We will vote on the performance gap. This is a high --

DR. ZUN: Now that I have the right measure, the performance issue -- I have a question about, because even though the performance measure gap is small, the question is: is there evidence to explain why? And the reason I ask is: is this a -- can we do something about the performance gap, or is it set that there's always going to be X number of people who will not take their medicine, or they won't take their medicine because of some other reason.

And so, you know, it sounds like there's an important measure, but the performance gap, I'm not sure I saw any

information to explain why it's not 100 percent. Where the gap is, that 5 percent -- is it because they are intolerant to those meds, they can't get their meds, they are so -- I don't know. If we're putting that measure out there, what's that group? What's that population?

CO-CHAIR BRISS: So does the developer want to comment on that?

DR. SCHOLLE: So it could very likely be that that was a patient choice, and a desire not to be on an antipsychotic. So it's hard to interpret that lack of an antipsychotic medication.

And like Harold said, I mean, our intention with this measure was really to pair it with the -- to look at access to this medication, and then to use that information to be able to understand the continuity information, to see whether that would help us. And do you see better continuity in states where you see higher access rates, or

does it work opposite of that? Just looking at the results seems to suggest, you know, the states that have lower use performance rates also have lower continuity rates, which suggests that, you know, there's a similar kind of activity going on.

CO-CHAIR BRISS: So I think I'd like to try to vote this. So evidence for performance gap.

MR. WILLIAMSON: We'll now vote on the performance gap. This is a high, moderate, low and insufficient vote. Begin voting now.

We have 1 high, 11 moderate, 7 low, and 1 insufficient.

DR. SUSMAN: Can I just ask, Helen, when we have these paired measures, the issue oftentimes might not be a performance gap in the first measure that you're sort of setting everything up with. So in considering the methodology, I wonder if we should really be looking at this for the future. I think

here it's fine. It passed. But I would feel bad about something not passing because the baseline sort of that sets up then the performance gap is shot out of the water.

DR. BURSTIN: It's an excellent point. I think the issue though is there's lots of different kinds of paired measures, and not all of them in fact set up the second measure, but in fact offer two rates on a similar thing that, you know, that you'd want to see them together.

For example, volume and mortality from cardiac surgery procedures, something like that, you'd want to see those together, but it's not as if you wouldn't want to them have volume. It gets a little complex, but I think the point is well taken.

DR. SUSMAN: Yes. Yes, I mean, just thinking about maybe an A or B to sort of try to allocate them to one or two buckets.

CO-CHAIR PINCUS: I would agree with that, to actually sort of think about the

two different categories of paired measures.

DR. BURSTIN: Yes.

CO-CHAIR PINCUS: Those that can be independently looked at and those that can't.

CO-CHAIR BRISS: Okay. So moving to evidence. And, yes, there may be some discussion on this point.

CO-CHAIR PINCUS: So again, we come to the issue that Jeff just raised. For this measure specifically, the sort of attribution of evidence that a single antipsychotic prescription is going to have an impact is probably small. On the other hand, in thinking of it as being paired with the second measure -- which is looking at sort of consistency of antipsychotic prescription over time for people with chronic schizophrenia -- then it's a totally different ball of wax. So I guess in some ways, Helen, it comes back to you in terms of how we should rate that.

DR. BURSTIN: It is complex.

CO-CHAIR PINCUS: So I mean the way I've been thinking about it is that these are paired measures and, you know, they go together. And it's very useful to have this measure to just look at, you know, what's the overall denominator that the second measure is looking at? Because if some states had very low single prescription rates, that's a problem in and of itself, and you want to sort of adjust for that.

CO-CHAIR BRISS: And in terms of the evidentiary question about whether treating people and consistently treating people with antipsychotic meds, appropriate people with antipsychotic meds, I think we may have dealt with in the last review. And I hope we don't have to re-litigate that in every measure today.

And at least for me -- I guess I'm taking off my chair hat for a second. At least for me, I would be okay in this paired measure that -- if you think treatment is a

good thing, getting started on treatment is a good thing --and then continuing on treatment is a good thing. And these measures taken together kind of answer that.

And so I wonder if we could kind of move on from that. I think that there are additional evidentiary questions with this measure, like the age cutoff, that seemed to me to be harder. So, Caroline?

DR. CARNEY-DOEBBELING: I've already voiced the age cutoff, but I was curious from the developers: Many of your other behavioral health-related measures go back to an indexed event, a new event for a diagnosis. And there's an acute phase treatment and a continuous phase treatment.

This measure was set up a lot differently, probably because of the issue of the 25-year-old, and you wouldn't have necessarily a first episode or an indexed hospitalization in that period. But I'm curious why they were constructed differently

than the ADHD measure and the depression treatment measure.

DR. SCHOLLE: So this measure looks a lot more like our measures for people with diabetes and heart disease, where it's a lifelong chronic condition approach rather than depression or ADHD that might have an episodic treatment approach. I don't think the evidence for continuing to treat depression with antidepressant medications after the symptoms from an episode have resolved -- I don't think that that evidence is strong to continue, or it may be disputed. So certainly that's the part of what's going on.

So in this case we were trying to align with other measures that look at medication possession ratio, as in our asthma measure that's looking at poor people with asthma of a certain level of severity. Then they should stay on a controller medication over time, and you're looking at the number of

days covered. So it's really treating this as a chronic lifelong condition rather than an episodic condition.

CO-CHAIR BRISS: So your card's been going up and down. Are you satisfied?

Tami?

DR. MARK: Sorry for asking so many questions. I've spent years and years looking at claims data and mental health diagnoses, so I get into the nuances of all these things.

So I guess, you know, I think there's one issue, which is the evidence that people with a clear diagnosis of schizophrenia benefit from long-term use of antipsychotics, and then there's this other issue of whether someone with one diagnosis in schizophrenia in a claims database should be getting an antipsychotic.

And given that, I think some of the issue around the 25 and the weighing, it's how you're weighing this diagnoses knowing

that -- I'm not a clinician, but knowing -- you know, seeing a lot of single diagnoses of schizophrenia in the claims data and thinking about it clinically, that someone might show up at a hospital with a drug psychosis or, you know, show up with dementia and get a schizophrenia diagnosis, you know, how do we count those kind of inappropriate or misdiagnoses in these measures?

CO-CHAIR BRISS: Nancy?

DR. HANRAHAN: I think that's certainly a question I had in mind, because I've used these data, too, to answer research questions. But what really clarifies -- and I want to just check this out to be sure I'm thinking straight about this -- is that what's so appealing about this particular measure is because it's a state level measure. So in other words, at the state level we're getting a sense of how well these individuals that have this diagnosis -- given the fact that there is a margin of error in that that's

pretty strong -- are getting the standard-of-practice medication.

So instead of looking it from the individual level and the pros and cons of whether I should take the medicine or shouldn't take the medicine, whether I have had one diagnosis or another, this measure is really moving it up to a population level. And then there can be decisions made, as they said in here, about how better to set up systems so that these individuals get the support they need, versus how can we hold the clinician accountable for whether or not that person takes their medicine or not. Does that make sense? Yes.

CO-CHAIR BRISS: Back to Caroline.

DR. CARNEY-DOEBBELING: But if these get entered into the HEDIS data set, they'll be used to judge health plans. So they'll drill down definitely beyond the state level, and plans may then use them to drill down even further.

DR. HANRAHAN: But if it's determined -- and this is a question, that's really --

DR. CARNEY-DOEBBELING: But if this measure's intended for the HEDIS data set, then it definitely will be used at the health plan level.

DR. HANRAHAN: But if it says that it's going to be an analysis at the state level and it's defined as a denominator --

CO-CHAIR BRISS: So we need to ask NCQA to answer that question.

DR. SCHOLLE: So the measure as presented to you is defined for the state level. As I mentioned in the introduction, these measures have been proposed for inclusion in HEDIS, but that's not presented in what you have, because they haven't been approved for use in the health plan level. And there are ways that it can be used. The likelihood of this measure being useful at an individual provider or program level is

challenging because of the denominator size, I think. And I think it might be used more in a quality improvement context at that level than it is in a public reporting context at a population level, like a health plan or a state level.

I did want to just clarify that the way the denominator specifications read, it's one inpatient or two outpatient diagnoses. And again, you know, as we were looking at it, we were thinking age 25. You know, we're trying to minimize those errors of putting people in the denominator. That's why age 25: inpatient. Then we're pretty confident that somebody's doing a good diagnosis hopefully.

We did do some sensitivity analyses, and really changing the way that we defined the denominator didn't change the number of people who got into the denominator very much. So when we looked at two outpatient, whether we included primary or

secondary diagnoses. So we felt that it didn't change it much, to use that one inpatient. Actually, the bulk of people who came in to the denominator, come in through the two outpatient diagnoses.

CO-CHAIR BRISS: So, Jeff?

DR. SUSMAN: So just to confirm with NCQA, I mean, the real action here is in the pairing of the measure and citing up really the second measure. I mean, it's not to look at this data per se. It's to set you up to be able to look at the second persistence.

CO-CHAIR BRISS: Nancy, are you still trying to speak again, or is your card just up? It's okay.

All right. So on the evidence I think we've already decided that medications for people who need medications are a good thing, right?

We've had the discussion about the age issue. And so as I understand it,

essentially the argument that's being made by the measure developer is that they've tried to balance essentially sensitivity and specificity sorts of issues, and they've picked 25 for this measure set to be a little more specific, perhaps at a cost of some sensitivity. Right?

And so, are there any other evidentiary issues that ought to be put on the table before we vote evidence?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's try a vote.

MR. WILLIAMSON: We will now vote on the evidence. This is a yes, no, insufficient question. Begin voting now.

The measure passes: evidence 19 yes, 1 no.

CO-CHAIR BRISS: So reliability and validity of the measure.

CO-CHAIR PINCUS: Let's go to reliability. And the way which this was

tested was basically to look at the test/retest reliability across states, and there was some variation that was found.

And maybe, Sarah, could you sort of discuss the reliability findings on this one, and maybe also in the context of it the second measure as well, to just get -- you know, to distinguish the reliability, you know, between the two measures?

DR. SCHOLLE: Sure. And I'm sorry, the form doesn't allow us to put in our pretty pictures. We're working on that.

So basically, you know, there's a challenge when you try to examine reliability. We had a limited number of states, so the signal-to-noise approach to testing, we did not apply the one that CMS used. And instead we looked at: over time did we feel like this was consistent. And in this measure we had, I think, 15 states, 15 or 16 states that could report it. It's probably on that chart.

But in most of the states, all but

one state, it was either no quartile change, you know, in the ranking of the states or one quartile change. And there was one state that changed dramatically and had a three quartile change, but that state actually had very small numbers. And so we weren't able to look at reliability in that state, you know, from year to year.

So we think that that may be a data problem. And similarly on the medication, the continuity measure, again we see more states. The states that were able to calculate, that were able to be included in this quartile analysis to see if they stayed in the same quartile -- compared to other states over time -- all the states but one did. But several states dropped out because their denominators were small. And I think the states that tended to have more of a shift tended to have smaller numbers. And so, we thought that that probably contributed to it. We saw small numbers in some states. We saw

variations in the proportion of people who were eligible for this measure across states, and some states had very small numbers.

Remember, we were looking at Medicaid fee-for-service data only and through the MAX files. So actually, we didn't have the Medicare data, so we weren't able to look at dual-eligibles, and we weren't able to look at some states' specific kinds of codes for behavioral health care, because those activities are not in the MAX files. So they might contribute to this.

I think our sense is when we reviewed this with our panel, they felt like -- this was good evidence of reliability as far as we could tell from the data source, but we have to realize this is for claims to -- and this is not intended to say this right for one person or another. It's really at a population level. Is it fairly consistent over time? And our group felt like it was.

CO-CHAIR BRISS: So when you talk

about small numbers by state, what kind of numbers are we talking about?

DR. SCHOLLE: You know, less than 50.

CO-CHAIR PINCUS: So, Sarah, there's one question you can clarify. Is this measure intended for Medicaid fee-for-service only, or Medicaid for the entire state Medicaid performance?

DR. SCHOLLE: Right. Okay. So the issue is: does testing in the fee-for-service claims data give us enough confidence for applying this measure and Medicaid programs generally? And I think the answer is: yes, you know, claims data from a state program -- the claims data that you have is a fee-for-service -- you know, if you're a state that's fee-for-service only, those claims data you can calculate yourself.

If you're in a state that uses managed behavioral health or managed care, then states can either ask their health plans

to calculate the information, or they can ask for encounter data to do it. We just did not have those data to be able to apply the measure. But our experience from HEDIS is that, you know, states use the HEDIS health plan specifications in their fee-for-service claims data. And really what they're changing is the definition of continuous enrollment in the state. So we have pretty good confidence that those claims-based specifications work pretty well for states.

What our testing in the fee-for-service -- the Medicaid Extract file -- does, is allows us to compare and see what's happening across states. And so we had 15 to 20 states that were incorporated. You know, trying to get managed care data from all those different states would be quite a challenge. But we're able to see that there is variation, that there are some states that have smaller denominators that may have to do with their eligibility requirements for Medicaid. But

our group felt confident that this testing in the MAX files would help us.

We also talked with state Medicaid medical directors about the implication of these measures, and they felt that the specifications were things that they could do. Where they felt like they couldn't apply our specifications, or that they would get different information, they told us that. And that's why we don't have psychosocial measures for psychosocial treatment, because the Medicaid medical directors were very clear, we're not going to be able to do that, and it wouldn't be fair if you compared one state to another on access to a sort of community treatment or something.

So we had the fee-for-service data to get quantitative results. We used the focus groups with our Medicaid medical directors and mental health directors and managed mental health care officials to help us understand: would these specs work in

other settings.

CO-CHAIR PINCUS: So to clarify with regard to the reliability issue specifically which we're talking about, that the intent overall will be to apply this across the full state Medicaid program, managed care, fee-for-service?

DR. SCHOLLE: Yes.

CO-CHAIR PINCUS: But you tested it in the fee-for-service data that were available. So one can imagine that some of the issues around the reliability from year to year could affect changes in the overall proportion and nature of the managed care programs in relationship to the fee-for-service programs --

DR. SCHOLLE: Right.

CO-CHAIR PINCUS: -- in the state, which would affect the numbers in the denominator.

DR. SCHOLLE: Right.

CO-CHAIR PINCUS: So I'm just

trying to understand where the instability is.

DR. SCHOLLE: Right. Right, we saw really big differences across the states and the proportion of people who had dual-eligibility and therefore couldn't be in our denominator either, because we didn't have Medicare data. So there's some messiness there.

CO-CHAIR BRISS: So did you try retesting your reliability stuff in places that had sufficient numbers? You know, the less than 50 testing makes me a little queasy. And so, I --

DR. SCHOLLE: No, so the data that we presented -- I mean, when we had small numbers, we excluded those states. And so you get the smaller number of states that are included in that test/retest reliability, where there were few. But even when they just cross the line of our threshold of reporting them, we still expect to see a little bit of messiness. New Hampshire, you know, was

messy, right, because they had small numbers across. Wyoming was only present in this measure, so you get a sense of the -- not a lot of schizophrenics diagnosed in the Medicaid data.

CO-CHAIR BRISS: Anybody else have questions or comments about this?

(No audible response.)

CO-CHAIR BRISS: Let's try voting reliability, please.

MR. WILLIAMSON: We will now vote on reliability. This is a high, moderate, low, insufficient rating. Begin voting now.

We're still waiting on one response, so if we could please --

We have 1 high, 15 moderate, 1 low, and 2 insufficient.

CO-CHAIR BRISS: Validity, please?

CO-CHAIR PINCUS: In addition to some of the similar data that was presented previously, there was also some testing, as I understand it, that was done with regard to

looking across states with regard to hospitalization rates in relationship to the rates for -- again thinking about it for both the first measure as well as the second measure, as well as doing face validity testing across the expert groups and among Medicaid directors and other relevant groups.

CO-CHAIR BRISS: Questions? Comments? Concerns? Yes?

DR. HANRAHAN: I'd just say that, you know, I think it's really to the benefit of the development of this measure that they used focus groups to validate, because it really does enhance the face validity.

CO-CHAIR BRISS: So let's try voting. Oh, sorry. I have trouble looking at both sides, actually.

DR. ZUN: So the question of validity, I'm a little concerned about it because of the question of we don't know why they're not taking their meds. So can we make a validity decision if we don't know the

negative part of this? You know, we're judging it on their taking it or not taking it, but if they're not taking it, we don't know if there's a good reason why. And this is an appropriate -- are we measuring the right thing?

CO-CHAIR BRISS: That's a question for the developer. Have a comment?

DR. SCHOLLE: So if I had my druthers, for measuring outcomes for people with schizophrenia, I'd be looking at some sort of a functioning measure that would allow us, like we do in diabetes where we look at control of blood sugar and cholesterol and blood pressure and other things. But in behavioral health conditions we don't have a tradition and we may not have the tools to measure symptoms and functioning over time. And with the schizophrenia population we'd also probably be wondering a little bit about whether there is a level of improvement or a level of functioning at which one would want

to do that. So we're stuck with what's in the claims data that we can measure and where there's evidence base for a treatment.

And so, that's why these measures are trying to get at use the best evidence we have about what is good care for people with schizophrenia. If we wanted to focus on, you know, that -- if this were specified within an electronic health record where you could document patient refusal or if you could document the clinician's reason, then we would love to do that. But being able to actually measure care for schizophrenia in a health plan or a state using claims data, I think our experts felt, wow, this is worthwhile.

And our public comment from the HEDIS work seems to suggest -- and we also did public comment on these measures for states. And the public comment was very positive. We really need measures like this for schizophrenia. And from claims data, claims data are feasible, but we don't have anything

about that interaction, about what the providers was thinking or what the patient's reaction was.

CO-CHAIR BRISS: So, Tami?

DR. MARK: To sort of follow up on that comment, I think one issue is what is the potential harm if not having 20 percent of people adherent is a good thing? You know, most of that 20 percent has a good reason for not taking the medications, what is the harm in moving that to 90 percent and potentially over-prescribing? And I guess so the context, part of where I'm coming from is I do feel like there is an issue of over-prescribing of antipsychotics. It's not necessarily in schizophrenia. I think that's well established, but there are other areas where we're seeing increasingly concern about prescribing of antipsychotics in terms of dementia, in terms of bipolar disorder, in terms of sleep conditions, in terms of children.

And, you know, you couple that with the difficulty of diagnosing schizophrenia, particularly on one inpatient hospitalization and then you say we're going to push this measure over 80 percent and, you know, not really knowing if there's a good reason why that 20 percent is not taking it. You know, that's kind of my public health concern about this measure.

DR. SCHOLLE: Just to clarify, in the Medicaid data that we looked at, the average was 64 percent. We got two-thirds of people with schizophrenia who met our denominator criteria and were consistently on the antipsychotics. And it varied. You know, as low as 48 percent in one state to about 80 percent in New Hampshire. From 48 percent in Mississippi to 80 percent in New Hampshire. So you get a sense that there is a lot of variability in access and continuity of these medications.

CO-CHAIR BRISS: So, I think I'd

like to take David and then --

DR. PATING: Yes, so I just want to -- Dr. Zun's comment and Dr. -- and Tami's comments was that I'm struggling with kind of the so what, and then the policy implications of this as a stand-alone measure. And what I really liked yesterday is when the Joint Commission gave you a measure and a sub-measure. I think that actually it would be better to report these out as a measure and a sub-measure, because this as a stand-alone has policy implications, but you know, it's like what is NQF trying to say, that we've got a stand-alone measure, which goes to Dr. Mark's kind of comments. Do we want everybody on antipsychotics?

So I just think if they're really going to be paired, the analysis and then the approval, it should have been done together as a sub-measure.

DR. HANRAHAN: Taking into consideration all that you've said, Tami, the

data shows that two-thirds of people with schizophrenia or with serious mental illness do not receive or do not access treatment. So that is just a profound number. And I think that the presentation here of this particular indicator really matches with how unmeasured and how untouched we have gotten or we have not gotten to touch this problem.

So in that regard, yes, there's other ways that this measure will take on its own life form. But given that two-thirds of people with serious mental illness do not get access to adequate treatment, I think that this is really a strong support for taking this measure as is and intuitively moving forward and then moving it through the NQF process.

CO-CHAIR BRISS: I feel like some of our discussion is getting progressively farther away from validity, so I'd actually like to try to vote the validity stuff. And some of the issues that we're talking about

now seem to either fit better into usability or feasibility issues. So could we vote validity, please?

MR. WILLIAMSON: We will now vote on validity. This is a high, moderate, low and insufficient rating. Begin voting now.

We have 18 moderate, 1 low, and 1 insufficient.

CO-CHAIR BRISS: So usability?

CO-CHAIR PINCUS: In terms of usability, again a lot of this was determined through the focus groups with the different potential users in terms of, you know, ranging from state Medicaid officials and practitioners and including consumers in the focus groups. And there was endorsement of the notion of its usability for improvement as well as for accountability.

CO-CHAIR BRISS: Additional comments on usability?

(No audible response.)

CO-CHAIR BRISS: Why don't we try

voting?

MR. WILLIAMSON: We will now vote on usability. Begin voting now.

We have 4 high and 16 moderate.

CO-CHAIR BRISS: And let's move to feasibility, please.

CO-CHAIR PINCUS: So feasibility from the point of view of it being a claims-based measure makes it very feasible. But I guess the question I had with regard to feasibility is the issue of combining fee-for-service and plan-level data to get aggregation at a state level and the feasibility of that process.

DR. SCHOLLE: So actually the states are getting a lot of experience with that right now for the children's core set, the Medicaid children's core set. And I speak from knowledge. NCQA has a sub-contractor to the technical assistance contract that CMS provides to help states implement the specifications. It's a challenge because

states often have people in fee-for-service, primary care case management, other kinds of managed care arrangements. And what's envisioned is to get to a reporting that is representative of the state.

I think the claims-based measures are the easiest to implement in that way because you essentially can calculate a weighted average of people who meet the criteria in the different populations and the states should be able to know who is in managed care for different periods of time.

So we've seen states be able to apply specifications very similar to this at the state level combining across different data sources. It's not easy. And states are learning --

CO-CHAIR PINCUS: And also, what about duals?

DR. SCHOLLE: Well, that depends on the extent to which states have access to the Medicare data. And as you know, CMS had

made great strides in the past year to provide the access to the Medicare data to the states so that they could -- the states can combine the Medicaid and the Medicare data. And we're hearing from states, they're great delight at being able to do that. But I think states are really just learning how to do that and how to do it in real enough time to be able to do quality improvement from the results. But they're learning and anxious to do it.

DR. CARNEY-DOEBBELING: If you look at state by state and the growth of managed care, there are RFPs that have been answered throughout the country and more open right now to move the aged, blind and disabled populations and the duals into managed care programs. So more and more we will see less and less fee-for-service-type of data because groups like those that are typically in fee-for-service are M block being moved into managed care, which goes back to my earlier statement about this ultimately likely

becoming a HEDIS measure upon which health plans will be measured. And having led those efforts to bring those type of data together, they are huge. They require intensive resources.

The CHIPRA measures are still voluntary reporting. They're not required. Most states are reporting them, but it's a huge amount of effort to do that. And every state today looks a little bit different. So what may be in Montana's fee-for-service population may be very different than what's in Iowa's. So it's hard to compare states state by state by state because it's not always apples to apples.

DR. SHEA: I just had a question to understand because of the states. So there's 20 states that are in this, and are they able to easily submit this data and so forth in terms of the feasibility?

DR. SCHOLLE: So these are intended for voluntary use and we developed

this suite of measures thinking of the Medicaid core set and thinking of allowing -- getting to standardized measures that states could use, so for populations that have been under-represented in other national reporting activities like HEDIS.

So have we tested these in states that have a lot of managed behavioral health care or managed health care for Medicaid in people with schizophrenia? No, we have not done that testing. Are we confident that these specifications would work? Yes. I think our main concern would be about the denominator within any given health plan, and particularly we are aware of the -- moving more people with serious mental illness into these plans.

And we have proposed these for HEDIS for Medicaid plans. And we had very positive response to including these in the HEDIS for Medicaid plans. So we think it's a reasonable application that they can be used.

DR. MARK: Just to clarify, when they moved to managed care I think you would still get their fee-for-service claims for the prescription drugs. It's usually separately -- correct me if you're wrong. So you're going to get the prescription drug claims. So the issue is really the same that we had on the other measure, which is that you're going to get the antipsychotic. It's just you're going to have a smaller denominator because you're not going to pick up all the inpatient/outpatient claims where you're going to get the diagnosis of schizophrenia.

DR. SCHOLLE: It depends on whether the state is successful in getting decent encounter data from their managed care plans as well.

DR. MARK: Well, for the drug claims?

DR. SCHOLLE: No, the managed care encounter. So if the state has worked -- you know, so some states like New York and

Pennsylvania have been very successful in getting good encounter data from their managed care plans. And so they can actually -- they can combine or they can calculate that --

DR. MARK: Right.

DR. SCHOLLE: -- measures within the managed care encounter data from the health plans.

DR. MARK: But you're almost always going to get good drug data from all the states regardless of the managed care component. It's not that you're not going to get the fee-for-service drug data. It's really the other data.

CO-CHAIR BRISS: So anybody else with any issues that haven't already been discussed?

(No audible response.)

CO-CHAIR BRISS: So let's try voting feasibility, please.

MR. WILLIAMSON: We will now vote on feasibility. Again, this is a high,

moderate, low and insufficient rating. Begin voting now.

We have 4 high, 15 moderate, 0 low and 1 insufficient.

CO-CHAIR BRISS: So any last comments before we do the overall vote?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote on the overall suitability for endorsement. This is a yes/no question. Begin voting now.

We are still waiting on one response.

Could everybody try one more time and then we'll see if registered? There we go.

And we have 18 yes and 2 no.

CO-CHAIR BRISS: So I recognize we're running a bit behind. Because these two measures are so similar and they have many of the same issues, I'd like to quickly try to

get through the second measure in this set. I hope we can not re-litigate stuff that we just litigated. And I think that we should be able to make that go quickly. And we recognize already that there are overlap issues between the first and third measures that we're doing this morning. And so, as I understand it; Helen, you can correct me if this is wrong, but as I understand it, we're supposed to take this third measure on its own measures and the developers will work together subsequently to reconcile with -- to harmonize the measures, rather.

DR. BURSTIN: Yes, although I think we'll also try to have a discussion once we finish this next one to say what we think in fact are the best elements of both. So as they're combining into a single element, into a single measure, we actually get the best of both measures.

CO-CHAIR BRISS: And we'll do the best of both worlds discussion after a break,

not before, but first let's try to get through assessing the measure.

CO-CHAIR PINCUS: Okay. So impact is the same set of issues in data as we discussed with the previous two measures.

CO-CHAIR BRISS: All right. So anybody have comments before the vote?

(No audible response.)

CO-CHAIR BRISS: Hearing none. High, moderate, low, insufficient.

MR. WILLIAMSON: We'll now vote on impact. Begin voting now.

We have 15 high, 4 moderate, 0 low, and 1 insufficient.

CO-CHAIR PINCUS: Yes, although I'm not very reliable in putting on my microphone. But the gap issues are the same largely for the measure previously discussed in terms of -- because here, don't forget, we're dealing with the consistency of medication prescription.

CO-CHAIR BRISS: Right, and the

gap in the last one was actually not a lot, right?

CO-CHAIR PINCUS: Right.

CO-CHAIR BRISS: And I don't remember the gap. The data for the gap on this one was a little bigger, presumably?

CO-CHAIR PINCUS: Yes, they're bigger and they're bigger than on the measure two before because of the Medicaid population.

CO-CHAIR BRISS: So anybody want to comment further on this before we vote this one?

(No audible response.)

CO-CHAIR BRISS: So to the vote, please.

MR. WILLIAMSON: We will now vote on the performance gap. Again, this is high, moderate, low and insufficient. You may begin voting now.

We have 11 high, 6 moderate, 3 low, and 0 insufficient.

CO-CHAIR PINCUS: Then with regard

to evidence, the evidence here is focused not on a single prescription, but on the issue of for people who sort of are in the denominator, whether or not they've received consistent medication over time with a medication possession ratio of 0.8 or higher. And so, that this is looking at the use of maintenance treatment for schizophrenia.

CO-CHAIR BRISS: So I think the maintenance treatment issue has already been adjudicated and things like the age issue have also already been adjudicated. So anybody want to raise anything that we haven't already talked about?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote.

MR. WILLIAMSON: We will now vote on the evidence. This is a yes, no, insufficient evidence vote. Begin voting now.

We're still waiting on one response, please.

We'll try one more time. If your battery's dead, you should have a blinking red light, but I think everything was fine. We got it.

Okay. So we have 18 yes, 0 no and two insufficient evidence.

CO-CHAIR PINCUS: Reliability issues were the same essentially as for the measure we discussed previously in terms of looking at the states, across the states, and the same set of issues with regard to the Medicaid fee-for-service data that was available.

CO-CHAIR BRISS: So anybody want to make further comments before we vote?

(No audible response.)

CO-CHAIR BRISS: Good. Then let's vote, please.

MR. WILLIAMSON: We will now vote on the reliability. This is a high, moderate, low or insufficient vote. Begin voting now.

And we have 2 high, 16 moderate, 1

low, and 1 insufficient.

CO-CHAIR BRISS: Validity, please?

CO-CHAIR PINCUS: Okay. And again, similar validity testing was done both in terms of looking at the associations with other sort of concurrent types of measures, as well as in focus groups.

DR. ZIMA: I have a question.

CO-CHAIR BRISS: Please?

DR. ZIMA: And this is a question to developer. On 1936 is there a typo under 2-B-2.3 where the results are presented? It looks like high end use of antipsychotic continuity is correlated with cardiovascular screening and follow-up hospitalization, but you were looking at rates of hospital and ER use. On the testing results, 2-B-2.3 for 1936.

CO-CHAIR BRISS: I'm sorry, mic, please?

DR. ZIMA: Oh, with the other measures? Okay. Okay.

CO-CHAIR BRISS: I'm sorry, so can somebody summarize what that discussion just was?

DR. ZIMA: I was concerned somehow that there was a typo, but what they did is they jumped ahead and they looked at a measure we're going to discuss.

DR. SCHOLLE: We were showing correlations among the measures within the suite as well as, I think, in a -- we also have validity results. Did we show the -- if you look at the validity testing, we also had information on its correlation with hospitalizations.

CO-CHAIR BRISS: So, Jeffrey?

DR. SUSMAN: As we were just discussing, if you're getting regular care, it's more likely you're going to be screened and have higher utilization of other services.

CO-CHAIR BRISS: So anybody else have comments on validity?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote on validity. This is a high, moderate, low or insufficient vote. Begin voting now.

We have 3 high, 15 moderate, and 2 low.

CO-CHAIR BRISS: So moving to usability, please.

CO-CHAIR PINCUS: Again, usability, the same set of issues as we just discussed with the previous ones, you know, endorsement by the various focus groups and stakeholders for both accountability and improvement.

CO-CHAIR BRISS: So anybody have issues to discuss that we haven't already discussed?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote

on the usability. This is a high, moderate, low or insufficient vote. Begin voting now.

And we have 5 high and 15 moderate.

CO-CHAIR BRISS: And feasibility, please.

CO-CHAIR PINCUS: And feasibility, they're the same sort of issues with the one we just discussed previously with regard to the issues of combining data across Medicaid fee-for-service and health plans and the duals.

CO-CHAIR BRISS: Any issues to raise that haven't been raised?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote on feasibility. Begin voting now.

We have 1 high, 19 moderate, 0 low, and 0 insufficient.

CO-CHAIR BRISS: And moving to

overall approval, anybody have final comments before we vote?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote on the overall suitability for endorsement. This is a yes, no question. Begin voting now.

And we're still waiting on two responses. If we could get everybody to --

And the measure passes. We have 18 yes and 2 no.

CO-CHAIR BRISS: So that was breathtakingly efficient. I'll have to remember to hold people hostage before a break in my next chairing thing. So let's take an almost 15-minutes break and reconvene at five until, please.

(Whereupon, the above-entitled matter went off the record at 10:42 a.m., and resumed at 10:59 a.m.)

CO-CHAIR PINCUS: We're going to

get started by talking about the competing measure and harmonization discussion, so if people could begin to take their seats.

CO-CHAIR BRISS: And while we're getting people seated, just so that you can plan, I'm anticipating, because we're going to start losing people later this afternoon, that we'll want to manage lunch much like we did yesterday and take 10 minutes off and then eat and continue to work so that we can finish what we need to do today before we lose people for the --

CO-CHAIR PINCUS: And this time actually make it 10 minutes.

CO-CHAIR BRISS: Okay. So we're going to start off by having Helen kind of walk us through how you deal with this related and competing measure set of issues.

DR. BURSTIN: Great. So we skipped these yesterday. I know you talked about them in your pre-meetings, but just briefly, since we are faced now with competing

measures and this issue we think will be mitigated by the fact that the developers are going to work on combining it into a single measure, but just to remind you what we're talking about since it's relevant for later today as well, is that as we think about related versus competing measures, we're really talking about these two key issues; is it the same target population or a different population?

If it's the same target population and the same measure focus, then those are competing measures and those are the ones we really need to decide. I mean, essentially they're measures where there's shades of gray between them. It doesn't help the world when they're a different -- and even if they're at different levels of analysis; for example, a health plan measure versus a physician measure, clinician measure versus a hospital measure, it's still doesn't help anybody on the front line when they get measured in

different ways by different people. So that's what we mean by competing.

If there are different target populations but the same focus for the measure, then that's where we want to harmonize, meaning we understand that -- for example, we may get to this a little bit later, but as you looked at the smoking measures that you guys approved yesterday, there's a health plan measure, a clinician measure and a provider measure. There may be reasonable reasons why you need three of those, given the fact that they're very different data sets, different approaches. But you want to ensure at least that you're harmonizing on the measured focus. Are we defining things in a similar kind of way? So we'll get to that.

Next slide? So this is perhaps a little hard to see. It is for me. Is there any way to make it a tiny bit bigger?

So essentially, as we've been

going through this today, the first basic decision is does the measure meet all four criteria for NQF endorsement? So that's why you did a vote on both those measures saying yes they're both suitable.

So we then proceed to the next question, and the question is are there, you know, potential relating measures? We've already talked about that.

So comparing specifications is the next step. So at the conceptual level does it address the same concepts? We've talked about this. And then if they have the same concepts, can one measure be modified to expand to get the target population as indicated by evidence setting your level of analysis? That's really what we're talking about in this one. Can they take those two measures and in fact bring them together?

So the first question we would ask of this is, you know, if they are in fact similar and these are competing measures, can

they resolve stewardship for one measure? That's what they're going to work towards and that's, I think, what the recommendation of this group would be, to in fact recommend that one measure move forward.

So I think the key thing on the next one though, and we're really going to walk through in this issue is assuming that that wasn't possible and we needed to recommend the superior measure, which we're not going to do today, we would actually compare the measure evaluation criteria on each one of the ones you voted on to see if in fact there's one measure that's superior.

I think in this instance what we're going to do today is, instead, to take a look at this enormous sheet of paper that we're passing around; it would be great if you could share with the developers as well, of showing the measures side by side. So you could see perhaps are there elements of those two measures as they work to bring them into

a single measure that they can incorporate the best of each of those approaches, identify where there are particular concerns with those measures and then recommend the best measure.

And so, what we would usually do in terms of assessing for superiority is we would be walking through these. They are all tested, so that's easier. But you look under the second one there, reliability and validity, our preference of course is for measures of the broadest application that can get us the greatest populations and potentially address disparities as well, preference for measures that will be publicly reported, widest use or in-use. These are all in use, or should be soon. And then certainly preference for measures that are quite feasible. These are all claims-based measures, so I think we're quite equivalent there as well.

The one thing I'll also mention; and I believe we've got the developer on the

line, is there actually is in fact a third competing measure from Health Benchmarks, and the Health Benchmarks folks are on the telephone. That measure was not yet due for measure maintenance, so it'll come to you probably in the next phase of work. But to really be able to go through this discussion you need to in fact be able to see all three. And wanted to just have you have the chance to look at those, even though you haven't done a detailed evaluation of that third measure. These will certainly be information that Health Benchmarks would need to consider if they decide to bring the measure back forward for continued maintenance.

So I don't think we need to go into any greater detail there. Let's flip to showing the handout we just gave.

So on this enormous sheet of paper in front of you we have laid out to you -- thank you to Evan and Sarah for doing this -- laid out for you the different measures that

we're talking about today. And I don't know that we want to go through in great detail some of the differences here, but I think the key thing is as you look through this are there particular issues? And maybe just walking through it might be the simplest way.

So for example, if you look at the description, one of the first things that pops out is that the CMS measure is 18 and over and the NCQA measures are 25 to 64. So I think purely based on evidence is that a -- we've had a bit of this discussion to date. Is there a preference as they move these measures together to try to pick an age category based on evidence? And I'll stop there.

DR. SUSMAN: Now, I personally would like to see it at the 18 level for reasons we've discussed. Obviously there's pros and cons to this. And I think though that overall my sense is that the 18 and over cutoff makes conceptual sense and has a lot of validity, too.

DR. KELLEHER: Come back to a comment someone made earlier about the, you know, sort of clinical guideline about first break and early treatment, that I think we run the risk of substantially missing the boat for that group if we cut it at 25.

CO-CHAIR BRISS: Yes, essentially obviously one of the developers has started to make sort of evidence-based arguments. So you sort of talked about essentially the trade-offs between sensitivity and specificity and the potential benefits of earlier diagnosis versus the potential limitations of inappropriate labeling of young people perhaps. And so, I actually think you could -- the two developers might work together to kind of tee up the relevant evidentiary arguments better than perhaps has been done to date. As opposed to making a decision to day based on my gut, I'd rather see somebody make the evidentiary arguments in as clean a way as possible and bring that back.

CO-CHAIR PINCUS: Bonnie?

DR. ZIMA: Another advantage of going a little bit younger is you capture the transitional age youth, which is a big deal, and I think might also -- you could possibly cross-tab with eligibility codes for Medicaid to stratify by that population.

CO-CHAIR PINCUS: Does anybody else have comments with regard to --

Oh, I didn't see you.

DR. CARNEY-DOEBBELING: The trade-off for sensitivity and specificity might be looked at along with the algorithm for selecting a case, the one inpatient or two outpatients. So making that more stringent in the younger age may improve the specificity of schizophrenia in the younger age group.

DR. SAMET: On that note, I've been bothered -- I realize it doesn't do this by the 66 percent, or whatever the number is thrown out that has schizophrenia that never even get into this algorithm. And if -- and

I don't know if it does, but if by capturing them lower, if they have a first break, taps into that, that's another plus for that.

DR. HANRAHAN: Since we're offering our wishes here, I would like to also see the sensitivity and specificity testing between the groups to see if there are some groups that fall out of line with -- like for instance, by age. You know, before we set up a whole other process for adolescents and developing a separate indicator, let's do the testing, use the methods to determine that in fact there are sensitivity and specificity issues. And that goes for race or other disparities, too.

CO-CHAIR PINCUS: I'll call myself. I just with a quick look at this found five issues that -- in terms of need to be resolved. One is the 18 versus 25. And I tend to -- at least my recommendations, I tend to go with specificity over sensitivity here and would recommend that -- which is not to

say that the first break is not a big issue, but I think there probably ought to be a measure developed for first break specifically rather than try to stretch, you know, a measure that may not be ideal for that purpose. So I would strongly recommend that the measure developers consider developing some kind of first break measure.

The other thing, just the four other items I noticed that were different among these is the issue of having a paired measure that is having -- you know, capturing people who are just sort of -- just getting any sort of baseline kind of measure of any prescription and be able to compare it to the continuous medication one.

Number two, in terms of the denominator, there were differences in how the denominators would focus in terms of whether it required one hospitalization and two outpatient, or whether it was any two. And I'm not sure which is best, but I think doing

some performance, some sensitivity measurement around that would be useful.

The other difference I noticed that one required two prescriptions to get into the measure versus one prescription to get into the measure denominator. So again, to do some sensitivity analyses with regard to that.

And then I'm not sure whether the IMS Health, Health Benchmarks uses the 0.8 standard for medication possession ratio or not, or whether it just gives the overall average.

(Off mic comments.)

CO-CHAIR PINCUS: Huh? And maybe they could respond to that.

MS. FRANKLIN: Do we have anyone from IMS Health on the line?

I believe they had to drop off. She sent me a note.

OPERATOR: And if you need your line open, please press star one.

CO-CHAIR PINCUS: But anyway, that's something to look at in terms of -- you know, that's another point of potential standardization.

DR. PINDOLIA: Sorry. I thought we were just doing line by line, but if we're ahead to the numerator statement, I think it would be helpful for the group to understand what exactly is the proportion of day covered versus MPR, because those are two different type of calculations that are being used for 80 percent. If IMS isn't on the line, if CMS or NCQA can inform the group.

MR. CAMPBELL: Yes, this is Kyle Campbell from FMQA. We looked extensively at the differences between medication possession ratio and proportion of days covered when we arrived at this. Typically medication possession ratio, there are several flavors in the literature. Typically the days of supply are essentially summed for the entire period. So there is potential there to over-estimate

it here and it's when you have overlap in use of medications and switching and that sort of thing, whereas the days cover sets up arrays for each individual prescription and evaluates whether a day is covered with medication or not.

And I as I mentioned, with our algorithm what we do is we also adjust forward for those prescriptions where there's early refills. So we give credit for that as if the patient completely finished the first fill and then we adjust the start date of the next fill. So those are some of the differences between medication possession ratio and PDC. The other is how the period of measurement is defined. We've defined that as the index prescription, or the first prescription within the measurement year is the start date and then the end of the measurement period. In this case 12 consecutive months is our measurement period, is the period of time for which the PDC is assessed.

CO-CHAIR PINCUS: So I think what we wanted here is just to raise issues, because I don't think we're going to have time to have responses on each of these issues, because some of them will require data analysis and other sort of discussions. But I think what we want to do here is sort of raise issues and make sure that people look at them, you know, in the interim before they ultimately -- this ultimately gets reported back to us?

DR. BURSTIN: Yes.

CO-CHAIR PINCUS: Yes, because they report back to us. So David, then Tami, then Lisa.

DR. PATING: Hi, I just want to do a me too on the issue of the sensitivity. Even with two outpatient diagnoses sometimes you just -- the diagnosis of schizophrenia and you're going to track them for it, require them to be in the denominator for a whole year after that. Just it would be nice to have a

number on whether two is adequate or whether three is better, or one is enough, so -- because of the outpatient side of that.

DR. MARK: I think one measure used -- extended the coverage with the overlapping day supply, as you mentioned, and one did not. So that I think is a potential difference to look at. I think one excluded dementia and one did not.

And then just wanted to raise the issue again of if we should call this schizophrenia/schizoaffective as opposed to schizophrenia.

CO-CHAIR PINCUS: Lisa?

DR. SHEA: Yes, on the dementia point. And also one covered injectables and one didn't, and that can make a difference, too, in terms of continuity.

DR. SUSMAN: Just some further discussion among the developers about the feasibility/usability issues of combining state and federal data sets where some of us

were skeptical about how easy it was made out to be when we actually go to implement this more broadly.

CO-CHAIR PINCUS: Any other items that people want to bring up for the developers to discuss in the interim:

DR. NAEGLE: I --

CO-CHAIR PINCUS: Oh, Madeline? Okay.

DR. NAEGLE: I just wanted to reinforce Lisa's point that she notes that the injectable group was left out. And I would support that we try to get that group in when we're looking at adherence and that's why they're often on injectables.

DR. SHEA: One other -- the other measure that wasn't up for review excluded pregnancy, but the others didn't.

CO-CHAIR PINCUS: Vanita, do you have another point?

DR. PINDOLIA: It's for as they discuss this, I'm looking through the specs

and I don't see it in any of them, but if they could consider having if a medication hasn't been filled for at least six months or put some time window, then it falls out, because that's usually what we use to say that there is not an adherence issue. They've just been stopped. They no longer -- usually 180 days is traditionally used for that. That's something to consider at least.

CO-CHAIR PINCUS: Although I would worry about that if there was no sort of attendant outpatient visit during that time or they, you know, simply failed to engage the patient. So if there was an outpatient visit and no prescription, that's one thing. But if they just dropped out, then I would think there might be a failure to engage the patient.

DR. PINDOLIA: And you can combine the two.

DR. BURSTIN: Just as you're going through I think it would be best to try to

create a measure that can be used with the greatest number of levels of analysis. So specifically ensure we're getting clinician, health plan, state, state Medicaid plan, whatever the case may be. One measure applicable to all.

CO-CHAIR BRISS: Can we just give that advice to every developer in every subject?

DR. BURSTIN: Any questions from any of the developers about that discussion?

MR. CAMPBELL: Just one question. In the previous consensus project we discussed having a standard approach to adherence with NQF. And so I wanted to know, you know, at what point some of the considerations with regard to the methodological considerations that we make for this one chronic class of medications versus the rest of the NQF portfolio.

DR. BURSTIN: I think that's a great, Kyle, and I actually pulled up the

standard specifications for adherence measurement from the 2005 -- no, whenever that -- no, I was already here, so -- something like that.

We had a medication management project that went through exhaustively a whole series of adherence measures using both approaches. So I think it would be great if the developers combine your wisdom, take a look at the standardized specifications and see if there are some recommendations that could get made to bring it forward so we could actually bring that forward as a recommendation to this project. We could then add it to the measure guidance that we provide developers.

CO-CHAIR PINCUS: Cervical cancer screening, was that withdrawn?

So 1926 was withdrawn. Okay. Can the measure developer comment on the withdrawal?

DR. SCHOLLE: Yes. So since the

time that we submitted this measure the U.S. Preventive Services Task Force made new recommendations on cervical cancer screening for women. And so, NCQA is reevaluating our overall cervical cancer screening measure and this measure that is based -- that focuses on women with schizophrenia specifically. So we intend to bring this back after that reevaluation is complete.

CO-CHAIR PINCUS: Jeffrey?

DR. SUSMAN: I have to confess I didn't study this measure before it was withdraw, but just my bias is that we should be creating measures that are more generalized rather than to specific populations, so I would have to be convinced that there's some reason to look at cervical cancer screening differently in a population who has schizophrenia versus the general population of women. And if not, I'd prefer to have one way to screen for cervical cancer that's applicable to any population.

CO-CHAIR PINCUS: So that actually leads into the question I was going to raise. As we go into this next set of measures they're all -- not all, but several of them are specifically sort of a sub-setting or a stratification of an existing measure. And the question that I was going to pose to Helen is does that need to be proposed and gone through the entire process for NQF if you're essentially saying we want to sort of look at this like a disparities-kind of issue, that, you know, on a given preventive health measure is there a difference between this stratified group; i.e., people with schizophrenia as compared to the general population? Does that need to go through the whole process or can it simply be imbedded within the measure in some way?

DR. BURSTIN: Yes, and we've had a lot of these discussions. Actually I was talking with Sarah about it again this morning. I mean, I think there is a -- you

know, at times people could have different approaches to this. One could say that by pulling out a measure for a special population, particularly at a different level analysis, since these are intended to be state measures -- I don't believe the other ones rolled up specifically to state plans. You know, it certainly does call out the group in a way that's different than a strata.

But at the same time, there also could be an opportunity to make these sort of almost the analysis we talked about yesterday with the Joint Commission. You know, could these be measure 007, which is X screening for the general population? There's a, you know, XX7A specifically for schizophrenia. We're fine with that. We're happy to make whatever you think makes the most sense.

There were some slight differences to the measures we already have endorsed. So I think we at least need to talk about where there are differences. So for example, for

the diabetics, they have pulled in both A1c screening and LDL screening together for diabetics with schizophrenia, whereas those are different measures in our existing set. I personally like them together. It's probably time to have them together for everyone, but a bigger issue.

And there are -- again, this age cutoff issue we just talked about is another one where they've limited the populations going forward to be 25 and up to get at the issue of specificity to the schizophrenic population.

So probably need to talk through those issues a bit where there are differences. I guess, you know, as a general issue -- maybe could hear from Sarah, just getting her perspectives of where -- maybe even outlining where there are in fact differences of the remaining measures where you think we actually need to go through the exercise of truly, you know, reassessing a

measure already applicable for the full population. We did go back and check and as best we can tell there are no current exclusions in your currently endorsed measures for schizophrenics. So they're not left out of the current measure. They're just not called out in a specific way. So I'll let Sarah speak.

DR. SCHOLLE: So the rationale for this series of measures where we pulled in measures that are looking at physical or general medical needs for people with schizophrenia was because we know that their medical needs are often not addressed. And so there's a strong focus of our work. And really a strong focus of interest in our public comment and in our focus groups was we should be applying these measures to this population. At the time we submitted these measures, it was our understanding that we should treat them as separate measures and that there wasn't a way to kind of say and

here's a separate rate within the existing measures. So we'd be amenable to sliding it under there where that's possible.

Sometimes, and I'm not prepared to speak to this right now, but as we go through the measures we can call it out -- sometimes adding people -- having a rate for people with schizophrenia within that bigger group would require a separate effort. So for example, diabetes -- so actually even the cervical cancer screening, the diabetes screening measures. Those are exactly like the HEDIS administrative specifications for those measures.

So cervical cancer screening is a hybrid specification in HEDIS for health plans. Hybrid means you can either report using administrative data only if you feel confident, or you can draw a sample of people who meet the denominator criteria, and then you can show the numerator hits from your claims data and then go to charts to find the

other people.

Our concern is that if you did that for diabetes and you wanted to have within it a stratified rate for people with schizophrenia, you're 411 sample would include maybe one person with schizophrenia. So in fact, you would actually need a separate measure, separate reporting for that to work if you're drawing a sample in order to have enough information to report it. Sometimes there are other issues that have to do with how you define the denominator, whether you -- I think our -- if it had to do with what benefits were available and stuff like that. So there may be other issues.

But I think all the measures that are presented here for state-level reporting used a claims only specification, and that makes it different from the specifications that are presented for measures that have a hybrid specification at the health plan level that allow for drawing a sample and using

chart reviews to get to your denominator. So that would make it different. And, you know, we'd be happy to work with the NQF staff in trying to think about how to do this.

The list of measures that we pulled, I mean, we considered a number of physical health issues to be included here, and we our decision tree had to do with what measures -- was there an argument about a higher risk for people with schizophrenia? Was there an argument to be made about a disparity, that there's a disparity in care for people with schizophrenia? And so that's why you see the diabetes monitoring measures have to do with both concerns about the risk because people with schizophrenia tend to be on antipsychotic medications. So monitoring diabetes for that group is really important.

The cervical cancer screening was selected because of the evidence that -- prior evidence as well as the evidence from our field test about the truly low and

embarrassingly low rates of cervical cancer screening among women with schizophrenia and concerns about access to reproductive health care for that group. So there was a rationale.

And we didn't bring HIV screening. We thought about that one, but then there's really no evidence of higher risk or disparities, and so that measure didn't actually get into our testing phase.

So our panel was using that concern about higher risk, known disparity, and we presented all of these measures as separate measures, but we're happy to work with NQF to show, you know, how they could be a separate reporting rate. But know that in some cases that means it's actually a separate sampling process at different levels. So we'd have to think about how to represent that in the specifications.

DR. BURSTIN: And I think the final issue here is really the issue of the

end. So those measures go down to the clinical level currently of the NQF-endorsed measures. It doesn't sound like there would be a sufficient sample size, for example, for some of the clinicians to actually be able to get a rate on this. So again, it would be an issue as well for NCQA to give us advice about what level is appropriate for these measures given the sample size.

So I would suggest that we actually run through the measures. And, you know, we'll figure out the issues of whether they wind up being a subset of the original endorsed measure, or we could give it a new measure number; but either way, there's enough differences in them that I think it's worth at least having I think what could be a fairly quick discussion on some of these.

CO-CHAIR BRISS: I'm sorry--

CO-CHAIR PINCUS: --Just Peter then Jeffrey and Jeffrey. But realize that we're going to be running through these measures, so

if this has something to do with sort of how we run through those measures.

CO-CHAIR BRISS: Yes, so this is very quick. I just wanted to say that I'm very sympathetic to wanting to -- I'd be much more sympathetic to having some cervical screening measure that is then broken out by sub-population. I mean, it's straightforwardly true that sampling may have to be different among those, but at a minimum, sort of setting up, I suspect that within those measures now that there are all sorts of subtle differences in how the measures are spec'd and the defined details of the measure specs. And so at least lining those, the measure specs up so that you could really make more like apples to apples comparisons would be a major step forward.

DR. SCHOLLE: To the extent that we could rely on the existed specs, we did. And so, as we go through we can point out how they differ, and I'll rely on my team here to help point those out.

I did want to make one other comment. I mean, one of the reasons that it makes sense to have this Committee weigh in on this would be to say is this a rate that's important to reporting, I mean, so helping to validate that recommendation from our experts that we've actually heard supported in public comment that some of these measures are really important. Cancer, heart disease, diabetes are the leading causes of death in this severely mentally ill population. What measures could we apply? And having endorsement is important.

CO-CHAIR PINCUS: I mean, you know, we're going to have to go through these measures, so if there's something that's really important to make a point about the process, let me know.

So Jeff Samet, Jeff Susman and David.

DR. SAMET: Of course, really important. No, the thought is just that we --

you know, I'm wondering if we sub-population it to look at these issues depending on the fact that there's just disparity in that sub-population, or if there's something else going on, the something else being like the medications with antipsychotics that give you more of that diagnosis.

You know, there's disparity alone with -- you could find a number of things. I mean, we've talked about mental health substance use. Substance use, there's disparities around preventive services for cancer as well.

So that's sort of an uber issue to think about, but --

DR. SUSMAN: In addition to that I guess I'd make a strong push toward encouraging our developers to look at composite measures. Composite measures around cardiovascular health, routinely issued AF4Q, D5, looking at hypertension, diabetes, quality of care, etcetera. I think we really should

start bundling these together into meaningful groups. Clearly individuals with schizophrenia are at much greater risk, and let's be more holistic.

DR. PATING: Yes, lastly, along those same lines, I mean, I really like this direction for the policy implications about moving us towards behavioral health integration. With the ACOs there's no other measure that were moving towards integration except we're doing, you know, these medical exams on folks with mental health or substance abuse.

So I just would like to ask that the parallel process be looked at NCQA with regards to addictions as well so that we can move with both houses, behavioral health and to primary care.

CO-CHAIR PINCUS: So, yes, I think there's broad agreement that this is a very good strategy, a way of enhancing integration, shared accountability, and also that once you

think about how other populations and how one -- and that may come up in the discussion as we go forward.

But let's move on to measure 1932. Who's -- Nancy?

DR. HANRAHAN: So we have measure 1932 and 1934.

CO-CHAIR BRISS: But first let's hear from the measure developer.

DR. HANRAHAN: Okay.

DR. BURSTIN: I think we did.

CO-CHAIR PINCUS: Oh. Sarah, is there anything else you want to say about this specific measure? This is diabetes screening for people with schizophrenia or bipolar.

DR. SCHOLLE: Actually the diabetes screening measure is a new measure. Totally new. And this measure applies to people who have either schizophrenia or bipolar disorder and who were prescribed an antipsychotic medication. And so this is getting at the risk of metabolic disorders for

people who are placed on antipsychotic medications. So that's where the evidence is coming from.

So actually, in thinking about this measure, rather than looking at it related to the monitoring measure, I think I would encourage you to think about it related to the cardiovascular health screening measure, because those are the ones that have the same denominator: schizophrenia or bipolar disorder and antipsychotic medication. So it's that you were put on antipsychotics for a serious mental illness and that puts you at risk for cardiovascular or diabetes.

And so the diabetes screening is based on either the blood glucose or A1c test. And then the cardiovascular measure is looking at cholesterol tests. And those numerators will show up again or in a different form in the monitoring measures if we look at those as screening measures for this population.

DR. HANRAHAN: How different is

132 from 134?

DR. SCHOLLE: So the monitoring measure, it's the denominator that's different. Well, actually the numerator as well.

DR. HANRAHAN: Oh, I'm sorry. I see the monitoring versus screening now.

DR. SCHOLLE: Okay.

DR. HANRAHAN: Got it.

DR. SCHOLLE: Monitoring versus screening. Monitoring applies regardless of whether you had the antipsychotic medication. It's because you have the diagnosis.

DR. HANRAHAN: Right.

DR. SCHOLLE: Right? And that's the one that aligns with existing measures. The screening measure is really targeted to the risk from the antipsychotic medication.

DR. HANRAHAN: Thanks, Sarah. So the importance of this measure is very well documented. There's about two times the risk for diabetes in this population as the

population with schizophrenia in particular, and often bipolar illness gets medicated with the same antipsychotic medications. So the evidence is very strong that in fact this is an important measure to examine.

CO-CHAIR PINCUS: Is there additional comments or questions with regard to importance and impact? You're hesitating?

MS. PHILLIPS: I'm just wondering if we're worrying about the metabolic issues related to the antipsychotic medication, why are we limiting it to just two diagnosis codes? Because other diagnoses may be on these medications as well.

DR. SCHOLLE: We discussed this a lot. It came up in all of our focus groups and in our public comment I imagine it came up as well. So why not apply it to everybody who gets an antipsychotic medication? Our panel recommended that we think about it for people who are likely to stay on the medication, right, because it's the risk of being on it

for a long period of time. And so that's why we included schizophrenia or bipolar disorder where there was a concern about being on this medication or an expectation that you would stay on it, rather than just a single antipsychotic medication getting you into the denominator.

CO-CHAIR PINCUS: Peter?

CO-CHAIR BRISS: I don't think I was next.

CO-CHAIR PINCUS: Oh, Lisa?

DR. SUSMAN: I mean, my question is did you do any analysis with data around what the difference would be by spec'ing it out as the two diagnosis codes versus a broader group?

DR. BURSTIN: Use your microphone. Sorry. We can't hear you.

DR. SCHOLLE: I think that our testing focused on these denominators, pulling the denominators, because that's kind of the hardest step is to identify your denominator

population, then applying it.

CO-CHAIR PINCUS: Jeff Samet, are you -- have it up there, or from before?

(No audible response.)

CO-CHAIR PINCUS: Okay. Peter?

CO-CHAIR BRISS: Maybe I'm getting out of bounds about not evaluating the measure that's before me, but I mean, you know, if there are a set of things that one would like us to monitor in people with schizophrenia or people with schizophrenia that are on antipsychotic medications, why not do a bundle as opposed to taking one condition at a time?

DR. SCHOLLE: So could you take these two independent measures that we have that have diabetes screening and cardiovascular screening and make one measure that has two rates and then looks at whether both are met so you could say one, the other or both?

CO-CHAIR BRISS: Or a composite measure that says we would like you to do the

following five, or however many monitoring things in this population, and did you do them?

DR. SCHOLLE: Yes, you know, we're actually doing a lot of work on composite measures. I've been focusing mostly on children and looking at that. Now it gets to be a little bit of a challenge when you have you to kind of count up which measures apply, but I believe that the approach from NQF is that the individual measure has to be approved before we can create the composite. Is that no longer true?

DR. BURSTIN: It just has to be evaluated that it's appropriate within a composite. It doesn't have to be individually endorsed. So this discussion would be sufficient if you wanted to bring back --

DR. SCHOLLE: The measure is as did either of these --

DR. BURSTIN: Yes.

DR. SCHOLLE: You know, if you

wanted to bring in other things like cervical cancer screening or the diabetes monitoring measures, I mean, we're actually looking at different ways of calculating those composites. You can either create it at the state level where you just average the numbers, or you could create an individual person-based composite using an opportunities model of all the ones that apply to this person. What percent did they get?

Our experience is that -- and that actually helps you with some of the small numbers issues that we might face here, but it also -- I mean, there's for actionability people often like to see the individual items and to see, well, what's causing me to fail, or where is the biggest problem? So, but I mean, we'd certainly be amenable to thinking about how to group these measures into a composite.

You know, in terms of the claims data -- now remember, all of these measures we

designed and tested for claims data because we wanted states to use them immediately without having to do chart reviews. If we had had a broader scope that would allow us to do chart reviews, I think we would have been looking at other kinds of things like blood pressure control, and blood sugar control and, you know, the other kinds of composite measures that exist that get at more towards an outcome. BMI assessment was certainly something that was on our minds and was not feasible.

So really the question here is given that there's just kind of a limited number of items that one could do from the claims data, would there be value in creating a composite that just gets at cardiovascular and diabetes risk and maybe think about a person-based way of doing that where you'd either say, well, if they don't have diabetes, you put them in the screening measure.

Or if they do, you put them in the

-- you know, it gets complicated because the people who have diabetes -- I mean, the people with schizophrenia who are not on medications are not eligible for either the cardiovascular screening measure or the cardiovascular monitoring measure. You know, so people drop out. So, those were the kinds of complicated things we thought about as we --

CO-CHAIR PINCUS: I'm getting a little bit concerned that we're sort of getting beyond the issue of importance for this measure. I think there's important considerations that we want to bring out.

But maybe we can come back to them at the end and I -- because I think that ultimately, you know, how we think about this as a composite in relationship to other measures and how this relates to the varying denominators and so forth need to be considered.

Nancy?

DR. HANRAHAN: Yes, just one

comment --

CO-CHAIR PINCUS: Anything further about importance?

DR. HANRAHAN: -- about the importance is that most of the literature is related to schizophrenia and bipolar illness, about the importance of this particular indicator. That's why I would say that it's not generalized to all people.

CO-CHAIR PINCUS: Okay. So we're ready to vote on importance.

Oh, Leslie. Sorry.

DR. ZUN: Yes, I think I'm going to move over there so people can see me.

So I don't think there's a question about importance of diabetes screening, but the question is if we look at medical illnesses in the psychiatric population there's a number of disorders that are at a significant higher level than the general population. And so, this goes back to that connectedness or, you know, whether the

data's out there or not, you know? Because, you know, we know that cardiac disease is higher in smoking and substance use and those kind of things.

So, you know, does it make sense to just look at one item of importance rather than numerous items or -- thank you.

CO-CHAIR PINCUS: Are we ready to vote on the importance of this measure?

MR. WILLIAMSON: Okay. We will now vote on impact. This is a high, moderate, low or insufficient vote. And you may begin voting now.

And we're good.

CO-CHAIR PINCUS: We're on 1932.

Nineteen-thirty-two. Okay.

MR. WILLIAMSON: And we have 15 high, 4 moderate, 0 low, and 0 insufficient.

CO-CHAIR PINCUS: So, Nancy, let's move onto gap.

DR. HANRAHAN: I think the gap is

very well spelled out in the application by the increased risk and certainly the data that this population dies earlier than the general population by about 25 years. So there is evidence of a performance gap.

CO-CHAIR PINCUS: Any other comments or questions with regard to the gap? Lynn?

DR. WEGNER: If I can ask the developer, why did you choose four months as your point zero? Why did you not pick before at the initiation of medication?

DR. SCHOLLE: You have to give people time to get the medication, and so that you have to allow enough time to see the test after the medication is initiated, I think. As I recall, this is the --

DR. WEGNER: Could I also say that the -- did I get that?

DR. SCHOLLE: Where --

DR. WEGNER: Yes, go ahead.

DR. SCHOLLE: Could you just tell

us where you're reading from?

DR. HANRAHAN: I'm not reading from anything. I'm just asking a question about why you chose to start -- why you didn't choose to start before they started medication, before the medication started.

DR. SCHOLLE: Oh, to look for whether the test occurred before the medication --

DR. WEGNER: Right.

DR. SCHOLLE: -- rather than after it?

DR. WEGNER: Exactly.

DR. SCHOLLE: And should it count before or after? I think it's a matter of the window where you get it, but that's a good point. Our specification was trying to look at -- and can I ask one of you guys to take a look at the spec? Does it require that the test happen after or before?

I think it could happen either way.

DR. WEGNER: Okay.

CO-CHAIR PINCUS: So just to clarify, Sarah, so you're saying that it's four months on either side of the index prescription?

DR. SCHOLLE: Yes, let me just look. We're trying to understand where the four months is coming from, because we don't think that -- it's during -- it's one or more tests during the measurement year. There's not --

CO-CHAIR PINCUS: Nancy, where are you getting the four months?

DR. BURSTIN: Lynn.

CO-CHAIR BRISS: Lynn, I mean. Excuse me.

DR. WEGNER: It's 1-B-16, page 6.

DR. BURSTIN: Thank you, Sarah.

DR. SCHOLLE: Oh, oh, oh. That's the guideline recommendation. That's not the measure.

DR. BURSTIN: I see.

CO-CHAIR PINCUS: Okay.

DR. BURSTIN: That's good she found it. Good thing.

CO-CHAIR PINCUS: Okay.

DR. SCHOLLE: Okay.

CO-CHAIR PINCUS: So, Sarah, just to clarify: So this is during the measurement year and the measurement year begins with the prescription?

DR. SCHOLLE: It doesn't matter what order. It's any time during the measurement year. If they have the prescription during the measurement year, we're looking to see did they have the test any time during the measurement year.

CO-CHAIR PINCUS: What starts the measurement year?

DR. SCHOLLE: The measurement year. It's just the standard year. So it's a standard calendar year. We don't have a requirement that it's before or after it starts. It's just during a calendar was there

a prescription for the medication, and that gets you into the denominator. And the numerator --

CO-CHAIR PINCUS: Okay.

DR. SCHOLLE: -- did you have a diabetes screen?

DR. SAMET: So am I hearing that there's no sequencing here so that if I got a test done January 1 and then in September was put on the medication that that would count?

DR. SCHOLLE: Yes.

DR. SAMET: That doesn't make a heck of a lot of sense to me.

Well, but I mean, you wouldn't have the expenditure before you're doing the test. And the whole issue here, at least in large part, is that we're giving a medication which would increase the risk of the problems. And therefore, testing before doesn't tell us anything except, well, maybe this person was obese before they started or were not, or -- I mean, it's not bad, but it doesn't really

get to the point of why we're screening.

DR. SCHOLLE: But likewise, if you've been on the medication, once -- so I understand, if you just got on that one -- if you had that one time and you'd had an A1c test in January, would you repeat it, or the glucose test in January and you just got put on it, should it be repeated sometime after you get on the medication? But what we heard in our panels is that sometimes people would test before they put on the medication, as a baseline before they get on the medication. So and a lot of people are going to be on the medication for a long period of time. It gets to be very complicated. If we tried to tie it to the first prescription during the year, it makes it for a much more complicated specification.

CO-CHAIR PINCUS: So this is another example of sort of the sensitivity/specificity kind of issue. And while we're concerned about the medications,

there's also evidence that this population is, independent of the medications -- has a greater risk for diabetes and cardiovascular disease and so on.

You know, so I think that what Sarah's referring to is that, you know, the question is how much you capture. And it's a fairly low bar. It's, you know, establishing essentially a fairly low bar here, and the more you raise the bar, the lower the end you're going to get and the more complex it is, the administration of it.

So, I think, you know, one can think of it potentially as a kind of starting point. But I think this is something that we can discuss and debate and see whether it makes sense.

So, Caroline and Mady and Nancy.

DR. CARNEY-DOEBBELING:

Reconciling the comment that you just made with the comment that I was about to make, the specs say that it's only for at

least one claim for an antipsychotic medication. And so, that makes that bar exceptionally low because someone may get a claim briefly for bipolar but not need the antipsychotic ongoing for that. So that bar becomes extremely low. So reconciling that with what you just said, the other way to look at the measure, because there is an independent risk for metabolic conditions in schizophrenia is just to say diabetic screening for schizophrenia, period, irrespective of whether or not one claim for an antipsychotic was present.

CO-CHAIR PINCUS: Mady?

DR. CHALK: Which was precisely going to be my comment, but, you know, you have to make a decision about what you really think or what we really think is the most important risk factor. Is it the schizophrenia and bipolar disorder for obesity and diabetes, or is it the combination when you add the antipsychotic medication?

DR. CARNEY-DOEBBELING: Or does a -- a single prescription moves away from Peter's contention that these are chronic patients on chronic meds, because that's now how the spec is written.

CO-CHAIR BRISS: But the spec will pick up mostly chronic patients on chronic meds, right?

CO-CHAIR PINCUS: Yes. Lisa, Vanita, and then we'll come back to Nancy.

DR. SHEA: And this might be getting into the evidence piece, but the guidelines do support being on the antipsychotic plus having another risk factor where the illness itself isn't listed as one of those risk factors. So I was just wondering what the rationale was to sort of broaden the screening in this population.

CO-CHAIR PINCUS: Sarah, why don't we go through everybody and then have you respond?

Okay. Vanita and --

DR. PINDOLIA: I believe in other measures that we do with NCQA for HEDIS we usually tie a drug with a medical claim just to increase the probability that you're getting the right population. And I don't know if that's why this was linked. Or was it really linked to that because of the antipsychotics can increase your chance for metabolic syndrome and diabetes and you wanted to link that? So if maybe that could add some clarification of why you wanted the one drug. Because the one drug failed really makes no sense for this.

CO-CHAIR PINCUS: Okay. So, Nancy then Leslie and then we'll have the measure developer respond.

DR. HANRAHAN: So this is an indicator at a state level, and there is a performance gap according to 1-B-2 that is profound. You know, the mean value per state of the observance of getting this data is 12.1 percent and the maximum is 28.2 percent. That

is, you know, really --

CO-CHAIR PINCUS: And that's with a low bar.

DR. HANRAHAN: Yes. So, you know, I think that it's really good to get a sense of, you know, the specificity of this measure, but I think in many -- we're talking about a measure that is a state population level that really has some major implications for doing a quality indicator.

CO-CHAIR PINCUS: Leslie?

DR. ZUN: Is this the time to talk about the age group, or was that already discussed? Because if we're concerned --

CO-CHAIR PINCUS: We're really talking about performance here.

DR. ZUN: Well, but if you look at the -- if it starts at age 25, that's the performance as well as the tool. And if we're concerned about childhood obesity, why would we start at 25?

CO-CHAIR PINCUS: Okay. Sarah, do

you want to respond to that string of comments?

DR. SCHOLLE: Okay. So this discussion mimics every discussion we had with a focus group or a panel where we have different -- so I heard one recommendation we should screen everybody with schizophrenia for diabetes, and another recommendation that maybe we should focus -- if the risk is based on antipsychotic medication, then we should have people who've been on antipsychotics for a longer period of time. And so, and what you see is something in the middle, which is people that had one and had at least one prescription for the antipsychotic medication and have schizophrenia as a diagnosis.

So I think our panels have guided us to try to be somewhere in the middle and try to focus on something that's fairly easy to program in a claims-based measure, and because simplicity is valued in trying to apply these and especially if you are

concerned about people dropping out of your denominator and being at risk. So, it is what it is. That's the way the specification is delivered to you and we welcome the suggestion if you think it needs -- but this is similar to the kind of discussion that we've had before and this is where we landed.

In regards to the age group, again we were looking at this to focus on people with schizophrenia and at age 25. And so, clearly there are other issues and reasons to screen people younger, but that-- the rationale for this is based on having a serious mental illness and being exposed to an antipsychotic medication.

CO-CHAIR PINCUS: So, you know, we're actually voting on performance gap. And so are there any other comments on performance? Because I think some of the specifics, you know, we can get into in some of the other issues. Okay. So, Tami?

DR. MARK: Do you have a sense of

how well the glucose test and the diabetes test are coded?

DR. SCHOLLE: So you saw the -- and if you compared the results of performance rates for the glucose test in this population, it's at about 10 percent, we found in the claims data, compared to about 45 percent for the cholesterol test. So we talked about this with our panels about why it would be so much lower glucose tests could be done in the office and rather than being in a claim.

We were also concerned about that -- actually this whole panel of tests might be done in especially a mental illness setting where it's paid for through some other service and it's not being captured in claims data and that could be different from state to state, how they organize care, how they pay for care. And so that's a weakness of using the claims data.

Nonetheless, I think that the panel that -- and all our public comment, I

have to say, this is the measure everybody said important, important, important because of the risk of -- we -- of all the positive comments, the most positive comments came from the continuity measure and the series of measures related to diabetes and cardiovascular disease because of the higher mortality rates in this population.

DR. MARK: You know, there's a whole bunch of research now where people look, do chart review to validate, you know, what's in the claims. You just might look and see if anyone's done that on this piece.

DR. SAMET: I don't know if this will impact what you're using the claims data for this, because I don't know what happens when you have other point-of-care services, but hemoglobin A1c, which we're talking about here, is moving, isn't really there yet, but is moving in lots of settings to a point-of-care service. So basically it wouldn't require being submitted and therefore you may

not have a bill. And I don't know where that's going to be, but if this is all based on what may be the old system in a couple years, want you just to be aware. DR. SCHOLLE: And I don't know if I was clear. It's either the glucose test or the A1c counts to the numerator. And but we are aware of those problems. And these measures are specified here. For claims only where HEDIS measures have allowed for testing like this, we do allow for chart reviews just because of that same reason, that where it may not come through as a claim, and that would certainly be an issue. But overall generally these testing measures are often used in claims-based data only.

CO-CHAIR PINCUS: Okay. So let's vote on the issue of performance.

MR. WILLIAMSON: We will now vote on the performance gap. This is a high, moderate, low or insufficient rating. Begin voting now.

Okay. We're still waiting on two responses. There we go.

Okay. We have 15 high, 5 moderate, 0 low, and 0 insufficient.

DR. HANRAHAN: The evidence for this measure is based on Medicaid analytic abstract data which has, as we've already talked about, a lot of issues. However, it does capture the population and it also -- what's good about this indicator is that we have objective data, so we can go into HCPCS and other types of billing processes to gather in information about whether or not these lab tests were performed and paid for.

So at the state level we have evidence that in fact there is a large gap and that quality, quantity and consistency of the body evidence I'd say is in question. But, you know, I also believe that a measure at the state level has not been utilized before and for a lot of reasons that we've already talked about, the performance gap that we will get

data that will fill this evidence hole.

DR. KELLEHER: So did you have a -- your preliminary group, what did they think about it?

CO-CHAIR PINCUS: Dodi, what -- microphone?

DR. KELLEHER: I was just wondering, some of these had fairly good preliminary group evaluations and I'm just wondering what they were.

DR. HANRAHAN: I don't remember from the call that we had, but I think there was a lot of support for this measure. But the support is really coming from the belief that by creating quality measures that address the cardiovascular gap in care for this population is of value. So it's not that the evidence is great about this, because I don't think the evidence is great.

Okay. Preliminary we had four highs, two moderates and no lows or no -- that was for impact. Let me see. Evidence. Four

yeses and two nos. As a health outcome, absolutely not.

CO-CHAIR PINCUS: Right. That's a separate issue. You know, specifically looking at the issue of evidence and the association of the process and outcomes.

DR. HANRAHAN: Four yeses and two nos.

CO-CHAIR PINCUS: Yes, there were four yeses and -- and I guess the other issue is that it is also incorporated into multiple practice guidelines.

DR. HANRAHAN: It is. It is. They used the American Diabetic Association Practice Guidelines to build this measure.

CO-CHAIR BRISS: Okay. Vanita? Caroline and Jeffrey, are you making comments?

DR. PINDOLIA: Just one point of clarification for myself. When I was reading for the evidence, the Marder study is the one that was very specific for patients with schizophrenia and looking at BMI and diabetes.

I don't recall the ADA guideline specifically right now in detail to remember. Do they have a subgroup of schizophrenia guidelines built in? Because that's the only study that's being used to support the whole thing. Is that correct, or did I miss something?

DR. HANRAHAN: There are studies above this that are give evidence about the prevalence or estimates of prevalence of the disorder in this population. Down here there's not a lot of evidence displayed.

CO-CHAIR PINCUS: Yes, as I understand it, the Marder thing is not a study, but it's actually a --

DR. HANRAHAN: Practice guide.

CO-CHAIR PINCUS: -- sort of a recommended guideline.

DR. HANRAHAN: Yes.

CO-CHAIR PINCUS: Sarah, do you want to respond about the guideline recommendations that are specific to this --

DR. SCHOLLE: Right. So the

Marder guideline is specific about physical monitoring of patients with schizophrenia and recommends fasting plasma glucose level or A1c value monitored four months after starting an antipsychotic and then yearly, ongoing. And then the ADA guideline recommends testing for people who have one or more risk factors for diabetes. And so this comes in under those risk factors.

CO-CHAIR PINCUS: Other issues with regard to evidence? Jeff?

DR. SUSMAN: I think with the evidence that's presented and the evidence that's generally available there's a key factor or series of factors that are missing, which are the presence of the BMI or metabolic syndrome or some other thing that would trigger at least a narrower group of screening, I mean, if we're just trying to stick to what the evidence is that's presented to us.

So, you know, I think the

directionality, the intent is good, but I'm not clear that the evidence that's being presented are very high level, particularly as you've structured the measure given the limitations of the administrative data. I'd be happy to hear your response.

CO-CHAIR PINCUS: Why don't we wait and hear David's comment?

DR. EINZIG: Just a comment. I'm all for screening. It is very important. But just a question. Are people going to be falsely reassured with a normal hemoglobin A1c and a normal fasting sugar given the delay that can happen until the blood sugars actually go up, and was that discussed?

CO-CHAIR PINCUS: And, Peter?

CO-CHAIR BRISS: Yes, I sort of had the same reaction just on an evidentiary basis in terms of what's presented, you know, not nearly as strong as some of the stuff we reviewed yesterday, on which we were quite a bit harder, I think.

CO-CHAIR PINCUS: Sarah, do you want to reply?

DR. SCHOLLE: Okay. So the guideline recommendation that we're relying on is I think the Marder guideline about physical monitoring for people with schizophrenia, and which -- and remember, our denominator is people aged 25 with schizophrenia. So we're still -- it's that sensitivity/specificity issue.

We focused on annual testing rather than testing around -- you know, at a specific point in time after a medication is prescribed because that's challenging to implement. In our testing we tried to do that and there were a lot of issues around how do you define how long they've been on it and when do you do the test, and how do you implement that? And the year's worth of claims data where you have to wait until they meet the eligibility criteria.

So this was our best attempt to

try to address those issues of concerns about the risk of metabolic syndrome and diabetes and cardiovascular disease in this population. We were limited to claims data, so we don't have information on BMI. And so our choices were: do we use this measure because we're concerned about diabetes and cardiovascular risk in people with schizophrenia; or do we wait until we have EHRs in especially mental health clinics so that we can get data on BMI and from EHRs to be able to look at this; or do we require people to do a separate sample of people with schizophrenia to look at their health needs?

So those are the kinds of choices that we made, and we focused on a screening measure. So this kind of test is what is typically done in people who have diabetes from claims data. So this is aligned with our use of claims data to monitor this kind of testing. So we have pretty good confidence that claims data are weighed to monitor this

kind of testing.

Is this the right population? So I think the questions about evidence are is this the right population that should be tested for diabetes so that somebody can take action to try to address their diabetes? And the glucose test and the A1c tests are the tests that are available to do that. Now, whether the doctor or clinician would use that information to institute treatment or to monitor, that would be up to the clinician to monitor the patient.

CO-CHAIR PINCUS: I'm getting a little bit worried that we're sort of going back over stuff. So is there new stuff that people are bringing up? Jeff or Peter?

CO-CHAIR BRISS: Just one quick question. So the 25 comes from the precision of the schizophrenia diagnosis. It's a little younger, at least in general, patients than you generally see diabetes diagnoses, although it keeps pressing younger. So the other issue

about the evidence is: is that the right lower age bound if what you're really trying to do is pick up diabetes that ought to be treated?

DR. SUSMAN: Another way to look at this would be to do an analysis of the prevalence of your risk factors as Marder lays out and convince us, convince me at least, that there's such a large prevalence of these risk factors that your approach, using the limitations of administrative data, is valid. You know, there's 60 percent of the population has one or more of these other risk factors, therefore we're making the evidentiary leap that it's justifiable to do population-based screening on these criteria.

As it is now, I don't see that high level of evidence that we've seen in some of our other discussions. It's not that I don't believe it. It's probably what I would do in practice, but again, I'm just trying to make a distinction between levels of evidence.

CO-CHAIR PINCUS: Jeff Samet?

DR. SAMET: Just a brief point. I'm a kind of major adherent of the U.S. Preventive Task Force when I think about screening and talk about what's appropriate. And I was waiting for Bernadette to make the comment, but she didn't. So this doesn't come close to sort of stepping through those steps, and so that worries me a little bit.

CO-CHAIR PINCUS: Tami?

DR. MARK: It looks like the FDA also has recommendations regarding this. Did you cite those or --

CO-CHAIR PINCUS: Could you say a little bit more about that?

DR. MARK: I think the FDA made recommendations that -- I'm just looking it up -- that patients treated with atypical antipsychotics at that time, and it may apply to typicals now -- patients with preexisting diabetes, monitor or regulate for glucose control. Patients with risk factors for diabetes get fasting blood glucose at baseline

and periodically throughout treatment. All patients initiated on atypical antipsychotics get monitored for hypoglycemia and have a fasting blood glucose in patients who develop symptoms of diabetes.

CO-CHAIR PINCUS: Okay. Helen?

DR. BURSTIN: Two comments actually, not as an NQF staff member, but as somebody who used to oversee the task force.

So the first is that the task force makes recommendations specifically for the general population, so they wouldn't oftentimes go in and do a recommendation for a subset population at risk. So that's the first think.

The second thing is actually just more of a question as a primary care doc, which is that, at least in my experience and from the evidence I've seen, this is much more of an issue for those on new antipsychotics. And I wondered for the psychiatrists and others in the room whether that's something

that should be considered going forward for stratification. Because that's just where -- I mean, it's just been dramatically different watching patients over the last five years and the preceding decade of just dramatic weight gain in incredibly young patients with no other risk factors just ballooning in front of my eyes. So just a question for the psychiatrists. Sorry to be that explicit.

CO-CHAIR PINCUS: Okay. So I want to move this along because we do want to have lunch sometime today.

Yes, so Caroline, David and Nancy, again, sort of new issues that haven't been brought up before.

DR. CARNEY-DOEBBELING: Two quick comments. The first is what I'm hearing in the room is not that no one believes that this is the right thing to do, but that there hasn't been due diligence in presenting the evidence.

Secondly, I was going to echo the

same thing that Helen said. The risks to antipsychotics seems to be in the evidence weighted more towards second generation antipsychotics. So if the spec reads "any antipsychotic," that may not be the risk factor. However, schizophrenia in and of itself is a risk factor. So either link it to the right antipsychotic group or just make it generalized to the population of schizophrenia. And bipolar is a whole different discussion.

CO-CHAIR PINCUS: David?

DR. PATING: No, my comments were similar that the risks for diabetes are specific to the medicines and specific to schizophrenia. I think we're confusing what's in -- with trying to do like an indicated preventive health measure and a selective one and taking a medium ground. So are we measuring the medication effects or the general effects? I think this is a reasonable compromise, but I do think the evidences

support where this is going. But I think the measure itself has some probable usefulness.

DR. CARNEY-DOEBBELING: And if the measure is intended to be related to the fact that someone's on an antipsychotic, then I would recommend, I think as someone earlier did, not linking it to a diagnosis, but rather a treatment period of three months or six months, or whatever that might be on a second generation antipsychotic. So you pull in everybody who's at risk because they're on that drug.

DR. PATING: Yes, again, is it a target or is it more intermediately selected, and that's just not clear what is --

CO-CHAIR PINCUS: So, Jeff, do you have a comment or are you up there from before?

(No audible response.)

CO-CHAIR PINCUS: So let me take my chair hat off and I'll make a comment. Then, Nancy, maybe you can sum up.

From my point of view, it's clear that this is a high-risk group that screening makes sense for this group. The evidence is there. And I think the way in which this measure is specified represents a compromise across multiple different parameters that different people have different things that they would like to see in this measure that would ideally be piggybacked, you know, whether it's getting at the younger group or getting to a more specific group or leading to, you know --

But my reading of the evidence is that people with schizophrenia and bipolar disorder are at greater risk for this condition. Having medication enhances that risk and that there's multiple recommendations for this; you know, in some cases a broader population, in some cases a narrower population to be screened. And screening in and of itself is not going to solve the problem, but it's a necessary but not

sufficient component of that solving that problems. And so to my -- I think it makes sense.

DR. HANRAHAN: And briefly just --

CO-CHAIR PINCUS: Nancy, you can summarize and then let's get to voting.

DR. HANRAHAN: Yes. Well, I think you did such a great summary of that, Harold. I'm not going to repeat that.

I just want to point out that in regard to the evidence I think at the general population level it is well documented that type 2 diabetes is prevalent. And we have all the things that you just said about the specialty population, that to me is strong enough or evidence to at least note that this is of moderate to -- a moderate quality in the quantity and the quality of the evidence.

CO-CHAIR PINCUS: So are we ready to vote?

DR. HANRAHAN: Yes.

CO-CHAIR PINCUS: Okay.

MR. WILLIAMSON: We will now vote on the evidence. This is a yes, no or insufficient vote. You may begin now.

All right. We have 12 yes, 1 no, and 7 insufficient.

CO-CHAIR PINCUS: Okay. So let's move to reliability.

DR. HANRAHAN: Reliability is -- I think we've said a lot about the reliability, and I don't know we have to -- do we need to go over again, because it's the same data. It's the same, you know, argument for reliability. And, you know, I think it's probably not -- I don't need to say more.

CO-CHAIR PINCUS: Other people have comments or questions with regard to reliability?

DR. SUSMAN: Could you just say a word about the reliability? I mean, we've talked about all kinds of important issues, but the reliability of the measures, either test/retest, or however they calculated

reliability.

CO-CHAIR PINCUS: So if you go to 2-A-2.3, right up there, it's summarized.

DR. HANRAHAN: Sure.

CO-CHAIR PINCUS: On the screen, if you can see it.

DR. HANRAHAN: So the results showed that there was good test/retest reliability. Overall, 4 of 16 states had no change in performance quartile. That's pretty good. State performance for this measure correlated at a 0.33 level and accounted for 11 percent of the variance in the 2008 scores. It's not bad.

DR. SUSMAN: Well, what about the other eight states? I mean, that doesn't sound horribly convincing to me, I'm afraid.

CO-CHAIR PINCUS: So, Sarah, do you want to comment?

DR. HANRAHAN: Yes. And these are percentiles.

DR. SCHOLLE: As a whole there are

a number of things that could account for having lower or different rates and being in a different strata. Essentially, the way that -- we were limited in the ways that we could look at reliability. There's a number of ways to look at -- so the way that we chose was saying, okay, does this measure perform comparably from one year to the next?

CO-CHAIR PINCUS: Yes, this is exactly the same issue on a previous -- in terms of this is looking at a narrow fee-for-service --

DR. SUSMAN: I get that. I'm just --

DR. SCHOLLE: If you started in --

DR. SUSMAN: This was a different level of states that we're moving. I mean, really, there was 75 percent who've changed strata. And I believe it was much narrower the last time, but anyway.

CO-CHAIR BRISS: No, it was this bad once before.

DR. SUSMAN: Yes.

DR. SCHOLLE: And none of the states changed more than two strata. And remember, we're looking at diabetes screening. A state could have implemented a new program for their -- within a specialty mental health setting that -- where they were paying for it separately. We know those things have been happening. And that could account for changing whether you were in the top quartile or the next quartile. So practice could have changed and we wouldn't have reliability from one year to the next. It's limited. It's the only way we could get it from the data sources that we had.

CO-CHAIR PINCUS: Other comments, questions about reliability?

(No audible response.)

MR. WILLIAMSON: We will now vote on reliability. This is a high, moderate, low or insufficient rating. Begin voting now.

We have 13 moderate, 5 low, and 2

insufficient.

CO-CHAIR PINCUS: Validity?

DR. HANRAHAN: This is similar to the other studies. The use of focus groups to look at this to determine the face validity is really a nice method to use. And they -- the measure had a minimum value of 2.3.

I'm going to let you talk about this, Sarah. Just tell me what --

DR. SCHOLLE: So I think we looked at different ways of thinking about validity; obviously face validity from our experts and our focus groups, but we did look at how this measure was related to hospitalization. I want to call your attention to, you know, thinking about how screening is related to hospitalization. So what we found is that there was a higher hospitalization rate in states that had poorer screening, okay, so when the states that were in the bottom quartile of the screening performance had about 24 percent of their enrollees with

schizophrenia were hospitalized compared to 18 percent in states that were in the top quartile of performance on this measure.

DR. CARNEY-DOEBBELING: Hospitalized for diabetes or cardiovascular condition?

DR. SCHOLLE: It's for both.

DR. CARNEY-DOEBBELING: What were they hospitalized for?

DR. SCHOLLE: Oh, hospitalized for schizophrenia.

DR. CARNEY-DOEBBELING: But not for --

DR. SCHOLLE: We were looking at hospitalization as a could you make an argument that this measure, high performance on this measure was related to an outcome? Our outcome was hospitalization. What we found was it was and it --

DR. CARNEY-DOEBBELING: No, my question is in that group were they hospitalized for a diabetic or cardiovascular

complication, or was it for a psychotic break of some sort?

DR. SCHOLLE: Don't know. They were hospitalized. It looks like hospitalized for schizophrenia, but let me just double check.

CO-CHAIR PINCUS: I mean, this is a test of concurrent validity, so it's not a -- you know, and to see if the measures seem to go with other sort of things that you think would correlate. You know, it may have less to do with actually the performance of this act than it does with the fact that the people are more engaged in care. If they got a, you know --

CO-CHAIR BRISS: Or low-performing states are always low-performing states.

CO-CHAIR PINCUS: Right. But, yes, so in terms of -- but so again it's a modest kind of evidence for validity. But that's on top of the face validity in terms of the focus groups.

Any other comments with regard to validity?

(No audible response.)

CO-CHAIR PINCUS: Okay. Ready to vote?

MR. WILLIAMSON: We will now vote on the validity. This is a high, moderate, low or insufficient rating. And you may begin voting now.

We have 14 moderate, 5 low, and 1 insufficient.

CO-CHAIR PINCUS: Nancy, now usability?

DR. HANRAHAN: Let me just make sure I'm in the right place here. The measure was deemed useable and feasible. It comes from, you know, administrative data which is very accessible and documents really to a granularity level that's really good and it -- by documenting whether or not a lab test was done. It was also deemed feasible by focus groups, public comment and technical advisory

group.

CO-CHAIR PINCUS: Other comments with regard to usability?

MR. WILLIAMSON: We will now vote on usability. Begin voting now.

We have 4 high, 14 moderate, 1 low, and 0 insufficient.

CO-CHAIR PINCUS: Nancy had also discussed feasibility just before. Are there other comments with regard to feasibility?

(No audible response.)

MR. WILLIAMSON: We will now vote on feasibility. This is a high, moderate, low or insufficient rating. You may begin voting now.

We have 4 high and 15 moderate.

CO-CHAIR PINCUS: Before we vote on overall suitability for endorsement are there any final comments or questions?

(No audible response.)

MR. WILLIAMSON: We will now vote on the overall suitability for endorsement.

This is a yes, no question. Begin voting now.

We have 13 yes, and 7 no.

CO-CHAIR PINCUS: Okay. So when do you want to come back?

CO-CHAIR BRISS: Ask for public comment.

CO-CHAIR PINCUS: Oh, yes. We got to ask for public comment? Okay. Anybody on the phone or in the room -- well, anybody in the room that wants to make a public comment?

(No audible response.)

CO-CHAIR PINCUS: Okay. Anybody on the phone that wants to make a public comment?

OPERATOR: Star one over the phone to signal.

(No audible response.)

CO-CHAIR PINCUS: Okay.

OPERATOR: And no one has signaled over the phone.

CO-CHAIR PINCUS: Okay. So let's reconvene at a quarter of 1:00.

CO-CHAIR BRISS: And just bring food back to your place and then we'll eat it --

CO-CHAIR PINCUS: Okay?

DR. BURSTIN: All measure members of course are welcome to stay and eat. The few others that are here, just go ahead and eat.

(Whereupon, the hearing was recessed at 12:33 p.m. to reconvene at 12:45 p.m. this same day.)

A-F-T-E-R-N-O-O-N  S-E-S-S-I-O-N

12:49 p.m.

CO-CHAIR BRISS:  So we're going to restart with No. 1927, Cardiovascular Health Screening.  And so if the developer has any additional teeing up comments that haven't already been made?

(No audible response.)

CO-CHAIR BRISS:  So again, any additional teeing up comments from the developer?

(No audible response.)

CO-CHAIR BRISS:  Okay.  So since this has a lot of the same issues as the last measure, we're not going to do any additional teeing up.

So, Bonnie, maybe you can walk us through the conversation.

DR. ZIMA:  Did you want the developer to speak first?

CO-CHAIR BRISS:  Well, she's going to pass since it has many of the same issues

as others in the suite.

DR. ZIMA: Okay. I'll try not to be too redundant. Nineteen-twenty-seven is cardiovascular health screening for people with either schizophrenia or bipolar and who are also prescribed antipsychotic meds. This, too, is a process measure. Data source is administrative claims with a level of analysis at the state level. And of course the steward is NCQA.

Just in going over sort of the operational definition, the numerator is basically one or more LDL cholesterol screenings identified by a procedure code. And in our work group the developer helped clarify that that procedure code means done not ordered.

The denominator we've reviewed before. And again, some of the limitations include that persons enrolled in managed care Medicaid programs are not in here. However, I did see in the application that there is an

exception that beneficiaries for both behavioral health organizations or managed care organizations that were deemed to have useable data were actually included. Of course, you have the bias of not including people who lose coverage or are receiving Medicaid under other eligibility codes. The age range is the same as what we've discussed before. And again, how they created who meets criteria for schizophrenia or bipolar disorder is exactly the same as what we described.

I think the thing that's a little bit different is the denominator exclusions, which include persons receiving an intervention, whether it's an coronary artery bypass graft or percutaneous PCI during the measurement year, or at least one year prior. There's also outpatient or inpatient care for ischemic vascular disease or receiving care for CHF or history of prior MI during the measurement year.

I think the implications of this

additional bells and whistles is that it will exclude those with cardiovascular disease that have been detected and that this might also impose a bias towards those who are more likely to access care for a heterogeneous group of reasons but be continuously insured,, but this is actually a very common limitation for claims data.

As far as evidence, really the rationale for high impact is twofold on significance. One, that there's greater lifestyle risk factors among this target population; two, that there's high non-treatment rates for hypolipidemia among persons with schizophrenia. There was no mention of whether this is true for persons with bipolar disorder. And that as stated before, some antipsychotic medication classes have greater risk of elevated LDL and triglycerides.

There's an assumption here that improved screening will lead to proper

diagnosis and treatment, but this is a common conceptual leap. And oftentimes in these debates the rationale is that nevertheless improving screening is an important first step.

And that's it for impact.

CO-CHAIR BRISS: Yes, so comments on impact?

(No audible response.)

CO-CHAIR BRISS: Yes, I guess I have a quick question on the impact. So in the general population if I were going to pitch cardiovascular health screening, this wouldn't have been the first priority on my list. And in fact, it wouldn't have been in the top few. So we know that --

CO-CHAIR PINCUS: What do mean by "this?"

CO-CHAIR BRISS: Cholesterol. Yes, so in the general population hypertension's a bigger deal. In this particular population tobacco is a huge deal.

So maybe it's just an alignment issue between the title of the measure and the issues that you're addressing. And I understand that the meds weighs cholesterol and/or triglycerides, but can you comment a little on why you chose these particular targets of change?

DR. SCHOLLE: Because we're using claims data the only one of those three risk factor we could look at is cholesterol test. From the claims data we considered -- and we had lots of encouragement of thinking about a tobacco assessment and counseling measure and that's -- we actually investigated whether that could be done through the claims data, but we heard clearly that it not -- and you wouldn't be able to know who was a smoker anyway. So that current HEDIS measure on smoking cessation is a serving measure. So that didn't seem feasible as an approach within this suite where we were actually directed to develop measures for claims data.

And likewise, BMI. There is a BMI

assessment measure for the general population, but that is a hybrid specification and we have very low rates from plans that choose to submit administrative data only to HEDIS, have essentially zeroes and they submit those. So since we are working with claims data, we could not focus on BMI and tobacco.

CO-CHAIR BRISS: Any other questions or comments?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote impact.

MR. WILLIAMSON: We will now vote on impact. This is a high, moderate, low or insufficient rating. Begin voting now.

We're missing one response.

Okay. We have seven high, nine moderate, two low, and one insufficient.

CO-CHAIR BRISS: Performance gap, please?

DR. ZIMA: The performance gap on this was not very wide. The 25th percentile

was 42 percent. Median was 46 percent. The 75th percentile was 51 percent. The range was 7 to 63 percent and estimates of precision were not presented.

CO-CHAIR BRISS: And it's low in every age group. Comments? Yes?

DR. CARNEY-DOEBBELING: I know that we're not supposed to necessarily discuss how hard it is to get a measure done. However, that being said, given how low this measure is in the general population of people, try getting someone with schizophrenia to come in fasting. It's nearly impossible.

And I know that we still need to look at things to move this forward, but there are a whole bunch of process issues associated with getting this test done as reflected by the general population. And so, I have concerns, serious concerns about using this to grade states and to grade health plans right now until we can come up with methods or payment structures or redesign of clinics that

make it much easier for persons with schizophrenia to access and LDL-C.

We have gone back and forth with physicians in the State of Indiana. And because most of the cholesterol tests are run as a panel -- LDL-C in and of itself doesn't have to be fasting necessarily, but the others do and they typically get run as panels. So the advice back to the patient is you have to come in on another day and get this test done and be fasting. It's a very difficult thing to do in this group.

I was just reminded that's a usability issue. I'm sorry.

CO-CHAIR BRISS: So this is specifically about the performance gap. Anybody else have issues specific to the performance gap?

That's okay. We'll eventually get to usability.

(No audible response.)

CO-CHAIR BRISS: Let's try voting

on performance gap, please?

DR. SCHOLLE: I just want to clarify. Is your concern about the fasting, a fasting cholesterol, because the measure is an LDL test and it doesn't require the fasting?

DR. CARNEY-DOEBBELING: That's what I just said. But most of these are done as a panel of HDL, VLDL, LDL, and those others do require fasting. So to comment, they get aggregated at the billing level often. Docs rarely will just order an LDL as a stand alone and the advice typically and what happens in most FP and in general internist's office is to come back and get the whole cholesterol panel done, which has to be fasting. So then they would have to specifically order an LDL-C without the others when the others are important as well.

CO-CHAIR BRISS: So let's table further discussion on that, any needed further discussion on that until we get to actual

usability issues. So anybody else want to comment specifically on the gap?

DR. PATING: I'm just actually not quite sure what the gap is. I see the performance statistics, but -- and I saw there's an article that says 25 percent less cholesterol screening in schizophrenics than the general population. Is that the gap we're looking at, or is there any other number? Because they're producing a lot. Compared to what? I guess that's what I'm looking at in terms of the gap. Compared to what we think it should be, or general population?

CO-CHAIR BRISS: All right. I see that Helen isn't here. I thought gap in this context could either be performance gap in this population, which it seems to me to clearly meet, or performance gap across populations, on which I didn't see data presented here. So but at least the first one the performance is low in this population.

CO-CHAIR PINCUS: I think it's

absolute and in comparison.

CO-CHAIR BRISS: Yes?

DR. HANRAHAN: My interpretation is that the gap is at the state level, that the attention to this population's cardiovascular health is not being attended to, and that's a huge gap.

CO-CHAIR BRISS: At every level. So are people ready to vote on that issue?

(No audible response.)

MR. WILLIAMSON: We will now vote on the performance gap. This is a high, moderate, low or insufficient rating. Begin voting now.

I think we're still missing one response. Okay. Yes, we found it.

All right. Eight high, seven moderate, two low, and two insufficient.

CO-CHAIR BRISS: So, Bonnie, evidence?

DR. ZIMA: Okay. Evidence? Okay. I think this might also be a growth point

given that NQF has sort of raised the threshold for evidence on this one. And because it was a process measure under the new criteria, there was more emphasis on the developer to demonstrate quantity, quality and consistency. And it looks to me like -- at least of the literature presented it appears mostly on significance or impact.

The application does correctly note that this measure does assess an opportunity for treatment, but there was nothing in the application on the evidence between a relationship between adherence and desired outcome of improved treatment or diagnosis.

CO-CHAIR BRISS: Comments? Questions?

DR. SAMET: I guess I would invoke our thinking about the sort of science behind recommendations of prevention. And this one, from my read on it, comes even shorter way down the path than the previous one that we

approved by -- you know, kind of applying again the U.S. Preventative Services Task Force. Is this something that if not picked up -- the hardest one is usually if you don't screen for it and it eventually comes to be revealed, if not addressing it at a time when you could have detected, would it have made a difference? And I don't know, I mean, there's no -- they're not even close to having any data on that subject. So it makes me think this just may not be ready for prime time.

DR. SUSMAN: So I very much resonate with what you're saying, Jeff. And yet at the other hand, I would say from a general population perspective we know the general population has a huge performance gap in getting cardiovascular measures completed, that if you were to look at a composite of measures that everybody would agree should be indicated that are U.S. Preventive Services Task Force indicated, you know, it might be somewhere in the 10 to 15 percent range to get

all the things done that are supposed to be done.

One could logically, I think, take the leap of faith that this population would have even a harder time getting those things done on a routine basis. But I'll admit there's no evidence presented in this application that really speaks to the issue at hand, which if anything I wish that the sponsoring organization would just take a little bit more time to provide the information asked for because we're trying to then have to fill in this gap by our own anecdote rather than what I think clearly exists out there.

CO-CHAIR BRISS: Other questions or comments or concerns?

DR. SCHOLLE: I apologize for not being able to repeat the full review of evidence. And often what we find in behavioral health conditions is that the evidence we'd like to see doesn't exist. And

so we're creating evidence based on the recommendations from a general population and risk groups.

So the logic for this measure is that the U.S. Preventative Services Task Force strongly recommends screening men and women at different age groups for lipid disorders. And part of that, the recommendation is lower -- I mean, the evidence, the grade of the recommendation is lowered depending on the age group and then the task force brings in a recommendation based on risk for cardiovascular disease.

And so, I just want to make -- I don't know where -- so what we have is the task force recommendation that talks about risk status, right? And then we have guidelines for schizophrenia and a guideline for bipolar disorder that says that people with bipolar disorder and schizophrenia are at risk for hyperlipidemia. So what we've tried to do in our language is to say, well, we

don't have a lot of guidelines. We don't have a lot of systematic reviews that apply to this specific population. But, we have a general population recommendation that is grade A or grade B from the task force and we have recommendations or guidelines for these specific mental health conditions that say this group is at high risk. And so that's how we pulled together the recommendation, the evidence for this.

So what's the quality and the consistency? Well, the quality of the task force recommendation is excellent. What's the quality of the guideline recommendations within the conditions? It's pretty good, but there's not very many of them. It's not like there's a lot of research that supports it. So when we get to conditions like schizophrenia and bipolar disorder, we're really trying to weave together information from the general population focused on this group. And so, it does raise concerns about

-- we'd like to have more evidence before to support our measures. On the other hand we have a lot of evidence about the impact of this condition in this population.

So our expert groups and our stakeholder groups and our focus groups all weigh that together and say we know that people with schizophrenia and bipolar disorder have higher rates of cardiovascular disease. The test, while it does present some challenges, is actually -- that attention to it is a relatively inexpensive step to trying to identify people. And so, that's how they have to weigh the evidence against the impact and the feasibility.

CO-CHAIR BRISS: Any other questions, comments, concerns on the evidence?

I'm sorry, a couple. Caroline?

DR. CARNEY-DOEBBELING: If it's a general population measure, then I'm wondering why we're creating a specific measure for schizophrenia. And in NCQA's other book of

HEDIS measures for LDL they're linked to preexisting cardiovascular disease and to diabetes. They're not used ever as a stand alone LDL-C for the general population of any sort.

DR. SCHOLLE: And it has to do with identifying people at risk, because we don't have -- so in the claims data -- so the risk factor for -- in this case it's schizophrenia or bipolar disorder.

DR. CARNEY-DOEBBELING: Sure, I understand that, but then that would beg my original comment, which is why don't we have this for the general population of anyone in the general population who has a risk factor?

DR. SCHOLLE: Because we don't have the information about the risk factor. And here the risk factor is the diagnosis.

DR. CARNEY-DOEBBELING: But you would if you're driving it by claims. So if there are other risk factors that could be claims-driven, you might have it.

DR. SCHOLLE: Well, we've thought about it. We actually tested that, but if we were to go, the places we would go would be smokers and people with -- overweight or -- and those are not in claims data.

DR. SAMET: Just a brief comment. It almost seems like it's a National Quality Forum decision whether you want diagnosis-specific-type of preventive services as -- I mean, we can make our collective opinion, but it's almost an uber issue, it seems to me.

DR. BURSTIN: Our preference; and we just talked about this a little bit earlier on competing and harmonization, is we want measures applicable to the broadest possible populations. However, in this instance there actually is not an endorsed measure for the broadest possible population. It is a difficult measure to construct.

Curious to hear if Peter wants to talk about the need for a measure just like this for the Million Hearts Campaign in fact,

and there is still challenges with the evidence. So this has been an ongoing issue.

And if Bernadette even wants to speak to some of the ongoing issues around the cardiovascular screening for the general population for the task force.

But I just want to remind you that, you know, you do have the opportunity -- we have not invoked it yet, thankfully, as part of this process, but there is an exception, potential exception to the empirical body of evidence. And so, that is specifically if there's no empirical evidence, expert opinion is systematically assessed with agreements that the benefits to patients greatly outweigh potential harms.

So I would prefer that you actually vote on what you think the evidence is. And if you went to then invoke the exception, we can vote on the exception. I just want to at least put that forward as an option. We don't do it very often, but in

instances like this where there's probably not going to be a whole lot of specific empirical data on schizophrenics and LDL screening for schizophrenics, but at the same time there's good evidence of a high rate of cardiovascular disease unrecognized in schizophrenics, that would be exactly the kind of instance where we actually left this exception in place.

DR. SUSMAN: You know, I think what would be helpful in general; not just for this sponsor of a measure, but others, is to draw out the causal pathway in each step along the line more explicitly, not just through citation, but more explicitly put the evidence in your submission so that when we're looking at it, we have that. So in this case there is a high rate of risk factor prevalence and disease prevalence and premature outcomes that are bad, that everybody would agree about. There's a high outcome prevalence of cardiovascular disease in the general population. Then with a clear causal pathway

it really doesn't take a big leap of faith to say, well, yes, this makes sense.

I would make a distinction, as I think you did when you opened, that there's a difference between health disparity issues and those where there's something about the population that we're looking at that makes them at increased risk. And I think cardiovascular disease, there's something about this population in general that probably increases their risk whether it's due to their medications, their rates of obesity and so on, and smoking, you know, all the other stuff that we all know about. And for that reason I would be more supportive of looking at this as a specific population.

The other disparities issues, I would do I think what Peter had suggested, which is to have the measure cervical cancer screening in the sub-populations or stratification. Enough said.

CO-CHAIR BRISS: Yes, so I'm going

to take off my chair hat for a second. So on this one my evidentiary sense would have been that we know in the general population that cholesterol screening linked with treatment helps, right? You know, so we know that in this population that performance rates are at least or lower, lower that risk based on medications and other things is at least as high or higher. And we know that outcomes are at least as -- they're definitely worse, right? So on an evidentiary basis I'm not troubled by generalizing general population cholesterol screening works to this higher -- probably higher risk population.

Now I am a little troubled by this measure. The claims-based nature of this measure is sort of -- this one has a drunk at the lamppost problem to me. We're looking at the third or fourth most important driver of this population's bad cardiovascular outcomes because that's what we can find in the claims data, right? And so, I'd much rather

eventually get to a more holistic, more relevant measure that actually gets at the bigger drivers, even if it's harder to get.

DR. PATING: Okay. Yes, I actually would agree with that. What am I trying to say? Just the causal connection between getting this lab value and having, what I think is what you're pointing to, a larger review, cardiovascular review of risk factors presumably by an internist or somebody with, you know, more -- because I don't know if a psychiatrist would have the whole -- the skill to go through the whole cardiovascular risk set. The causal links are just not there. And so, it's sort of a tail wag the dog. If this tail was connected to the dog, I would say okay, but I'm just -- the evidence is not there that there's a head on the other end of the dog. You can quote me on that one, yes.

DR. MELYNK: In the U.S. Task Force we always use these analytic frameworks

and really looked at the supporting evidence all along the pathway. And I think it's super important here to look at what are the greatest predictors of heart disease in these patients. And I'm not sure just looking at an LDL is going to give us what we really need. So that's my real concern regarding this particular measure.

DR. HANRAHAN: Just briefly, we're dealing with administrative data here, which is the best we can do. You cannot get from administrative data any kind of cardiac risk profile. It's just not possible.

If we could, the American Heart Association has a wonderful risk profile that we could utilize. And I am really worried that because of that we won't move forward with this population that has such high risk and is known to be so vulnerable to these kinds of conditions. And again, it's a state-level indicator. It's not at the individual level or the provider level. It's at the

state that's going to use that to either promote their programs, evaluate their programs based on that.

DR. MARK: Yes, I mean, I understand the point that this is maybe not the best thing to be measuring, but I think if you think about it as a proxy for now they're getting some kind of medical care, someone's paying attention to their whole health as opposed to maybe just giving them medication in a, you know, psychiatric setting, then maybe it's a useful measure even though it's maybe not the best measure.

CO-CHAIR PINCUS: You know, basically I agree, Peter with your judgment about the evidence assessment, but I also would just combine Nancy and Tami's point that -- because I think there are indirect benefits of doing this that are likely to accrue in terms of the greater degree of connection between general medical care and psychiatric care that again are going to be -- are not

directly related to treatment of cardiovascular risk, but that by having this as a measure it will stimulate that, and that it will stimulate states to actually do that now whether that falls into, you know, an exception category or whether that falls into sort of evidence.

But I also feel -- you know, my own view about the evidence is that, you know, that this is a very high-risk category. There are people that -- you know, I would put the evidence from my view that the evidence sort of just makes the threshold, independent of this sort of indirect theoretical notion.

CO-CHAIR BRISS: So I'm going to keep going around the table and call on myself.

So the other question that I wanted to ask about this is in the general population the people for whom the evidence is best that cholesterol screening and treatment helps I think is in people who have already

established vascular disease. So the other evidentiary question is why do we exclude the people from the denominator for whom the evidence of benefit might be the best?

DR. SCHOLLE: That's the monitoring measure. So they're there, they're just not in this measure, because we separated out screening from monitoring.

CO-CHAIR BRISS: Right.

DR. SCHOLLE: And remember, for this measure this is screening for people with schizophrenia or bipolar who also have an antipsychotic medication. So again, in this measure we tied it to the antipsychotic medication. It's the same denominator as the diabetes measure.

CO-CHAIR BRISS: But you can't be monitored if you're not screened, right? And it's routine -- so I've been suffering with -- can you tell I've been working on cardiovascular stuff a lot lately? And so, even among highest risk people in the states,

if you look at national data on screening and treatment, especially for cholesterol, it's just ghastly. I mean, even with people in the highest risk. So in this population I'm not at all convinced that people would get -- in the general population I can sometimes make a straight-faced argument that, look, everybody in America gets screened with an LDL. I'm not so sure in this population that I quite believe it. So are we sure that excluding the highest risk people from the denominator of this measure actually makes sense?

DR. SCHOLLE: It really comes down to how do we define -- I mean, where do you want to split the measures? Okay? So what we chose to do is to take everybody -- so all those exclusions in this measure are the people who are in the cardiovascular monitoring measure. So if you put those two measures together, you have almost everybody with schizophrenia. You're leaving out the 5 to 10 percent of people with schizophrenia who

don't get an antipsychotic. They're the people that, because they are truly not in this denominator, or of either of the measures that -- as we were doing it, we were trying to parallel our measures where the existing measures for monitoring of diabetes and heart disease where they exist. Right? So those two measures that we're going to look at that are about monitoring, those use the same denominator criteria as the existing HEDIS measures and they add diabetes or schizophrenia. So then, it becomes, you know, that sub-population under the bigger population. So we chose to do it that way. And then we chose to create a screening measure because it's screening if you don't have the diagnosis yet. So that's the way that we split it up that way. If it makes more sense to do it differently, then I think we'd be interested to hear what you're trying to work with -

CO-CHAIR BRISS: So I just want to

make sure that I understand. So a person with a diagnosis of coronary disease who wasn't appropriated screened is the person that I'm worried about. We've excluded that person from the screening measure. I don't see how you would get into a monitoring measure if you weren't screened. So that person gets lost, right?

DR. SCHOLLE: So let me just be clear. In our terminology, so screening is you don't have the diagnosis yet or at least you don't have it within the data that we're looking at. We can't find the diagnosis. Monitoring could be the same kind of secondary screening test, but for people who have the diagnosis. So that's how we defined it. So you call it screening, but we call it monitoring. Okay.

MR. WILLIAMSON: We will now vote on the evidence. This is a yes, no or insufficient rating. You may begin now.

We have eight yes, three no, and

seven insufficient. The measure fails on evidence.

DR. BURSTIN: And at this point the question is does anybody want to invoke the exception? Since there's been enough discussion, I think it's worth talking about and voting on.

DR. SAMET: I move that we should have a discussion about the exception for the -- well, I mean, I think you and Harold actually made the case in your previous -- we might invoke. So I don't really have much else to add. I think it is a unique population. I think the data is not there. That's why it didn't get across. But there's lots of risks, there's lots of reasons to think it might be helpful. And I was really moved by the comments as well about as a proxy almost for being in care and there's such a difficulty getting this population into the general medical care setting. Those are the things that move me.

DR. PATING: Yes, my interest would though be to ask the submitters to come back with just a more explicit logic model. I mean, just because it could be a valid indicator. It just may not be ready for NQF consideration right now. We could look at it again when the logic model -- because I feel like we're thinking here as a group, and I think this should have been done in a committee, just to be a little more explicit. It could even be just showing us a chart of how these different things connect.

And then also I'd be wondering, you know, if we're looking at really primary care screening, why can't we just go and measure whether they saw their primary care doctor and code it as a CPT visit in the last year and the standard of care would be hard and long at least minimum kind of screening in that visit. You know, I don't know what the other options could be. I just would like it kind of fleshed out because I feel like we're

thinking the work here in the group.

DR. SAMET: So, David, I was with you up until you said primary care. I mean, the reality is there are good studies looking at, one, utilization of primary care services; and then do primary care clinicians actually for this population provide the services that are indicated by well-recognized bodies. And the answer sadly is not. I think there's a lot of things that take us; and I count myself as a primary care clinician still, off of doing the things that we probably would do routinely because we're overwhelmed by all the other stuff that's going on.

So just getting to the issue at hand of an exception, I'm not in favor of an exception, even though I voted yes to see this go forward on the basis of albeit less-explicit evidence. I think we have a process which is rigorous and I'd hope that we could really adhere to that rather than trying to exception this out.

CO-CHAIR BRISS: Yes, I'm --

DR. BURSTIN: I'm sorry, just a point of clarification. This is actually a part of our process. So I don't want to make it seem like we're doing this on the fly. We actually had an evidence task force who, you know, went through this exhaustively and really put this in as an exception in areas I think somewhat like this. So I don't want you to feel like this is something new and different. It is in fact part of our process intentionally to allow expert opinion and expert consensus opinion to in fact move areas where we think risks are -- you know, that benefits significantly exceed risks.

CO-CHAIR BRISS: So --

DR. BURSTIN: And just; I'm sorry, one more process point to David. You can only evaluate the measure before you. So we can't be talking about other potential measures that could come forward. They're not before you today. So it's really just about this one.

DR. PATING: No, I was just --

DR. BURSTIN: Yes.

DR. PATING: -- thinking of the goal --

DR. BURSTIN: Yes.

DR. PATING: -- you know, just to flesh out the logic model.

CO-CHAIR BRISS: So I'm going to take off my chair hat for a second. But in terms of invoking the exception, it's hard to not creep into what else might we do. So truth is I'm wearing one of my hats, as I'm still a safety net internist, right? And so, as a safety net internist I'm very skeptical that just the act of getting an LDL is likely in the kind of population that's hard to reach, right, and for whom it's hard to coordinate care. I'm very skeptical on its face that the act of getting an LDL is going to provoke the kind of coordination of care for a more holistic set of cardiovascular indicators, that it's really going to get us

to where we want to be.

And so, I'm sensitive to not wanting to let the perfect be the enemy of the good, but I'm not sure that this is quite up to the level of good to me yet and I'd really like somebody to think about what's the measure or set of measures that would provoke the better coordination that this population really needs and deserves.

So, now soap box over and back to chair role. So, I'm just going to keep going around the table.

DR. KELLEHER: I just wanted to comment in terms of looking at an exception that I get concerned that in this very imperfect world of creating measures that if we are over zealous about insisting that there be a body of evidence before there's been a chance to develop a body of evidence, that we will never go forward. And for me, I think we're there.

DR. SUSMAN: I'm trying to figure

out -- I could take your comment either way, that you --

DR. KELLEHER: Oh, all right. Well --

DR. SUSMAN: -- are in favor of the exception or not and I'm just trying --

DR. KELLEHER: I am in favor of an exception. And I think Harold said it best, so maybe he'll repeat it on his turn.

DR. MARK: Yes, I'm in favor of an exception, too. And not knowing enough about the cardiovascular screening recommendation, I'm trying to understand why others aren't. So if the recommendation is that basically everybody get these glucose tests at some point and then within -- I'm sorry cholesterol tests at some point, and then the question is whether this -- so there's a big circle and the question's whether in that circle this little sub-circle should also get that. It just seems obvious that you're just saying, well, some little population within everybody

should -- we need to do a special test on them. So maybe I'm missing something here.

And then also the other question is what kind of evidence would need to see? I mean, what would the study look like to be convincing?

DR. SUSMAN: Well, I think there are a lot of barriers in the implementation that go beyond just getting the test. I mean, let's just follow it through. So you get the test. Then you have to have the patient back in. You have to ascertain whether the individual meets criteria for medication. You prescribe the medicine. Has to get filled. There has to be persistence of medicine. So there's that whole set of issues.

And if we're using this as a proxy, is this a reasonable proxy for cardiovascular care? And, you know, on an evidence basis, although I can see both sides of this; and I'm arguing in my own mind with that, I'm not so sure, given the difficulty of

assuring care and actually getting to an outcome of improved cardiovascular health, that this is the way to measure it or the way to help us along that pathway. At least I think it's reasonable to have concerns about that one way or the other.

DR. MARK: So what would the evidence look like if you wanted to support that case?

DR. SUSMAN: I mean, for me it would be looking at a whole series of measures which are much more robust that would require data abstraction, and I would just push on NCQA and HEDIS and our other measure developers to say, you know, look, it's been too long that we've accepted this dictum, well, it's administrative data. That's all we can do. Sorry. Here's the imperfect -- you know, make do with it.

I think it's time to really take a stand and say, come on, this is what we need to do. We need to hold ourselves to a more

high standard than we've been willing to in the past. End of soap box.

CO-CHAIR BRISS: Jeff. It's easier to just go around the table.

DR. SAMET: So it's sort of point/counterpoint with Jeffrey.

Well, to say that we have to hold NCQA to higher standard but we don't have to hold medical care to a higher standard seems -- you know, it's too hard to have them get in the system and follow on through, there's something that doesn't work there for me.

Well, so I think this is a difficult one and I think it's right on the edge, to be truthful. But I almost think it's worth trying for a couple years. Not that they shouldn't push forward trying to do what they're trying to do, but we're hearing repeatedly that nothing's happening in this -- you know, and I know it's the case we never get them into primary care because they never show up. So it's like they're not even

present. And if this could be one piece that's an exception that moves forward for a few years and tries to advance the field a bit, I'm okay with it as an exception.

DR. CARNEY-DOEBBELING: I'm still not clear if we're voting on the exception or having the discussion beyond what would be if the exception was voted on in one way or another or still --

CO-CHAIR BRISS: We're trying to get us to voting on the exception.

DR. CARNEY-DOEBBELING: Okay. So in that case I would say if part of the reason for the exception is to use the LDL-C as a measure of a proxy for primary care, where is the evidence for that even in the general population, let alone this population.

And by way of background, I'm an internist psychiatrist. I ran a med psych unit. I did all kinds of stuff. I wrote a bunch of papers, some of which I'm sure Harold reviewed along the way, about the comorbidity

and the higher death rates and all of those kinds of things. I know that this is the right thing to do.

But I also know from wearing my other hat of being part of a Medicaid office and a health plan office that when you start putting measures that don't clearly make good sense for changing an outcome into the burden of what a provider has to do all day every day, you miss the point of moving things forward because you make them angry.

And if what we really want is integrated care, let's come up with a measure of you have to code obesity and overweight, you have to code nicotine dependence because then you start getting to those risk factors, and/or you start working on another level to change policy to actually pay for integrated care state by state by state. Where is the evidence that adding one more measure is a proxy for what really needs to get done?

CO-CHAIR PINCUS: So I mean, I

respectfully disagree, you know, not wearing my chair hat, because I actually think the logic has been displayed considerably and I agree with Jeff, the second Jeff, in the sense that, you know, this does I think, you know, cross the threshold, but barely, and justifies the exception. And that's what the exception was made for, because number one, especially in this situation, because we're dealing with the measure at hand, the measure at hand is a state measure and the intent is to stimulate actions by the state to improve systems.

And so, given that consideration, you know, obviously that goes down to the practitioner level ultimately, but it really will -- you know, the intent to bridge those systems, and I think ultimately that's what you want to do. And it follows, at least for me, the notion that this clearly -- you know, sort of the circle within a circle notion that these clearly are individuals who are at higher risk and that a screening test is a

necessary but insufficient part of the process of improving their care.

CO-CHAIR BRISS: So now I'm definitely wearing my chair hat. So I think most of the arguments that can be made have been made. I think that I'm now hearing the same arguments and I'm agreeing with all of them. And a foolish consistency is the hobgoblin of little minds, right? And so, if anybody else has anything that hasn't already been said -- you said there were a couple of people down on this end that were trying --

Okay. So I think people know how they feel. So yes is an exception and no is --

MR. WILLIAMSON: I'll read it aloud. "If there is no empirical evidence; for example, only expert opinion, and expert opinion was systematically assessed with agreement that the benefits of the measured process or structure to patients greatly outweigh potential harms, is there an

exceptional and compelling reason that the measure should be considered further?"

So we'll be voting yes or no. One is yes, two is no. And you may begin voting now.

I think we're still waiting on one -- or is it down to 18 now?

Okay. Actually we're good.

So the measure passes the exception. We have 10 yes, and 8 no.

We will now move on to reliability.

DR. ZIMA: Okay. Reliability. The measure specs were I thought very clearly written, the reliability testing. What they did is they used data from 16 states of the 22 states. And in our work group call we asked, you know, why were the states not there. It was because the sample size was too small in the denominator. So the data's limited to 22 states that had completed data.

Just a quick question to the

developer, and that was was the 22 also the same as the number of eligible states, or were there some eligible states in this reliability testing that did not have complete data?

DR. SCHOLLE: There were some states that did not have complete data. Let me just check.

DR. ZIMA: So we had --

DR. SCHOLLE: I built a pretty chart. Three states did not have complete data to allow for the reliability test.

DR. ZIMA: Okay. So we had three plus six, so we had nine states that did not have complete data in the total sample? Is that how that works? I only bring that up as a usability issue in thinking about states being able to use this.

I think reliability was based on stability of performance at the state level and slightly more than half, 56 percent, of the states, nine over six, found no change between the two years. Correlation was

moderate at 0.43.

CO-CHAIR BRISS: So perhaps it looks to me like we've passed two measures today that performed less well on these kind of measures than this, right? So does anybody want to make the case that this one isn't as good as the two that we've already passed?

(No audible response.)

CO-CHAIR BRISS: So let's try going straight to voting.

MR. WILLIAMSON: We will now vote on reliability. This is a high, moderate, low or insufficient rating. You may begin voting now.

We have 0 high, 14 moderate, 3 low, and 0 insufficient.

CO-CHAIR BRISS: So moving onto validity, please?

DR. ZIMA: Okay. Validity was also based only on face validity by a multi-stakeholder technical advisory group plus public comments plus focus groups from several

organizations including the Medicaid Medical Directors Learning Network, managed behavioral health organizations, state mental health commissioners and medical directors. Concurrent validity was based on correlation with other quality indicators related to screening, which was found to be high, and use of hospitalization ED use for schizophrenia. And the developer argued that there's a negative relationship between the screening. And there was an assumption that hospitalization ED use for schizophrenia may be an adverse event.

I found it sometimes a little bit of a stretch how it supported the validity of LDL screening, improving diagnosis and treatment. And again, I think we kind of butt our heads up against some of the limitations of using claims data. Potential threats of validity were not examined.

CO-CHAIR BRISS: Comments before we vote?

(No audible response.)

CO-CHAIR BRISS: Let's try voting, please.

MR. WILLIAMSON: We will now vote on validity. This is a high, moderate, low or insufficient rating. Begin voting now.

We have 0 high, 13 moderate, 3 low, and 2 insufficient.

CO-CHAIR BRISS: We're inevitably so much faster when we get exhausted.

Usability?

DR. ZIMA: Okay. On usability we had some discussion in our work group about that there might be higher use of ED and hospital in some states because there's more specialty mental health services, better access to care for persons in crisis. Low rates of use in some states could also mean that there was a shift in mental health services for SMI population to other sectors like jails and prisons. And so this also raised the question of whether the findings

and adherence to this measure were relative easy to determine and were meaningful.

As far as feasibility, again 72 percent of 22 states had complete data using this claims data.

CO-CHAIR BRISS: Questions? Comments? Concerns?

(No audible response.)

CO-CHAIR BRISS: Let's move to voting. Sorry. I'm sorry.

DR. SAMET: I'm just ignorant. So that last number you gave, the 72 percent had data that could be used? Is that what you said? Is that good? I mean, I don't deal with this claim stuff. Or is that not good?

DR. ZIMA: I don't know. This is only again sort of summarizing what was in the application such that it was 16 out of 22 states had complete data using the MAX claims data. So that's 72 percent. And sample sizes were not presented for the denominators for the 22 states.

DR. SAMET: So I'm just asking someone who knows claims data. I mean, I don't know whether to consider that as good usability or --

DR. CARNEY-DOEBBELING: If the measure is intended to be used at the state level to compare states, then I would suggest that that's moderate, at best moderate.

DR. SCHOLLE: So we're using the fee-for-service claims data extract. And so the states where we could not do that, there are two limitations about how this could be used at the state level that would make it more -- would provide larger robust samples or denominators for the states. One is the state could use it for both their fee-for-service and their managed care population. We couldn't test that. They could use it for people who have dual-eligibility by looking at the data that are in Medicare and Medicaid.

I understand that all those things are hard for states to do, but that's what

states are doing to report the children's core measures and what they're doing to try to monitor their dual-eligible population. The idea is to present a measure that is relevant to this population in that category. We do know that we lost a lot of people dropping out the dual-eligibles who would have been eligible for this measure. So we really think that -- and states told us that they would apply it to their dual-eligible population.

CO-CHAIR BRISS: Caroline?

DR. CARNEY-DOEBBELING: Point/counter-point. Behavioral health is especially different on the managed care level. So I don't think you can also make a leap to say that states will easily produce this data from managed care because there are behavioral health carve-outs that may prevent even the ease of getting that data in in the first place and it's not been tested yet. You've not tried to collect that data.

There are already issues

collecting the data when they are available, let alone in a managed care environment where most of the ABDs and duals will be moving as we read the policies coming out of the Government. I don't think we can comfortably say that this is easy to get done in the real world and that states can easily get it done.

CO-CHAIR BRISS: Jeff, are you trying to get back in?

(No audible response.)

CO-CHAIR BRISS: So anybody else, comments before we vote?

(No audible response.)

CO-CHAIR BRISS: Let's try voting usability, please.

MR. WILLIAMSON: We will now vote on the usability. This is a high, moderate, low or insufficient rating. You may begin voting now.

We have 1 high, 12 moderate, 5 low, and 0 insufficient.

CO-CHAIR BRISS: Feasibility,

please? Bonnie, comments on feasibility?

DR. ZIMA: No, I think I actually lumped usability and feasibility together given that they're so intertwined on the state level data.

CO-CHAIR BRISS: Okay. So anybody with additional comments that haven't been made?

(No audible response.)

CO-CHAIR BRISS: Hearing none -- ah, yes?

DR. PINDOLIA: So to what Caroline had said earlier, I think we're going to have even a bigger compounded effect of not only having fasting data. There are people coming back with fasting labs to get their full scope and then have -- or possibly have just LDL without having a fasting level separate because more and more health plans are starting to charge co-pays for labs. So now they would be possibly charged with two co-pays and they're probably even less likely to

come back.

CO-CHAIR BRISS: Any other questions, comments, concerns?

DR. SCHOLLE: I don't think that Medicaid charges a co-pay.

DR. CARNEY-DOEBBELING: Oh, no, and some managed -- in Medicaid expansion there can be co-pays charged. And there are co-pays charged for pharmacy across Medicaid, too. It's different state by state by state.

DR. SCHOLLE: But I'm not sure for the testing piece, though. I would say that, you know, among the measures in the public comment -- you know, so to determine whether this is feasible and usable by state Medicaid programs and by health plans, we rely heavily on the public comment that we get, and public comment for this measure, like the previous measure, was very positive. So if there were challenges in doing this measure, we would have heard about it.

CO-CHAIR BRISS: So let's try to

go to a vote, please, on feasibility.

MR. WILLIAMSON: We'll now vote on feasibility. This is a high, moderate, low or insufficient rating. May begin voting now.

We have 12 moderate and 6 low.

CO-CHAIR BRISS: So that takes us to the overall suitability. One is yes and two is no.

MR. WILLIAMSON: Final comments?

Oh, I'm sorry. Final comments?

(No audible response.)

CO-CHAIR BRISS: So one is yes and two is now.

MR. WILLIAMSON: Okay. We'll now be voting on the overall suitability for endorsement. You may begin voting now.

The measure passes 10 to 8.

CO-CHAIR BRISS: So the next one is 1933.

DR. PATING: Dr. Briss, could I just ask that our comments regarding potential looking at the validity of the care path be

passed on and --

CO-CHAIR BRISS: Yes.

DR. PATING: I don't know how -- what way that would be done, but --

DR. BURSTIN: They're sitting behind you, first of all.

DR. PATING: Yes. No. So --

DR. BURSTIN: And they'll be responding --

CO-CHAIR BRISS: They've been listening? You have your backs to them and they've been listening very carefully and taking notes.

DR. BURSTIN: They're all lined up behind you.

And actually it's probably just helpful here to remember where we are in the entire consensus process. I mean, all that will happen at this point is we will draft a report with your preliminary recommendations, the commentary, etcetera. It will then go out for public comment. They'll then have another

chance, and the developers will as well, to respond to public comment. So I suspect we'll get a fair amount of public comment on this measure. So we'll be revisiting it soon, which is good.

CO-CHAIR BRISS: So the next measure is 1933. Any opening comments from the developer that we haven't heard already?

We're doing some reordering so that we can capture discussants before they leave. So any comments from the developer?

DR. SCHOLLE: Just to orient ourselves, so 1933 is the cardiovascular health monitoring measure. So this is the measure that takes the existing HEDIS measure for looking to see whether people who have established cardiovascular disease, whether they have at least one cholesterol test during the year. In the HEDIS health plan measures this is paired with a measure that looks at control. And so, we did not propose that control measure because it requires chart

review. And in developing these measures we were asked to focus only on claims data because of the expectation that those would be more easily used by state Medicaid programs. But I believe that the measure is exactly as the HEDIS measure except for the denominator definition of schizophrenia.

DR. BURSTIN: And the age cutoff is different, is that right?

DR. SCHOLLE: The age cut-off --

DR. BURSTIN: Again, the 25 cut-off.

DR. SCHOLLE: And so, our data, in HEDIS, the HEDIS rate is 26 percentage points higher than the rate that we found for the schizophrenia population in our field test, so the HEDIS Medicaid rate.

CO-CHAIR BRISS: So I'm sorry, you've said a couple of times today that you were asked to -- and I may have been distracted while I was trying to slavishly pay attention to what staff was telling me to do.

So when you say you were asked to focus on claims data for several of these measures, you were asked by whom?

DR. SCHOLLE: We were directed by our funder, ASPE, but focus on measures that could come from claims data in order to make the measures feasible for states to report from claims. And that's consistent with our experience of working with states on the children's core set. Where the measures have required doing chart review, states are having a hard time if they're not geared up for that already or can pass it on through their contractors.

DR. BURSTIN: This is I think an interesting issue, this one and the diabetes one to follow. The diabetes one to follow is a little bit different because it combines two existing endorsed measures of A1c and LDL for diabetics. But I almost wonder if because this is truly essentially the identical measure to what is already endorsed for the

general population; going back to I think Tami's point earlier, this is what the universe is and this is the schizophrenics in the middle.

If there's no reason that anybody on this Committee thinks that there is a different evidentiary base, reliability, etcetera for this measure, it's not clear to me that it needs to be -- I'm not even clear this needs to be a separate measure, to be honest. I still think this actually would be a very nice strata within the existing HEDIS measure, and I just think it's something we could talk about. The only difference truly is the age cutoff of 25 versus I believe -- what was the other one?

DR. SCHOLLE: So this measure is proposed 25 to -- that's not what the denominator says, but just to be clear, 25 to 64, and the measure in HEDIS is 18 to 75. And remember, that's because we're doing that now. In the process of aligning with our CMS

colleagues, we are talking about lining that up as well. So your point is well taken, Helen.

CO-CHAIR PINCUS: That's the definition of the stratum if you were doing it the way you said, which would make sense if that's the intention.

CO-CHAIR BRISS: And in addition to what Helen said, it strikes me that -- so if this is essentially paired with the last one that we discussed in some sense, that we should have dealt with a lot of the details.

So, Helen, are you suggesting that we don't review this measure, or we quickly review this measure, or do you have a recommendation for us?

DR. BURSTIN: This is really something I think we probably need to talk offline with NCQA. I think my recommendation would be if it's truly the identical measure in every other way that's already been endorsed, I don't see any reason why we need

to review the details of this measure in terms of the evaluation of the criteria. And maybe we just jump to a discussion perhaps of is this something better done as a stand alone? Is there any reason why the identical measure with the identical information should be a stand alone, or is this something better served as a strata within the current endorsed measure?

DR. CARNEY-DOEBBELING: Is it a hybrid measure?

DR. BURSTIN: This one's not.

DR. SCHOLLE: I actually believe that the HEDIS measure is a hybrid measure.

DR. CARNEY-DOEBBELING: Yes. So is the strata problem you wouldn't get necessarily enough schizophrenics to report them out separately if you relied on pulling them out of the general population because it's a hybrid measure?

DR. SCHOLLE: It's in the stratified measure to report it for

schizophrenia. And the other question comes is do you want this identified as a schizophrenia -- a measure that you should use and report specifically for people with schizophrenia. So I wonder -- and to have it paired with the previous measure.

So, you know, as we went through this, we were looking at it saying let's build a suite of measures for people with schizophrenia. So we looked at existing measures and new measure concepts for this population. So it would be good to know whether this Committee would recommend this as one that we should consider for people with schizophrenia. If the evidence and feasibility and reporting and all that is essentially the same, then maybe the question to ask the Committee is does it make sense to have a strata for people with schizophrenia and then we work with NQF to figure out, well, how do we represent that in the list of endorsed measures?

CO-CHAIR BRISS: So it looks like there are several people with cards up. Let's try to spend five minutes at the beginning trying to answer that, the sort of contextual question and then we'll make a decision based on that about whether we want to go through the rest of the drill.

So, Jeff, do you have comments about that?

DR. SUSMAN: Yes, I basically feel like we should be headed more towards having the overall measure and then doing the stratification by whatever characteristic we want. So I would leave that to the NQF staff to figure out with NCQA how that's done and you get the right n and all that. But I think that makes more sense to me. Same with cervical cancer, wherever else you happen to go with this.

DR. CHALK: Given that there will come a point relatively soon where we will want the population that's addicted to opioid

-- that is opioid dependent or alcohol dependent stratified within an overall measure like we're talking about, I would support what Jeffrey just said. That's my leaning.

CO-CHAIR BRISS: I'm not going to make a comment as a Committee member and not as a chair. I also favor movement toward a stratified measure. This is based on general cardiovascular stuff. Again, I like the upper age range of the general measure better than this measure, again because there's a whole lot of cardiovascular morbidity and mortality in that older age band and there's no good reason from a cardiovascular health standpoint to exclude the highest risk people.

DR. KELLEHER: This may be sort of an aside comment, but I was wondering in your -- I know this stands alone, but in your screening measures you included a population on antipsychotics that included diagnoses of schizophrenia and bipolar. And yet in your monitoring measures -- I'd just like to know

why you didn't have more sort of synchrony with that population?

DR. SCHOLLE: It's really a matter of time and resources. We started this work focused on schizophrenia. We were able to recover and to bring the bipolar group into the antipsychotic measure, but we did not have time to go back and redo the entire evidence, all the work that preceded that for the bipolar population. And so, we only got as far as doing that for the screening measure.

CO-CHAIR PINCUS: I agree with the sentiment about making it a segmentation, but I would also urge NQF to actually formalize this issue of segmentation so that when the measures are published as being endorsed there's somewhere where it kind of gives more specificity that it has been recommended or that it can -- you know, somehow that people see this as a potential subsidiary measure.

CO-CHAIR BRISS: So what should we do at this point, Helen, about this measure?

DR. SCHOLLE: To be clear, we have to get specs for that denominator. If we're going to recommend it, we need specs that go along with it. So it needs -- it's not just -- yes, we'll talk about how to do that.

DR. BURSTIN: Right. I mean, I think it's fine if the Committee wants to just, you know, for the sake of completeness just very quickly run through their criteria here, knowing the evidence is in fact the evidence to the entire population, measure in use for many years, good -- you know, and I think this could just be rather rapid. If that would help us, you know, figure out next steps and have it blessed, I'm fine with that. But it should be a pretty quick discussion.

CO-CHAIR BRISS: So given that, let me suggest that since we've just done a measure that's going to be very like this, let's try to run through the drill on this measure and try to not re-litigate things that we just did with the last measure. Okay?

So with that, Bonnie, are you leading this one, too?

DR. ZIMA: I just also want to share a little side bar that I had with Sarah earlier and that was that the sample size in this data is really schizophrenia and diagnosis of cardiovascular disease, not just schizophrenia. And I want to give them a chance to make that correction.

CO-CHAIR BRISS: Maybe I didn't need to say anything about impact that hasn't already been said. Let's vote, please.

MR. WILLIAMSON: We will now vote on impact. May begin voting now.

For impact we have 10 high, 7 moderate, 1 low, and 0 insufficient.

CO-CHAIR BRISS: Does anybody need to say anything about minding the gap that hasn't already been said yet?

(No audible response.)

CO-CHAIR BRISS: Let's vote, please.

MR. WILLIAMSON: We will now vote on the performance gap. You may begin voting now.

We have 6 high, 11 moderate, 0 low, and 1 insufficient.

CO-CHAIR BRISS: I'm hoping we can get away without a long discussion of evidence. Bonnie, would you like --

DR. ZIMA: Evidence. Ditto.

CO-CHAIR BRISS: Anybody have anything to say that hasn't been already said?

David?

DR. PATING: I just want to clarify. Can you just, Bonnie, maybe explain how this indicator is different than the 1927? The 1927 you're screening just once, but there's an annual screening or -- and this one's a maintenance?

DR. ZIMA: Different populations.

DR. PATING: Different populations?

CO-CHAIR BRISS: Essentially with

existing cardiovascular disease.

DR. PATING: Okay.

DR. ZIMA: Denominator.

CO-CHAIR BRISS: And truth is, from my perspective that might make the evidence a little better.

So with those caveats, anybody else have anything new to say about evidence?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote on the evidence. Reminder, this is a yes, no or insufficient vote. You may begin now.

We have 15 yes, 1 no, and 2 insufficient.

CO-CHAIR BRISS: Bonnie, is there anything new in reliability of the measure?

DR. ZIMA: Nothing new on the issues. The findings are a little bit different. Less than one third, 31 percent, of the states by a vote of 16 found no change.

CO-CHAIR BRISS: Will you scroll to the results, please?

DR. ZIMA: I guess the other way you could say it is that almost 70 percent of the states had some type of change in one direction and extent of change not defined.

CO-CHAIR BRISS: And again, we've seen this kind of pattern of data now at least three times today. So anybody need to say anything else before we vote?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote on the reliability. This is a high, moderate, low or insufficient rating. You may begin now.

We have 1 high, 13 moderate, 4 low, and 0 insufficient.

CO-CHAIR BRISS: Validity? Anything new?

DR. ZIMA: Not really.

Correlation remains high with monitoring.

CO-CHAIR BRISS: Comments?

(No audible response.)

CO-CHAIR BRISS: Let's vote, please.

MR. WILLIAMSON: We will now vote on the validity. This is a high, moderate, low or insufficient rating. You may begin now.

We have 2 high, 15 moderate, 1 low, and 0 insufficient.

CO-CHAIR BRISS: Usability? Anything new?

DR. ZIMA: No. Similar concerns on usability and feasibility.

CO-CHAIR BRISS: Okay. So anybody want to say anything else before we vote?

(No audible response.)

CO-CHAIR BRISS: Let's vote, please.

MR. WILLIAMSON: We will now vote on the usability. Please begin now.

CO-CHAIR BRISS: Helen, is there a "Guinness Book of Record" for rapidity of approval from a standing start?

MR. WILLIAMSON: We now have 2 high, 12 moderate, 4 low, and 0 insufficient.

CO-CHAIR BRISS: And feasibility, anymore comments before we vote?

(No audible response.)

CO-CHAIR BRISS: Let's vote, please.

MR. WILLIAMSON: We will now vote on the feasibility. Please begin now.

We have 1 high, 12 moderate, 5 low, and 0 insufficient.

CO-CHAIR BRISS: And overall approval. Anybody have closing comments?

DR. PINDOLIA: I had one question for clarification. So based on the discussions that we had and what NQF is going to go talk to NCQA, if this is endorsed, it's endorsed with that conversation or it's endorsed as a stand alone on its own --

DR. BURSTIN: Yes, basically all you're doing right now is saying it's suitable for endorsement and we'll work out the details --

DR. PINDOLIA: Okay.

DR. BURSTIN: -- of whether it's in fact we think a subsidiary measure under the existing measure. But we at least want to have it blessed that you think it's suitable, it meets the criteria.

CO-CHAIR BRISS: Anybody else, questions or comments?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote on the overall suitability for endorsement. This is a yes or no rating. You may begin now.

We have 16 yes and 2 no.

CO-CHAIR BRISS: And you can make up a lot of time if you do endorsements like

that.

So can we go to No. 1934?

MS. FRANKLIN: So did the developer want to tee up this one, 1934?

DR. BURSTIN: It's interesting, this is not exactly the same only because it's actually combining two existing HEDIS measures into one in this instance. Correct? So the question would be, Sarah, et al, would this be acceptable as potentially strata under each of those measures? Okay.

DR. SCHOLLE: And just to clarify, the importance of the rate that we found was -- what is it? Where's -- which is the HEDIS rate? The HEDIS rates are in the range of 70 or 80 percent, and the range for this was 50 percent.

PARTICIPANT: So significantly lower.

DR. SCHOLLE: So, yes, we would be comfortable putting these as rates under each of those individual HEDIS measures for

diabetes.

CO-CHAIR BRISS: I think we will likely have surfaced most of the issues that can be surfaced. Let's try to do another abbreviated process with this one and see how it goes.

DR. BURSTIN: And we'll do the same thing at the end of this. If you guys deem this measure as suitable for endorsement, we'll work with NCQA to come up with subsidiary measures under the diabetes LDL and the diabetes A1c, although certainly one might think that maybe it's time to put them together for everything.

CO-CHAIR BRISS: So, Lisa, you want to tee us up, please?

DR. SHEA: Sure. Well, this as we said looks at individuals who have diabetes and schizophrenia and wants to make sure that they have at least one hemoglobin A1c and one LDL-C done during the year.

The evidence in terms of the

impact is high, as we've heard before. In addition, there are studies cited that shows that about a third of people who have both conditions do not receive the treatment.

CO-CHAIR BRISS: So anybody need to make additional comments that haven't been made about impact?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote on impact. This is a high, moderate, low or insufficient rating. And you may begin now.

We're missing two responses. Yes. Oh, okay. So we're missing one response now. There we go. Yes, we just got it. Yes.

All right. We have 10 high and 6 moderate.

CO-CHAIR BRISS: So minding the gap?

DR. SHEA: So as we heard from the developer, there is a gap. They do provide

data from this database that we've talked about that shows in general half the folks are getting this and that there were disparities in terms of the African-American population.

CO-CHAIR BRISS: Questions? Comments? Concerns?

(No audible response.)

CO-CHAIR BRISS: Let's vote, please.

MR. WILLIAMSON: We will now vote on the performance gap. You may begin voting now.

We are still waiting on one. There we go.

We have eight high, eight moderate, zero low, and zero insufficient.

CO-CHAIR BRISS: Anything we haven't heard before on the evidence front?

DR. SHEA: No, I think in general it's the same body of evidence.

CO-CHAIR BRISS: So let's vote again for the fourth time this afternoon on

the same body of evidence. How's that?

MR. WILLIAMSON: We will now vote on the evidence. This is a yes, no or insufficient vote. You may begin now.

We have 13 yes, 1 no, and 2 insufficient.

CO-CHAIR BRISS: Reliability and validity, please, Lisa?

DR. SHEA: So similar types of reliability testing were done, as we heard, similar to the cardiac measure. And I'm looking here to get the specific numbers here. So in this one actually it did do a bit -- there was more stability so that 9 of the 16 states, or 44 percent, had no change in the performance quartile between the two performance years. And the R for that was 0.45.

CO-CHAIR BRISS: So again these data look a little familiar. And this is toward the upper half of this flock, so anybody got comments?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote on the reliability. This is a high, moderate, low or insufficient rating. You may begin now.

We are still waiting on two responses.

And we have 1 high, 15 moderate, 1 low, and 0 insufficient.

CO-CHAIR BRISS: Validity?

DR. SHEA: So regarding validity, the same type of face validity was assessed, and again, the group found that this was a helpful and useful measure. There was again the same sort of validity being done in terms of looking at -- regarding other screening measures.

CO-CHAIR BRISS: Questions? Comments? Concerns?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please.

MR. WILLIAMSON: We will now vote on validity. This is a high, moderate, low or insufficient rating. You may begin now.

We have 1 high, 15 moderate, 1 low, and 0 insufficient.

CO-CHAIR BRISS: And I'm suspicious that we may have a few issues that we may have heard before on usability and feasibility. Anything new?

DR. SHEA: No, in general it's the same data that was reported by the states and the panels in terms of the usability and feasibility.

CO-CHAIR BRISS: So any comments on usability before we vote?

(No audible response.)

CO-CHAIR BRISS: Hearing none, let's vote, please?

I'm sorry.

DR. ZUN: Perhaps this is just a

clarification. Is there a problem with usability if it doesn't clarify which LDL test? It says a LDL test. Are there multiple different tests that can be performed? So there's going to be difficulty using one test versus another, or --

DR. SCHOLLE: The specifications will define which CPT codes count and which tests count.

DR. ZUN: Because when I looked at the beginning it said "or an LDL test." It doesn't say which. Or am I confused?

DR. SCHOLLE: LDL-C. Is that the question?

DR. ZUN: I'm sorry?

DR. SCHOLLE: It's LDL-C test.

DR. ZUN: Okay. I thought it said something.

CO-CHAIR BRISS: So with that, anybody else, comments before we vote?

DR. ZUN: Okay. I'm sorry. It says "one or more of the tests." So you're

just implying that there may be multiple of the same test?

DR. BURSTIN: In a given year, yes.

DR. ZUN: Okay.

CO-CHAIR BRISS: So with that, let's go ahead and vote, please.

MR. WILLIAMSON: We will now vote on the usability. This is a high, moderate, low or insufficient rating. You may begin now.

We have 0 high, 17 moderate, 0 low, and 0 insufficient.

CO-CHAIR BRISS: And feasibility?

DR. SHEA: Ditto.

CO-CHAIR BRISS: Anybody have comments other than ditto?

(No audible response.)

CO-CHAIR BRISS: Let's vote, please.

MR. WILLIAMSON: We will now vote on the feasibility. This is a high, moderate,

low or insufficient rating. You may begin now.

We have 1 high, 15 moderate, 1 low, and 0 insufficient.

CO-CHAIR BRISS: And overall suitability, any last closing comments?

(No audible response.)

CO-CHAIR BRISS: Let's vote, please.

MR. WILLIAMSON: We will now vote on the overall suitability for endorsement. This is a yes or no question. You may begin now.

We have 17 yes, and 0 no.

CO-CHAIR BRISS: So we've gotten up to the last break of the day. Let's take 10 minutes and then we'll come and finish these up.

(Whereupon, the above-entitled matter went off the record at 2:26 p.m. and resumed at 2:44 p.m.)

CO-CHAIR PINCUS: Why don't we get

started? We have three more measures to do. So the next measure we're going to do is the emergency department utilization for mental conditions by people with schizophrenia, measure 1938. And so to hear from the measure developer and then Les is going to take the lead.

DR. SCHOLLE: So this measure evaluates whether people with a schizophrenia diagnosis have an emergency department visit for mental health. So this was part of our suite of measures to try to understand something about access to care for people with schizophrenia, and our expert group thought this was a critical way of assessing a poor outcome or poor access to care for people with schizophrenia. We went back and forth about whether this should be an emergency department visit for any problem or for mental health and then ended up with a measure focusing on mental health.

And so, as a state-level measure,

as a population-level measure or a health plan measure the idea is that it's really showing you where you have problems in the service system so that people with schizophrenia are ending up in the emergency room with their schizophrenia care rather than in outpatient care. It's clear that sometimes those emergency department visits are necessary, so this is a measure where you'd be looking to be able to make comparisons across organizations or states rather than saying this was always a bad thing to go to the emergency department, but the group felt strongly that we should be looking at this as a way of evaluating poor access to care, the bad outcome.

DR. ZUN: So I was asked to be the reviewer or presenter of this 1938. And so, as we walk through this, I'm going to try to leave my bias aside and do my best to discuss the measure here.

So the measure is looking at emergency department utilization for mental

health conditions in the subset of population of schizophrenia. And it is for those that -- again, we're back to that age group; 25 to 64, with a diagnosis of schizophrenia in an emergency department visit.

And so, let me go through the impact possibility and evidence, if I might. Okay. So we know from the evidence out there that patients with schizophrenia frequently used an emergency department, and we don't know if that's for good reason or bad reason. We do know that many of them have comorbidities and substance use.

There were three references noted on 1-A-4. I just happen to have pulled two of those three references and would like to quote -- actually, two of the three were by the same investigators. One was a VA population and one was a general population. And their conclusion in one paper was, "The relative rate of emergency department use may be suggestive of inappropriate use or may reflect

perceived barriers to care." And the other quote: "The overall increased use of services could be driven by increased severity of both mental illness and medical disorders." So the evidence doesn't seem to go along with the contention of the technical advisory group that put it together.

So as we go through -- I think that's just the 1-A component, so I'll stop. Thank you.

CO-CHAIR PINCUS: Any comments or questions in response to the developer?

DR. SCHOLLE: So I mean I think the concerns about evidence among the measures that we've presented today, this is the one that I think that we felt had the weakest evidence in terms of how do you know that this is something bad?

Now in our work and in our discussions with states and other stakeholders, trying to get a handle on potentially avoidable acute care like this is

of critical interest for states. But for a particular person, knowing whether or not that ED visit for a particular person is good or bad is fraught with problems. And we don't have a risk adjustment approach for saying is this appropriate for somebody who's more or less severe? And certainly trying to think about should we focus on -- the evidence review is focused on medical conditions, which was the original focus of the measure. And then it changed based on the field test.

CO-CHAIR PINCUS: Comments from the panel? Peter, Jeffrey, Tami.

Oh, no, we come back to Les.

CO-CHAIR BRISS: Yes, he'll get the last word whatever happens. So in this measure what's the sort of marginal utility of an ED utilization measure for this particular sub-population as opposed to a broader population? Can you give me a sense of the -- and so I'm not too troubled by the fact that this is likely a population that uses the ED

a lot and that some of it may be over-utilization. But what's the marginal gain about breaking out this particular sub-population and is there a broader measure that we could be looking at that might be stratified into this population?

DR. SCHOLLE: There is a broader measure that looks at emergency room utilization in a general population. This is different in that it's saying -- but I do not believe that it's endorsed. This measure is more like measures -- it flows more from the logic of potentially avoidable hospitalizations and potentially avoidable care.

And so, that's the logic that supported this. It's coming from the desire to try to see whether the service system is working well. So it's really trying to get at access to care rather than delivery of evidence-based treatment.

DR. MARK: Yes, I agree the

evidence is very weak. You know, use of ED could be a measure of access. These folks could be suicidal and you bring them to the ED and higher ED use could be preventing suicide. So It could be a measure of good access. You know, my concern is that it may be easy to lower the rate of ED use without improving outcomes or quality of care at all.

DR. SUSMAN: I guess I'm trying to understand the rationale, and it isn't altogether clear to me from the developer, are you trying to uncover misuse, overuse, under-use? I don't see the driver here. And without that clarity of sort of model, it's hard for me to get incited about a measure.

DR. SCHOLLE: I think the value of this measure would be to provide a standardized way of reporting emergency room use that would allow you to make comparisons across states. It's the kind of measure that states want to look at. The limitation is that we don't have a good way to risk adjust

for severity for people with mental health problems and for schizophrenia specifically. So if we looked at it compared to other measures that get at avoidable hospitalizations, we don't have that kind of risk adjustment approach.

So the motivation for it is to come up with standardized specifications that would allow states to make comparisons across states in terms of their utilization of emergency room. So it's value is in allowing for fair comparisons, but the interpretation of the result and what that means would depend on how this -- on the states use it.

CO-CHAIR PINCUS: I'm going to call on myself as a -- take the chair hat off and then turn it back to Leslie to summarize before we vote.

DR. ZUN: Can I give you my opinion, too?

CO-CHAIR PINCUS: Yes, you can.

But my own view is that I think it

is useful to have this kind of data. I'm not sure it meets the criteria for a quality measure according to NQF. You know, in previous studies, at least that our group has done, we've sometimes separated sort of what we've termed descriptive measures from quality measures. And this seems to me to fall more as a descriptive measure where -- but in some ways I think the real issue from quality is this is kind of a very indirect measure of sort of disengagement from care or lack of access to care. And it would be better to have a better measure of disengagement rather than using this sort of utilization measure.

And so, I guess, you know, in my opinion not as chair, but as a member, I just don't think it meets the criteria to be an NQF-endorsed quality measure.

DR. ZUN: Okay. Now I'm putting on my hat to comment on it as the presenting this. I'm very concerned about this. As a patient advocate I can see what's going to

happen to this data. They're going to take this data -- and this is impact. They're going to take this data and say that ED is over utilized by schizophrenic patients with mental health disorders. And they're going to stop payment, just like they've done in a number of states. They're trying to stop payment for emergency department visits if you're in the Medicaid program because we don't think they're necessary.

I am very concerned that this in fact will not provide the quality results that -- I think we all agree that that is a two-edged sword. I am very concerned that we're going to be looking at the other side of this coin where we're going to discourage schizophrenic patients from using emergency department services when they're in crisis because the states are no longer going to pay and discourage that use. I think that this is not the way to measure, not the way to impact, not the way to address the quality problem and

service delivery for this population.

Okay. I'm off my soap box now. Thank you.

CO-CHAIR PINCUS: So does anybody have something new to add?

DR. BURSTIN: Just a general comment, since the issue's been brought up of whether this actually overall even meets the sort of measure that could be brought into NQF for endorsement. Utilization measures that are clearly attached to a quality signal are fine. So for example, measures where there's an implication that -- for example, a readmission has a implication there potentially could have been a quality problem. We don't -- or a preventable ED visit. And we've looked at some of those measures. They have typically failed mainly because they're not yet reliably -- it's difficult to still assess reliability of preventability. That's a lot of abilities. But we do not actually have any measures that are pure utilization.

So I think the question you all need to determine at this table as the experts here and the multi-stakeholder experts is is there a quality signal here of a high ED use? Again, keep in mind this is a state to state comparison. We're not making inferences about providers, which I think would be far more problematic. It really is a significantly higher level of altitude measure.

CO-CHAIR BRISS: I have sort of a follow-on on Helen's comment. So I think I can imagine a utilization measure that was a cleaner signal of overuse of misuse, but I don't think that this is it.

CO-CHAIR PINCUS: As I understand it, if there's more, three or four as compared to one or two, then it doesn't move forward?

DR. SAMET: Can you just say the impact as it relates to this measure? Can you spell out what we're voting on right now? It's not the usual impact.

CO-CHAIR PINCUS: It's right up

there. It's whether --

DR. SAMET: No, I can read that, but sort of the impact of having a utilization measure about schizophrenics? Yes, I mean, it's whether a utilization measure such as this addresses a national health priority or --

CO-CHAIR PINCUS: Okay.

DR. SAMET: Yes, actually before you leave, could we maybe -- it's -- I think it would just be restating this --

CO-CHAIR PINCUS: Okay. Okay.

DR. SAMET: Can you guys vote before so we can -- I feel like we are in Chicago.

CO-CHAIR PINCUS: We're not voting early and often.

MR. WILLIAMSON: All right. Okay. We will now vote on impact. This is a high, moderate, low or insufficient rating. Begin now.

Okay. And we have 0 high, 1

moderate, 4 low, and 11 insufficient. The measure fails on impact.

CO-CHAIR PINCUS: Okay. So thank you. So we have two more measures. And Dodi is going to be lead for both of them. But first, they're both very similar measures that are proposed.

And I was wondering, Sarah, if you might sort of deal with both of them together?

DR. SCHOLLE: Yes.

CO-CHAIR PINCUS: So we're doing -- where did it go?

DR. SCHOLLE: Okay.

CO-CHAIR PINCUS: 1937 and 0576.

DR. SCHOLLE: Okay. So this is another one of those measures where we have a current HEDIS measure that looks at follow up after a hospitalization for mental illness. And then we also presented a measure that was stratified or sub-setted to look at follow up after hospitalization for schizophrenia.

So the measure, the first measure

has been in HEDIS now for about 10 years. The average performance rate at seven days is about 45 or 50 percent. So, and it has improved a little bit over time; those are the Medicaid rates, but really astoundingly poor. At 7 days, at 30 days the rate is closer to 70 percent.

We do see a disparity between our Medicaid -- the HEDIS health plan data and the data that covers all mental illness and all ages six and up. And the population that was sub-setted for schizophrenia where the rate for schizophrenia in this population, 25 to 64, that we measured from the Medicaid extract data we saw rates that were as much as 17 percentage points lower for people with schizophrenia. So that's follow up within 7 days or within 30 days of the hospitalization.

This measure does require that the follow up be with a mental health provider. Both require that. We've had a number of questions over time to include

follow up in substance abuse settings, follow up by telephone, follow up of other sorts. And our expert groups, both for the HEDIS measure and for the schizophrenia measure for states, have recommended that we continue to require that that follow up occur with a mental health clinician.

CO-CHAIR PINCUS: Sarah, could you just explain why you need to have a subset of just schizophrenia? Is there any difference in the behaviors or any difference in how it would be used? I mean, what's the reason for --

DR. SCHOLLE: The rationale is that the schizophrenia measure was part of that suite of measures for people with schizophrenia. So we would be happy to have that be a particular subset under the broader measure. So I think the logic that we talked about using with the other existing measures would work well here. And then the difference here is that that follow up after mental

health hospitalization measure for the general HEDIS set --

CO-CHAIR PINCUS: Okay. So then --

DR. SCHOLLE: -- is also being reviewed by this --

CO-CHAIR PINCUS: So then with Helen's blessing can we go ahead and just look essentially at 0576 with the assumption that there would be a substratum just focused on schizophrenia?

DR. BURSTIN: Duly blessed.

CO-CHAIR PINCUS: Okay. So, Dodi, do you want to start with impact?

DR. KELLEHER: This is a maintenance measure that was first endorsed in 2009 and is now up for maintenance review. Follow up after hospitalization for mental illness, not just schizophrenia. The measure assesses percentage of discharges in the age range of six years and older for those who are hospitalized for treatment and who had an

outpatient visit, intensive outpatient encounter or partial hospitalization with a mental health practitioner. And the rates reported are 30 and 7 days after discharge.

In terms of impact, this is a process measure that has data sources from claims, electronic clinical data and EHR. And multiple levels of analysis; clinician team, health plan, integrated delivery system, county, city, national regional and state. And there's ample evidence both that's been cited going back quite a few years. This actually has been a HEDIS measure since 1994, which sort of has a sad side to it since we don't seem to be getting very far with it. But that aside, there's plenty of information to show that this potentially could have great impact.

CO-CHAIR PINCUS: So there are any additional comments or questions with regard to the impact issue here.

CO-CHAIR BRISS: The triumph of

hope over experience.

CO-CHAIR PINCUS: Les?

DR. ZUN: I may be coming off in left field just a little bit, but you know, if we did emergency department follow up or a follow up after emergency department visit for mental illness, I think you'd be getting at the measure of accessibility and availability that we were trying to get at before, because I really think that's a much better measure of quality. So I'm sorry I digressed just a little bit off this measure, but I thought I had some valuable information.

And let me know if you need any help.

CO-CHAIR PINCUS: Other comments around impact?

DR. HANRAHAN: Yes, this is one of the gray areas or one of the more grayer areas. We just completed a transitional care study following people from inpatient to outpatient, and along the way we also looked

at Medicaid data and found that 93 percent of the people that are Medicaid-covered in the City of Philadelphia actually saw a primary care provider within the past year.

The other thing we found was that most of the people, when they left the hospital, the reason they had trouble or were readmitted was housing issues. So if they go out, either they don't have housing or they might be housed in group homes; and this population is more likely to be, or some kind of structured living situation. So I don't see that being easily captured in administrative data.

And we also found that the links to community following discharge really weren't a problem. There was no difference between our control and our experimental group in their links. And they were very good.

So, you know, after doing this study I have really -- I don't trust this information very much, that somehow the story

is that people don't get good follow up. I think it's confounded by enormously difficult social issues of housing and poverty and crime and unsafe neighborhoods. And I'm just dealing with one city and I really need to say that I can't call that a generalizable, but I have studied this area quite well.

So I would really say that I really don't know what the impact that this has other than for a lot of people collecting a lot more data about something that happens post-discharge that I'm not sure is really -- and for what, you know? If it's been around since 2004, I'm not sure it's really changing anything. So that's all.

DR. MARK: I mean, I can say I've used this data in various, you know, policy pieces and articles I've written and I found it very, very helpful to have this information. I also think, you know, it is discouraging to not see it move, but you know, we are getting a lot more energy around ACOs

and linkages and post-discharge follow up and transition care. So maybe now is the time when we'll see some kind of movement on it.

CO-CHAIR PINCUS: Jeff?

DR. SUSMAN: So my question; and either Dodi or the developer might be able to answer right off, is what diagnoses are you looking at and why did you exclude primary care? Was there an evidence basis for that?

I saw the list of diagnoses by, you know, ICD9 or -- but I don't know those off the top.

DR. SCHOLLE: So basically it's mental illness. So anything in the mental illness group. So it would include depression and --

DR. KELLEHER: Axis 1. DSM, Axis 1.

DR. SUSMAN: So, I mean, then it even further prompts my concern that by excluding primary care follow up, which is I understand part of this; if I'm wrong, please

correct me, is I think a much different paradigm than may of us are trying to create today.

DR. SCHOLLE: So I think the rationale, because we've had this discussion many times, is that in a world where it's really hard to get admitted to a hospital for a mental illness, and that is usually because of a suicide or some really difficult problem, that expecting a primary care provider to be able to handle that situation and handle follow up care is unrealistic. And so, that's where that discussion comes from.

DR. SUSMAN: I understand that. On the other hand, if there were appropriate coordination of care and communication, I would posit that it would be reasonable to follow up in other alternative settings. It's really the coordination of care, the discussion and communication that could either make or break such a transition.

DR. SCHOLLE: And again, this is a

measure that is comparative. When we look at the results, we see a lot of variability. Even though we see not much improvement over time, you know, the difference between the minimum and the maximum for Medicaid is -- for Medicaid health plans in 2011 was 11 percent to 87 percent with a mean around 45 percent. So we do see variation.

And measures like this help you understand how to compare things. And in different settings, you know, you might be able to say, well, that's because we have an alternative way of handling people who are discharged with a mental illness. But I think it really does point to a system that so far seems to be broken for most people and maybe there's a better -- maybe greater attention to this will resolve some of the problems and some of the issues.

Now we've tried to include care management and other kinds of encounters into measures like this. And our experience so far

has been that claims data are claims data and then those other data about the other kinds of encounters are somewhere else and are not frequently used enough for most health plans to combine the data. It's something that we're very interested in.

And as we see more information exchange and we see, you know, ACOs and different kinds of arrangements for people with dual-eligibility, we might get to a point where this isn't the right measure. What we need to be looking at is something like a care transitions measure or that gets at a patient reported experience. So we're aware of those things. But from claims data where you could just look to see who got hospitalized? Did they get something within 7 days or within 30 days? I mean, even when you look at the 30-day follow up, you know, the average for the 30-day follow up is 66 percent with a -- you know, between -- the 90th percentile is 82 percent. So we're still seeing -- you know,

even if you expanded that window to 30 days, because within 30 days certainly you'd expect the mental health professional --

CO-CHAIR PINCUS: It's a shame. It's really -- sort of the performance is shameful.

Jeff Samet, did you have --

DR. SAMET: No.

CO-CHAIR BRISS: Nancy, are you -- your thing is up. Do you still -- do you have another comment?

CO-CHAIR PINCUS: Okay. So any other comments on impact?

(No audible response.)

CO-CHAIR PINCUS: Okay. I guess we're ready to vote.

MR. WILLIAMSON: We will now be voting on the impact. This is a high, moderate, low or insufficient vote. You may begin now.

I think we're waiting on one.

CO-CHAIR PINCUS: Got it?

MR. WILLIAMSON: No, we're still missing one. There we go. I knew I wasn't going crazy.

All right. Here we go. We have six high, seven moderate, one low, and zero insufficient.

CO-CHAIR PINCUS: Okay. Let's move on to gap.

DR. KELLEHER: Well, and I think that we just went over the gap.

CO-CHAIR PINCUS: Yes. So is there any further comments about the gap?

CO-CHAIR BRISS: There's not a category for shameful.

MR. WILLIAMSON: We will now vote on the performance gap. This is a high, moderate, low or insufficient rating. You may begin now.

We have 10 high, 4 moderate, 0 low, and 0 insufficient. And one shameful.

CO-CHAIR PINCUS: Okay. Evidence?

DR. KELLEHER: Again, there's

ample citations and guidelines used for evidence that when there isn't follow up there's a poor outcome, and when there is follow up there's a better outcome. So, I don't know, do we --

CO-CHAIR PINCUS: Nice succinct statement of a summary.

DR. KELLEHER: And I haven't been quoting them, but we should look and see what our subgroup thought about all this, since I bugged other people about it. So on -- yes. I'm in the wrong place.

CO-CHAIR PINCUS: While Dodi's sort of --

DR. KELLEHER: We had six who thought the evidence was there and one that did not in our smaller group.

CO-CHAIR PINCUS: Any comments, discussion further with regard to the evidence?

(No audible response.)

CO-CHAIR BRISS: Okay. Ready to

vote?

MR. WILLIAMSON: We'll now vote on the evidence. This is a yes, no or insufficient rating. You may begin now.

And we have 13 yes, 1 no, and 0 insufficient.

CO-CHAIR PINCUS: So moving onto reliability.

DR. KELLEHER: The reliability was an estimate using a beta-binomial model, so essentially ratio of signal to noise with zero to one with 0.7 being considered very good. The rate for 30 days in all commercial, Medicaid and Medicare populations were 0.949 or better. The percentage members receiving follow up within seven days in all three populations were 0.95 or better.

CO-CHAIR PINCUS: Helen?

DR. BURSTIN: Question for NCQA. Just the level of analysis on this one also just says state, but I thought it was endorsed for health plan as well. Is that an error?

DR. SCHOLLE: It should have said health plan.

DR. BURSTIN: 1937 just has state. I assume that should be --

DR. SCHOLLE: Oh, I'm sorry. You're looking -- 576 is the one --

DR. BURSTIN: Oh, I'm sorry. I'm on the wrong one. Never mind.

DR. KELLEHER: So this is health plan only.

CO-CHAIR PINCUS: Yes, this is the health plan only.

DR. BURSTIN: Too many open.

CO-CHAIR PINCUS: Yes. Any comments about reliability?

(No audible response.)

CO-CHAIR PINCUS: Okay.

MR. WILLIAMSON: We will now vote on the reliability. This is a high, moderate, low or insufficient rating. You may begin now.

And we have eight high, six

moderate, zero low, and zero insufficient.

DR. KELLEHER: Validity. Measure was written, field tested and presented to CPM, Incorporated by HEDIS in 1994. Wow, and I remember that. That tells you how old I am. And so, given that there's actually some ongoing validity data, so if we turn to the results -- I think that's -- am I looking in the right place? 2-B-5.3? No? Could you scroll more? You'll see the 7-day rates for commercial, Medicaid and Medicare and then followed by the 30-day rates from 2009, 2010 and 2011.

CO-CHAIR PINCUS: Nancy?

DR. HANRAHAN: I have a question for the group over here. When I look at these rates, I really don't see much change over the three years that they're reported, which, you know, draws into question how useful the measure is, to me anyway. I see, you know, devastatingly poor follow up after the hospitalization.

And I know you said, Sarah, that, you know, as we get more involved in continuity of care and collaboration that that might change, but wonder what you think about that, or what the group has thought -- talked about.

DR. SCHOLLE: So actually if you looked at the full set of HEDIS measures and the longer time frames, look at blood pressure control and see this nice curve up. You look at diabetes control and you see the same kind of thing. You look at the behavioral health measures and generally you see kind of a -- not much improvement. Now what contributes to that, I certainly think there's a number of things that have to do with the way that we pay, and we have differences between how managed care separates out behavioral health versus general medical care. It has to do with psychiatrists not talking to primary care docs. It has to do at a number of levels patients not being -- wanting to address these

issues. Stigma I'm sure plays some issues.

Is it a valid measure? Yes. Is it making the world move? No, but that shows us that measurement alone isn't going to lead to improvement. And so the question is, well, this hasn't improved at all. Should we get rid of it? We have retired measures that have topped out. This is not one that's topped out, so that wouldn't be a reason to retire it.

I am encouraged though because I think there's more interest in behavioral health now than there has been. I think the reporting of measures through the Medicaid core set, the greater attention to behavioral health issues is that's contributing to why are other expenses, costs of care contributing -- I think there's a lot more interest in these measures. It's still going to be hard to improve it.

So we think that it's a valuable measure. We would like to see it remain

endorsed and we think that it's used in the core set and in other programs will help to draw some attention to it. And the kinds of interventions that are happening to try to improve care for people with duals and improve follow up after hospitalization for both mental health and general medical conditions may be -- that may be the kind of thinking that will help to spur this.

CO-CHAIR PINCUS: Lisa?

DR. SHEA: Just to follow up on that, anecdotally I know in my state now one of the major insurers is starting new programs where people who are in the jurisdiction will be offered case management services and so forth to link them to their next level of care which they haven't provided before. So it does seem to be spurring the insurers to provide some care.

CO-CHAIR PINCUS: And I think that's happening in New York State also. There are initiatives to actually make -- put

some -- you know, as Sarah said, you know, that it's one thing to measure something, to actually put policy and practiced teeth into it, and it is just now where that's actually beginning to happen.

So why don't we vote on validity and then go back to your issue?

DR. BURSTIN: Okay.

CO-CHAIR PINCUS: Okay? So why don't we vote on validity?

MR. WILLIAMSON: We will now vote on validity. This is a high, moderate, low or insufficient rating. You may begin now.

We have five high, eight moderate, 0 low, and 0 insufficient.

CO-CHAIR PINCUS: So the issue that Helen had was when we previously voted on reliability -- and also just in general the statement in front of the measure is that it goes all the way from state and health plan down to clinician. And the question was to what extent has reliability been appropriately

tested at the clinician level?

DR. BURSTIN: And further, the submission form, now that I'm on the right form; I knew there was something here, says, "This measure has not been tested by NCQA to distinguish individual clinician-level performance." So just to qualify, we then can't endorse it at that level. So we would need to have that modified, not have every box checked unless you have additional evidence to bring forward.

DR. SCHOLLE: I know that we've been asked to specify it at the clinician level for electronic health records, and that's probably where the specification came from. I just don't have any that address this measure, I believe. So that's probably where that came from. But we don't actually collect data at the clinician level, so we will -- can remove that.

CO-CHAIR PINCUS: So let's move to usability.

DR. KELLEHER: Usability. The current use is for public reporting, regulatory accreditation programs, quality improvement and benchmarking, external benchmarking over multiple organizations and then internal quality improvement within a specific organization.

CO-CHAIR PINCUS: Further discussion on usability? Peter, then David.

CO-CHAIR BRISS: Yes, I'm very stuck on -- I can't imagine that this measure could have utility at the individual clinician level. Maybe there are half a dozen psychiatrists in the country who see enough patients in the people to be able to make this a meaningful measure, but it would be good for NCQA to actually think about that. I can't imagine that it can make sense.

DR. BURSTIN: And actually, just to speak to that, in the next round you guys will be reviewing the Joint Commission inpatient psychiatry measures, which I believe

several are about transition to outpatient. So I think the last time we discussed this we encouraged Joint Commission and NCQA to talk about this. It seems like there are some real opportunities there to kind of link some of those measures and get at provider-level measures that in fact allow you to make that link, but certainly not clinician would be hard to do.

DR. EINZIG: So I'm speaking at the level of a pediatrician managed health psychiatrist and wondering about the criteria of should this be separated for children versus adults? Majority of kids don't see child psychiatrists or psychologists. Most psychiatric care is provided by the pediatrician primary care doc. So I'm thinking -- so if a kid gets discharged and if they have to see a behavioral specialist, somebody that they don't have any relationship with, no prior, you know, experience with, could that do more harm than good if it's not

a good fit?

So in other words, if they see that person and it's not a good relationship, not a good rapport, would that turn that patient family off to receiving further care down the road versus referring to the primary care doc, where they have that relationship since birth, and then go from there?

DR. SCHOLLE: So a couple things. One is that the reporting is stratified by age, so it's reported both for children and for adults and then combined, and it's in the children's core set for Medicaid.

I think the argument about is primary care follow up sufficient/adequate for children, again, I mean, I would have to go back to the argument that I used just in general. I don't know why it would be different for kids. If kids are sick enough to be hospitalized for a mental health problem, then I'm not sure pediatrician -- I agree we've got the coordination with the

pediatrician, but having care in the pediatric office where there isn't a mental health clinician to manage the follow up is, our panels would say, not enough for a kid who's sick enough to be hospitalized.

DR. EINZIG: But there aren't child psychiatrists there. That's the other issue.

DR. SCHOLLE: But it's not just psychiatry that counts. I mean, it's mental health clinicians. So other kinds of clinicians would count. I understand the shortage of child psychiatry, but I think the issue is about a licensed mental health clinician.

CO-CHAIR BRISS: I had sort of a similar issue about -- I understand your arguments about a mental health professional. I was wondering about -- have you had feedback about rural areas? I mean, there are some rural areas where -- that have sort of -- for adults that have sort of similar capacity

issues. And I wonder if we could actually be creating unintended effects by trying to fit every square piece into a round hole.

DR. SCHOLLE: So remember, it is a population-based measure. So like, I mean, all of the measures that we talked about today, they're really population-based. And I don't think 100 percent is the goal for everything. You know, but certainly 45 percent sounds pretty lousy to me. So I think we're somewhere in between there. Are we trying to shoot for 100 percent? I don't think so.

So in terms of the unintended -- the rural -- and we've had recommendations for incorporating tele-monitoring, like so where there's an opportunity for that. If those visits get billed to the health plan in the same way that a face to face visit were billed and we could track it in the claims data, then that could be counted.

I think what we've been struggling

with is how does that happen and does it get documented? And is that in a place where it's going to show up in the claims data for a health plan, or for a Medicaid program or whatever? So we're open to that.

And I think if we started to see that this measure started to decrease because those alternative activities were happening, and we could document where it's going, I think we would look at that. We don't have a sense that that's widespread and that's contributing -- that it's -- that we're not counting, you know, people who are getting adequate services. And we still get the sense that we just have lousy performance on these measures.

CO-CHAIR PINCUS: Tami and then back to Dodi.

DR. MARK: Yes, I mean, I'm sympathetic to this view that we should debate the definition, but I think we also need to weigh it against the fact that we have very

long-term trend data. And if I think about the number of measures, which we have quality measures going back 10 years for mental health, it's probably this one and maybe another two, you know? There are so few measures for mental health quality over time.

So even if this is an imperfect one, at least it's, you know, had a long trend on it. We can see what happens now if we do all this stuff related to follow up and transition care in ACO. So there's an argument to be made for keeping the imperfect ones so you can follow it over time.

CO-CHAIR PINCUS: Dodi?

DR. KELLEHER: So to piggyback, I agree it's worth keeping, but I'm thinking that, or I'm hoping maybe; because I'm an optimist at heart, that with the maturation of the use of electronic health records, with the new models coming out with the practice patterns changing, a lot is going on that we're not quite there. And I think that's

been apparent actually over the full two days, you know, where we want to go to that next step. But I think if this is endorsed, probably before you come on back, I think it would be worth the while to review all those areas and decide whether this really is usable and feasible in the form it's been in for almost 20 years.

CO-CHAIR PINCUS: So we're ready to vote on usability. Oh, I'm sorry.

DR. NAEGLE: I just have one little comment and I wanted to follow up with David's point, so, and I think these points are well taken. So when it does come back or in the time, intervening time, if we could give special consideration to population needs, one of them being children and families. The other being older adults, especially older adults with depression who do not seek care and psychiatric services even after they're hospitalized. So thinking about how we may be missing a number of populations

I think would be helpful.

CO-CHAIR PINCUS: All good points. Vote on usability?

MR. WILLIAMSON: We will now vote on usability. This is a high, moderate, low or insufficient rating, and you begin now.

All right. We have 3 high, 10 moderate, 0 low, and 1 insufficient.

CO-CHAIR PINCUS: Feasibility?

DR. KELLEHER: I don't have a lot to say about this. Okay. I think it's -- I'm like Harold. I keep forgetting my microphone. Let's see. Sorry, I've lost my place.

CO-CHAIR PINCUS: Well, in some ways, you know, the fact that it's been collected for 20 years --

DR. KELLEHER: Right, we've sort of --

CO-CHAIR PINCUS: -- suggests that it's feasible.

DR. KELLEHER: -- gone over it. You think?

You know, and I think the concerns have been raised as well when we were talking about reliability in terms of where this needs to go.

CO-CHAIR PINCUS: Any other issues about feasibility?

(No audible response.)

CO-CHAIR PINCUS: Vote?

MR. WILLIAMSON: We will now vote on the feasibility. This is a high, moderate, low or insufficient rating, and you may begin now.

We have seven high, six moderate, zero low, and one insufficient information.

CO-CHAIR PINCUS: So overall suitability for endorsement. Other comments? Anything additional that people would want to add? I don't know, Madeline, was yours up from before?

DR. NAEGLE: Thank you. Not sure it belongs here, but just to reinforce, maybe while we're finishing this up, to think for

our developers and also for our considerations that we're really not getting at information about highly vulnerable groups who receive disparate care. And I would include not just the frail elderly, but the sexual minorities who really don't rise to any kind of level of our being able to assess where we are in terms of our data collection, and even measure development, I would suggest.

But also; and then this isn't only because I'm old, I think we need to begin to expand our strata when we -- even with people with schizophrenia, you know, understandably, who live 25 years less than the general population. I think we need to get beyond 64 and we need to begin thinking about elderly people between 65 and 75, and 75 and 90, because those numbers are growing and we are not ready to manage their general care, certainly not ready to manage their behavioral health needs. So those are my thoughts.

CO-CHAIR PINCUS: Other items?

Other issues?

(No audible response.)

CO-CHAIR PINCUS: So ready to vote on overall suitability for endorsement.

MR. WILLIAMSON: We will now vote on the overall suitability for endorsement. This is a yes or no rating, and you may begin now.

And we have 13 yes, and 1 no.

CO-CHAIR PINCUS: So does this mean we're done?

MR. WILLIAMSON: Just about.

DR. BURSTIN: The only question is if there are any specific advice on the specific strata for schizophrenia in this measure. I know there was the age issue. And I don't know if there's anything else. Otherwise, we can bring that back to you offline.

CO-CHAIR PINCUS: And I guess just one other issue just with regard to this measure, Sarah, is, you know, in terms of

since you're already stratified, there may be other strata to consider that came up in the discussion, including mortality of a potential strata.

DR. SUSMAN: You know, one other issue. Again, I'd be interested in seeing if you can provide it, and it's relatively easy, is just the difference in the measure if you specify psychiatry mental health versus primary care. Because I just feel like we have been pushing on this trying to develop a mental health behavioralist outlook on this, and it isn't getting us very far. And maybe we need to think more about team-based care and other alternative approaches.

DR. KELLEHER: And sort of the last suggestion, and this comes out of my own experience in community mental health for many years, is at least in California there's a lot of sub-acute care that's 24/7 and it's just as imperative there that there would be a good transition of care and follow up as there is

in the acute hospital, and I've never seen that addressed. I don't know if it is addressed somewhere else, but I haven't seen it. And I think that sort of goes to, you know, again other levels of care and that could either be defined as acceptable outpatient follow up or needs to be addressed in terms of the denominator and the numerator.

DR. SCHOLLE: And just so that you know that all the recommendations for new measure concepts or new ways to thinking about this will not go unheeded, we actually are working with Mathematica and ASPE and SAMHSA on a project right now to look at new measure development related to behavioral health. And so, many of the topics that you've recommended are already in our evidence review process. And I'm not sure they're going to survive discussion based on our understanding of your evidence requirements, but we'll do our best.

CO-CHAIR PINCUS: So it sounds like we're done. I think from my point of

view I'd like to really thank the Committee for really its incredibly hard work getting a lot done very efficiently, and really taking everything very seriously in terms of how to think about these issues which have important national impact, and to thank the staff and the measure developers for the work that they've done, all of which really, you know, are required for this process, and to thank my co-chair.

CO-CHAIR BRISS: hanks to everybody from me, too. I have to stop one last time and ask for public comment one last time.

CO-CHAIR PINCUS: Oh, yes.

(No audible response.)

DR. ZUN: It's quite a reflection on the chairs as well that we can be so efficient.

CO-CHAIR BRISS: Thank you. Hearing no public comment, I think we're adjourned.

MS. FANTA: And one last thing really quick. We sent out an email last week about the follow up calls, so we've scheduled that based on everyone's -- well the majority's availability. So that next call will be April 24th, which is a Tuesday, from 12:00 until 2:00. So hopefully you can all make it. And I'm sure on behalf of the entire --

PARTICIPANT: 24th is --

MS. FANTA: I'm sorry?

PARTICIPANT: -- next Tuesday?

MS. FANTA: Next Tuesday, yes. I just want to thank you all for your thoughtful participation and for coming out to D.C. Thanks.

DR. BURSTIN: It's not clear we need it. We could do some of this on email if we need to. We'll get back to you. You guys are busy.

(Whereupon, the meeting was adjourned at 3:45 p.m.)